Table 19.3. Analysis classified by state in water (1.0 g/L) (GHS, Bottom-up, TEA)

Regulatory System	Water Solubility ≥ 1.0 g/L	Water Solubility < 1.0 g/L
Accuracy	47.5% (29/61)	42.3% (11/26)
Sensitivity	43.2% (19/44)	85.7% (6/7)
Specificity	58.8% (10/17)	26.3% (5/19)
False Negative Rate	56.8% (25/44)	14.3% (1/7)
False Positive Rate	41.2% (7/17)	73.7% (14/19)

Table 19.4. Analysis classified by state in water (10.0 g/L) (GHS, Bottom-up, TEA)

Regulatory System	Water Solubility ≥ 10.0 g/L	Water Solubility < 10.0 g/L
Accuracy	44.0% (22/50)	48.6% (18/37)
Sensitivity	38.5% (15/39)	83.3% (10/12)
Specificity	63.6% (7/11)	32.0% (8/25)
False Negative Rate	61.5% (24/39)	16.7% (2/12)
False Positive Rate	36.4% (4/11)	68.0% (17/25)

Table 19.5. Analysis classified by state in water (100.0 g/L) (GHS, Bottom-up, TEA)

Regulatory System	Water Solubility ≥ 100.0 g/L	Water Solubility < 100.0 g/L
Accuracy	44.4% (8/18)	46.4% (32/69)
Sensitivity	30.8% (4/13)	55.3% (21/38)
Specificity	80.0% (4/5)	35.5% (11/31)
False Negative Rate	69.2% (9/13)	44.7% (17/38)
False Positive Rate	20.0% (1/5)	64.5% (20/31)

Table 19.6. Analysis after cut log D (2.88) (GHS, Bottom-up, TEA)

Regulatory System	logD ≥ 2.88	logD < 2.88
Accuracy	40.9% (9/22)	47.7% (31/65)
Sensitivity	100.0%(4/4)	44.7% (21/47)
Specificity	27.8% (5/18)	55.6% (10/18)
False Negative Rate	0.0% (0/4)	55.3% (26/47)
False Positive Rate	72.2% (13/18)	44.4% (8/18)

Table 19.7. Analysis after cut log D (1.70) (GHS, Bottom-up, TEA)

Regulatory System	Log D ≧1.70	Log D < 1.70
Accuracy	50.0% (19/38)	42.9% (21/49)
Sensitivity	78.6% (11/14)	37.8% (14/37)
Specificity	33.3% (8/24)	58.3% (7/12)
False Negative Rate	21.4% (3/14)	62.2% (23/37)
False Positive Rate	66.7% (16/24)	41.7% (5/12)

Table 19.8. Analysis after cut vapor pressure (6.0kPa) (GHS, Bottom-up, TEA)

Regulatory System	Vapor pressure ≥ 6.0 kPa	Vapor pressure < 6.0 kPa
Accuracy	36.4% (4/11)	47.8% (33/69)
Sensitivity	28.6% (2/7)	51.2% (21/41)
Specificity	50.0% (2/4)	42.9% (12/28)
False Negative Rate	71.4% (5/7)	48.8% (20/41)
False Positive Rate	50.0% (2/4)	57.1% (16/28)

SIRC Peer Review meeting minutes

Tuesday, March 3, 2015

Kojima	We hope to get your approval for this TG and to submit it for approval six months from now. This is the kick-off meeting, and this morning we will review the work thus far, and then begin discussions after lunch. Self introductions Review of Evaluation Criteria
Hoffmann	Just to confirm, we will discuss this week and have follow-up teleconferences as necessary to complete the peer review by August.
Kojima	That is correct.
Hagino	Overview of test method (See handout)
Hoffmann	Please explain the other endpoints in the SIRC-CVS study.
Hagino	Results were the same for the various endpoints. Very strong correlation between endpoints.
Jeong	What were the conditions for measuring OD?
Hagino	Optical density is measured in dry conditions
Hoffmann	The MHW study says acceptance as non-irritant without animal tests. Was that applied?
Kojima	Only draft. Not applied.
Hoffmann	Was the data analyzed before the validation study?
Hagino	Yes, it was.
Eskes	When was this data obtained, in 2009 to 2011?
Hagino	Papers issued in 2008 and 2010 evaluated 10% concentration.
Eskes	In the reevaluation (old data from MHW Project and new measured data from Shiseido) study, was TEA used as relative control?
Hagino	Yes.
Eskes	What was the concentration?
Hagino	They were neat chemicals.
Eskes	The first project from '91 to '99 used a different cut off?
Hagino	Both tested TEA, but not as relative control. Tween20 was used as the relative control.
Kojima	Japanese reviewers recommended additional study using relative control.
Hoffmann	Doesn't the introduction mention polyoxyethelene sorbitan monolaurate (Tween20) as reference substance?
Hagino	Yes.
Jeong	It is confusing because you said polyoxyethelene sorbitan monolaurate (Tween20) was used as reference substance.
Hagino	Need to consider neat chemicals for GHS.
Jeong	The concentrations mentioned in the report are confusing.
Takeuchi	Stock solution was 10%.
Eskes	Does slide ten shows neat or 10% concentration?
Hagino	Neat. Some of the data has TEA data and some does not.
Hoffmann	I don't understand how you can reanalyze the data without TEA data.
Hagino	TEA was used as a test substance although not as reference substance.
Hagino	Here is a summary of TEA protocol. (See handout of presentation.)
Takeuchi	Two runs means two plates or one plate?
Hagino	Two separate plates.
Jeong	You said TEA is best for relative control. TEA is irritant in EPA (Cat. III). That is a problem. Then isopropyl myristate seems a better choice than TEA.

Hagino	TEA is best of substances for which IC50 is available.
Jeong	Why not use cut off values instead of relative control?
Hagino	MHW Project members recommended relative control.
Jeong	I think relative control will introduce some problems going forward with many chemicals. I have a concern that there are many chemicals on the TEA borderline.
Eskes	Please show slide explaining sets and runs. Are the runs done on separate plates on different days? What is the difference between sets and runs?
Sunouchi	One run is one plate, tested on a different day from the other plates. Two or three runs per set.
Hoffmann	What does "series" mean in Quality Control 4?
Hagino	The dilution series on one plate.
Hoffmann	What is the background for the OD? In other words, the PBS. I want to know if it is subtracted from the reading before.
Hagino	It is very low.
Takeuchi	Is it low enough to be ignored or should it be subtracted?
Hoffmann	How did you establish quality control ranges for SDS and TEA?
Takeuchi	We cannot make any conclusion about the chemicals that give positive test results, so perhaps best to remove GHS Category 1 or 2 indication.
Eskes	When using DMSO, you first dissolve in DMSO and then add to medium?
Hagino	First, dissolve in DMSO and then add to MEM.
Jeong	Does this procedure allows different laboratories to use different vehicles?
Hoffmann	You can seen in Tables 1–3 that different laboratories used different vehicles.
Eskes	Was there any effect of using different solvents on misprediction? (false positives, etc.)
Sunouchi	Here is an explanation of the Validation Study (See Validation Study Report)
Eskes	Same chemical code for all runs?
Kojima	See page 3 and 4 of Appendix 3, which shows that there were lab codes assigned for each run.
Eskes	Where were the labs?
Kojima	Tokyo, Osaka, and Korea.
Eskes	Did they participate in earlier validations?
Kojima	No, they were all naïve laboratories.
Hoffmann	Did they participate on a voluntary basis or were they remunerated?
Kojima	Voluntary.
Eskes	Did observers take part in chemical selection discussions? May be important to indicate that they did not take part in those discussions, if they did not.
Eskes	Was there practical training or just video session?
Hagino	They watched the video and discussed the protocol, but did not actually perform the procedure at that time.
Eskes	Can you summarize the differences between the training protocol and the final protocol?
Eskes	Is there a limit beyond which the cells cannot be used? Is there a maximum number of passages? This should be reflected in the protocol.
Hagino	The stock solutions are prepared from the seed cells and must be used within 3 months of start of cultivation, but they should not be refrozen. Three months means about 24 passages.
Jeong	Is there a rationale behind the three months? Is that the recommended length?
Hagino	Quality of cells is checked per protocol.
Eskes	Do you have an absorbance spectrum for the crystal violet?
Hagino	Main absorbance is at 588 nm, but 570 nm is close enough to give same result.
Kosaka	You have data for both 588 and 570 nm. Are the results similar?
Sunouchi	Yes, from the lead lab as well.

Takeuchi	The protocol says only 588 nm, but the report mentions both 588 and 570 nm, which is confusing.
Sunouchi	The equipment at one laboratory was only able to measure at 570 nm.
Hoffmann	Amend protocol to allow also measurement at 570 nm.
Takeuchi	Perhaps some quality control criteria for TEA is needed to ensure quality at new laboratories in future.
Hoffmann	Is there a quality that needs to be maintained? If so, that should be indicated in the protocol. What was the quality for the validation study? Was it specified? Please specify in protocol.
Jeong	Same for the color of the test substances. There might be some interference.
Takeuchi	There might be some substances with a color that causes interference.
Jeong	Perhaps it should be specified that all color should be removed before measurement.
Hoffmann	Probably interference is not a problem, because liquid is removed before measurement.
Eskes	Can we see the individual animal data?
Kojima	There were three chemicals for which there was no animal data. We have data for 115 chemicals.
Eskes	Would like to see raw in vitro and in vivo data for TEA. It would be interesting to see this raw data and for the controls, as well. Data for the other chemicals can be provided later.
Hoffmann	If the data exists, it should be available.
Kojima	Bill Stokes mentioned that he collected data. The STE test chemicals are also similar.
Hagino	There is animal data for TEA tested at 10% and neat.
Hoffmann	Do we need access to all data or just to specific chemicals?
Eskes	We need at least TEA and controls. But would like to have raw data as much as possible. In vivo data for at least TEA. And then all validation data. Also, the raw data for tables 8.1, 8.2, and 8.3 showing OD range.
Hoffmann	When Lab C was requested to retest, what modifications were requested?
Sunouchi	We didn't send them any suggestions.
Yamashita	Explanation of tables 16 and 18.1 to 19.8
Eskes	How were the chemical classes of the substances defined?
Kojima	Based on INCI.
Eskes	There appears to be some relation with molecular weight. Do you have any ideas about why this might be?
Hagino	Here is why we set the applicability domain to molecular weight of 180 or higher. (PP presentation)
Hoffmann	You have observed this phenomenon before. Why were so many substances with small molecular weights included in the test?
Hagino	Because of the need to use substances for which in vivo data was available.
Hoffmann	That needs to be made clear in the report to show that you had a clear hypothesis and to avoid the appearance that you did data mining after the tests to improve results.
Takeuchi	Do you have examples of organic solvents that enter the lipid of the cell membrane?
Jeong	Perhaps need to clarify which substances enter the lipid and which do not, which will support your hypothesis.
Eskes	Why could you not calculate IC50 for P3-066.
Sunouchi	There was precipitation that prevented determination of IC50.
Eskes	In which case, indicate that in the report.
Jeong	On page 23, Table 11.2, there are many values beneath the 39.1 threshold, which

should be shown as < 39.1 rather than as an absolute value. And there are many other values in the tables that need to be corrected.

Endo Explanation of the mean value of <39.8 for substance P3-021 in Table 11.3.

Eskes Need to describe how the statistical analysis was conducted. There is no section describing that in the report. Also, do you know of any studies that show how using TEA or serum from different manufacturing lots affects test results?

Hagino The MHW Project and other in-house data includes different manufacturing lots. See page 39 of Appendix 5. The results are similar to results of this validation.

Hoffmann Did you see any irregularities in the curves used to determine IC50?

Hagino The data for this study was very regular,

Hoffmann Did you consider using nonlinear regression in determining IC50?

Hagino We got good results with linear interpolation and did not consider other methods.

Hoffmann Include in the protocol that the first IC50 intersection is used for non-monotone curves.

Hagino I would like to explain Section 5 Discussion (See section 5)

Eskes The focus of the discussion should be on understanding why misclassification occurred instead of how to improve accuracy by defining the applicability domain.

Hoffmann Interesting that IC50 of 1600 had similar results. Perhaps you can simplify the protocol by using IC50 of 1600 instead of relative control. How does the fixed cutoff perform with the reduced applicability domain? The positive and relative control are clearly defined.

Wednesday, March 04, 2015

Hoffmann TEA and animal data is requested as soon as possible.

TEA has a molecular weight below 180. Does that cause any inconsistency? Please think about whether that has any effect on the discussion. Do we have information on purity of TEA?

Kojima Purity of TEA is > 98% and molecular weight is 149.

Eskes It is difficult to find a suitable relative control. I don't think the molecular weight is that significant to the results.

Hoffmann The common impurities monoethanol and diethanol amine which are both corrosive, but 98% purity would seem to be high enough.

Eskes About the protocol, the reagents were hydrochloric acid and sodium hydroxide. PH adjusted.

EPA is mentioned but not explained. Should it be explained?

Takeuchi EPA is mentioned in study plan, so report should follow that.

Kojima EPA is primarily pesticides but this study focuses on cosmetic ingredients, so we decided to delete references to EPA.

Eskes Main focus should be on GHS. EPA is US only, so if we want a global standard, we should focus on GHS.

Hoffmann There is no mention of EPA in Study Plan.

Eskes But in second version and later, EPA is mentioned.

Eskes We have data for and can review the protocol based on GHS, but it would be difficult to do so based on EPA. But we can leave the EPA references in because it was in the Study Plan.

Hoffmann We should consider EPA unless directed by JaCVAM not to. But based on documentation alone, we need to consider it.

Analysis of EPA is needed unless we receive direction that it is not needed.

Jeong Protocol calls for 72 hour exposure, but volatile substances will evaporate before 72 hours is up. How do you determine contamination by volatile substances. Also, how

	do you expose a volatile substance for 72 hours?
Hagino	We have quality control data. We compare wells of negative control wells on opposite sides.
Jeong	That is an indirect method, but if it is really contamination, it would be detected in the next well. So the description needs to be improved. Also, for crystal violet, I found that it binds with DNA to stain. In this protocol, we dry the well to measure OD. But the cells should be confluent in the well to measure correctly. If not, observed values could be inconsistent. But the stained part will be nuclei, not cytoplasm. This could affect measurement. So there are advantages and disadvantages for TEA, but crystal violet does not stain homogenously.
Hoffmann	Have you observed this phenomenon?
Hagino	CRV stains entire cell. But sometimes not entirely homogeneously.
Jeong	I'm just wondering if there is any way to enhance accuracy.
Hoffmann	My impression is that measurements are relatively consistent.
Takeuchi	Do we have an example of what the stained cells in the well look like? Maybe CRV staining is very common, but need to add information about staining.
Hoffmann	What about the 72-hour exposure and what happens with volatile substances?
Hagino	By comparing negative control cells on either side of plate, we can see if there is any contamination from volatile substances.
Jeong	What is your experience with this kind of contamination?
Hagino	There was some contamination in our testing.
Takeuchi	Do you have an example?
Hoffmann	But if it's not cytotoxic, we won't see it. The applicability domain shows loss of accuracy for volatile substances. See Table 19.8. Which raises another question. Only 80 substances are shown in Table 19.8. Why not 115? This needs to be explained for all these tables. We need more explanation of how volatile substances are handled. We need an answer as to why exposure is 72 hours.
Jeong	If confluent and multilayered, then there is no problem. But if the cells become detached, then there will be blank spots.
Hagino	In vivo tests are non rinse and long-time application must be considered. So we wanted to have long-time contact with cells in our protocol. The reaction peak for in vivo tests is generally before 70 hours, so we used 72-hour exposure for our protocol.
Jeong	If cytotoxic, then we know after 72 hours
Hoffmann	So this mimics the exposure and conditions of the in vivo test.

Draft Peer Review Panel Evaluation of the SIRC-CVS:TEA Assay for eye irritation potential

Ver.1.0

Japanese Center for the Validation of Alternative Methods
National Institute of Health Sciences
Tokyo, Japan

11 March 2015

SIRC-CVS:TEA peer review: Draft initial PRP replies

Evaluation Criterion 1: A rationale for the test method should be available, including a description of the human health effect, a clear statement of the scientific need, and the regulatory application

1.1 This method is proposed as an alternative to the Draize eye irritation test, which predicts ocular hazard to humans.
1.2 It is proposed for use in a bottom-up approach for distinguishing non-classified test chemicals from classified ones according to GHS and for distinguishing Category IV test chemicals from other categories according to the EPA classification system.
1.3 Although the validation report specifies cosmetic ingredients, we understand this test method to be applicable to substances in general, because the validation included a broad range of chemical classes.
1.4 The report indicates the ethical need of the test method for animal welfare reasons, but it would be helpful to specify also scientific needs.
1.5 It would be helpful to describe advantages of this method vs. other existing in vitro test methods such as STE, RhCE, etc.
(NB: No local regulatory applications were considered by the PRP.)

Evaluation Criterion 2: The toxicological mechanisms and the relationship between the test method endpoint(s) with the biological effect as well as the toxicity of interest should be addressed, describing limitations of the test methods

Toxicological Mechanisms
2.1 The toxicological mechanism of the method is the cytotoxicity of substances to corneal cells
2.2 The advantages of using CRV (crystal violet) staining to assess cytotoxicity are described in the protocol, but should also be mentioned in the report together with disadvantages. In addition, it would be helpful to describe the mechanism of CRV stain in the report.

Relationship with Toxicity of Interest
2.3 It would be helpful to discuss the advantage of using corneal cells over other types of cells.
2.4 It would be helpful to reference the ICCVAM document that describes why cytotoxicity can be useful as an indicator of ocular irritation (Hamernik et al., 2006)
2.3 It would be helpful to provide a justification of the 72-hour duration of exposure in relation to Draize in vivo exposure, and including a discussion of volatile substances.
2.4 It would be helpful to explain the advantages of using a relative control (rather than a fixed IC50 value), as recommended by the MHW Project. The PRP discussed the relevance of TEA with respect to its in vivo and in vitro data and considered it an adequate relative control. It would be helpful to reflect these considerations in the validation report.

Limitations of the test method
2.5 Substances insoluble as per the described procedure cannot be tested
2.6 It would be helpful to explain why color interference does not seem to be a limitation.
2.7 It would be helpful to describe mechanistic limitations of the test method (e.g. that the method is unable to assess reversibility, etc.)

Evaluation Criterion 3: A detailed test method protocol should be available

3.1 The different versions of the protocol used in the validation study are available.
3.2 Differences between the training protocol and the final protocol were shown using blue and green font color, but that font color cannot be seen in the provided print-outs.
3.3 It would be helpful to state more clearly in the protocol the procedure used to ensure that cells are used within a maximum period of time. E.g., preparing stock solutions all from the original seed cells and ensuring that these stock cells are used within three months after thawing (to ensure that cells are not refrozen and reused for an unlimited period of time).
3.4 It would be helpful to review protocol detail on reagents, e.g., the meaning of (-) in "PBS(-)," concentration and purity, and on temperature(s) during procedures.
3.5 The numbering of reference figures needs to be checked.
3.6 Section 4.4 (5) should state that absorbance is preferably measured at 588 nm, but if that is not possible, then at 570 nm, in order to harmonize this aspect with the validation report.)
3.7 It would be helpful to improve the language of the test protocol by checking for grammar and readability, e.g., in Section 4.6, the sentence starting on line 2, or Section 3.7.1, the sentence starting two lines from the end of the paragraph.
3.8 For Tables 3 and 4 on page 16 of Appendix 1, it would be helpful to make reference to the relative control (TEA).

Evaluation Criterion 4: The within- and between-laboratory reproducibility of the test method should be demonstrated

4.1 Information on reproducibility is provided in Sections 4.5.1 and 4.5.2. In terms of concordance of classifications, intra-laboratory reproducibility was 100% (20/20) at each of the three laboratories, and inter-laboratory reproducibility was 90% (27/30)
4.2 Both intra- and inter-laboratory reproducibility met the respective success criterion (Section 3.1.4).
4.3 The fact that the three participating laboratories were all naïve and that no practical training was provided is a good indication of the robustness of the test method. This should be indicated in the validation report.
4.4 However, it is important to take into consideration that the same manufacturing lots of TEA and bovine serum were used. The validation report should discuss the potential impact of bovine serum and TEA from different manufacturing lots on the reproducibility. (e.g., data from page 39 in Appendix 5.)
4.5 It would be helpful to refer to Appendix 7, Figs. 1 to 20, and to discuss those.
4.6 It would be helpful to include frequency of runs not meeting criteria in Section 3.2.8 Quality Control.
4.7 It would be interesting to compare reproducibility of results of chemicals tested in both the reevaluation study (Appendix 5) and the validation study.

Evaluation Criterion 5: Demonstration of the test method's performance should be based on testing of representative, preferably coded reference chemicals

5.1 The coding of the tested substances was conducted appropriately. In particular, the coding of each set of test chemicals during Phases II allowed demonstrating the high reproducibility of the assay.
5.2 It would be helpful to check the consistency of the total number of test chemicals (presumably 115) used for the calculation of predictive capacity and the final numbers should be reflected in Section 3.3.2.
5.3 The distribution of test chemicals according to GHS categories was appropriate and provided an adequate representation of liquids and solids.

5.4 It would be helpful to explain why, although there was evidence from the reevaluation study (Appendix 5) that excluding low molecular weight alcohols, ester, ethers, etc. resulted in better predictive capacity, molecular weight was not considered as a criteria for selection of test chemicals.
5.5 It would be helpful to use a single, harmonized chemical identifier for each test chemical throughout the validation study report, including tables.
5.6 It would be helpful to indicate how the chemical classes were defined.
5.7 It is noted that the chemical selection appears to have focused on mono-constituent substances, which allows good comparison with existing Draize test data.

<u>Evaluation Criterion 6</u>: **Predictive capacity should be demonstrated using representative chemicals. The performance of test methods should have been evaluated in relation to existing relevant toxicity data as well as information from the relevant target species.**

NB: Preliminary answers based on data available at the first PRP meeting (without raw in vivo and in vitro data) are provided.

6.1 The study plan defines the rabbit eye irritation test as the reference test for comparison with the in vitro results.
6.2 We consider the rabbit as the relevant target species, noting that prediction of human eye irritation is the ultimate goal, acknowledging that no human data are available. The PRP noted that in vivo data have been reviewed by ICCVAM. Additionally, the majority of test chemicals were also used in other validation studies (e.g. RhCE and STE).
6.3 Predictive capacity for the unrestricted applicability domain was based on a sufficiently large and representative set of test chemicals, however, results did not meet the VMT success criteria (accuracy of 55% vs. 80% of the success criteria, and false negative rate of 41% vs. <5% of the success criteria).
6.4 It would be helpful if the numbers for a bottom-up approach would be included in the text of Section 4.6.
6.5 Of the various physical and chemical properties explored, only molecular weight was found to be a clear reason for misclassification, which was also indicated by the outcome of the previous reevaluation study.
6.6 Exclusion of test chemicals with a molecular weight of less than 180 resulted in a false negative rate (4.8%) that met the success criterion. However, accuracy success criterion was not met (71.4% vs. 80%). In addition, sample size was reduced to 42 test chemicals (21 GHS negative and 21 GHS positive).
6.7 Although the accuracy success criterion was not met, it is noted that this is mainly due to the false positive rate, which errs on the side of safety.
6.8 It would be helpful to consider also the data from the reevaluation study together with the validation study data for assessment of predictive capacity.
6.9 It would be helpful to describe the representation of the considered physiochemical properties for the test chemicals in the reduced applicability domain.
6.10 It would be helpful to indicate in the text of the validation report the identity of the test chemical, which yielded a false negative result with the restricted applicability domain of molecular weight below 180.

Evaluation Criterion 7: All data should adequately support the assessment of the validity of the test method

7.1 The PRP considers that the test method is adequately documented in a standardized protocol, taking into consideration the minor suggestions made in Criterion 3 of this document.

7.2 Data and evidence on reliability seems adequate, but the PRP would require additional information on results using serum from different manufacturing lots before the data is considered fully adequate.

7.3 For the restricted applicability domain, additional data and evidence on relevance (e.g., use of existing data, scientific rationale) would be required before the data is considered fully adequate.

Evaluation Criterion 8: All data from the validation study supporting the validity of a test method should have been obtained in accordance with the principles of Good Laboratory Practice (GLP)

8.1 Based on the information provided to the PRP, the study appears to be conducted in accordance with the principles of GLP.

Evaluation Criterion 9: Applicability domain of the test method should be defined

9.1 Different potential reasons for misclassification were investigated, including chemical classes and physiochemical properties, but not in vivo drivers of classification.

9.2 Only molecular weight was found as a possible reason for misclassification, and it was proposed to exclude test chemicals with a molecular weight of less than 180 from the applicability domain.

9.3 It would be helpful to provide a scientific rationale explaining why test chemicals with a molecular weight of less than 180 and that are irritants are not predicted correctly. (It was noted that specificity was not affected by the restricted applicability domain.)

9.4 It was noted that surfactants were the only chemical class with acceptable accuracy and false-negative rates. However, the sample size (7) was small. It would be interesting to provide a scientific rationale for mechanism of toxicity for surfactants.

Evaluation Criterion 10: Proficiency chemicals should be set up in the proposed protocol

10.1 Upon completion of additional information requested in Criterion 7 of this document, a list of e.g. ten proficiency chemicals could be suggested (e.g., by the VMT) even before the test method is ready for regulatory use.

Evaluation Criterion 11: Performance standard should be set up with proposed protocol

11.1 Because the assay does not include components, equipment, or other scientific procedures that are covered (or pending) by intellectual property rights, performance standards are not mandatory at this stage, but could be useful in case that similar or modified test methods become available.

Evaluation Criterion 12: Advantages in terms of time, cost and animal welfare

12.1 The report indicates the ethical need of the test method for animal welfare reasons, but it would be helpful to indicate that this test method is a potential partial replacement for the Draize rabbit eye test. In addition, it would be helpful to specify the time, throughput, and cost advantages of the test method.

Evaluation Criterion 13: Completeness of all data and documents supporting the assessment of the validity of the test method.

13.1 The PRP has requested access to the raw in vivo and in vitro data. Final assessment of completeness can only be made after receiving the information requested in this document.

Evaluation Criterion 14: Validation Study Management and Conduct

14.1 The validation study was considered to be conducted in accordance with internationally accepted principles (OECD Guidance Document 34).
14.2 Training and transferability was considered to be sufficient. These aspects were addressed in a dedicated phase.
14.3 The PRP considers the study to be well designed, having a subset of substances for reproducibility assessment and a different subset for predictive capacity assessment. Sample sizes were considered to be sufficiently large for the two subsets.
14.4 It would be helpful to describe how the sample sizes were determined.
14.5 It would be helpful to describe the data workflow.
14.6 Describe how the statistical analyses were conducted, including information about replicates, runs, and sets.
14.7 It would be helpful to indicate that the participating labs did not take part in discussions about selection of test chemicals.
14.8 It would be helpful to indicate the physical location of laboratories.

Other considerations

- This document consists of preliminary answers based on data available at the first PRP meeting (without raw in vivo and in vitro data)
- It would be helpful to show that the solvents used have no effect on the negative control, e.g., by exploiting existing data.
- It was noted that the reference substance TEA has a molecular weight below 180.

Conclusion

- to be drafted once the VMT has replied to the comments and suggestions made by the PRP

IL-8 Luc assay meeting minutes

Thursday, March 05, 2015

	Introductions
Basketter	Tomorrow will be good time to determine review parameters. Also want to discuss how to handle comments from people not in attendance at that time.
Kojima	This is the kickoff meeting. We look forward to receiving a final report incorporating comments from the Peer Review Panel in approximately six months.
Basketter	Revise the protocol implies clarification without modification.
Kojima	Yes, I hope suggestions are minor.
Basketter	Six months sounds like a reasonable timeframe, but we would like to finalize that after seeing how these discussions go over the next few days.
	Presentation: Introduction of the IL-8 Luc Assay but Prof. Aiba
Basketter	The floor is now open to questions. Thank you for a very clear and useful presentation. Aspects of the protocol change as you gain experience, and you spoke of 122 substances, from which detergents were removed. Were all substances tested with the final protocol?
Aiba	Yes. We will focus on Phase II and III when presenting final conclusions
Eskes	On slide 26, Criteria 3 is a combination of 1 and 2.
Aiba	Not a just a combination. We don't use the same criteria.
Eskes	You said viability testing is not needed. Why?
Aiba	We look at IL-8 and GAPDH, which indicates viability.
Basketter	What happens with very cytotoxic substances?
Aiba	Basically, the value of SLR-LA decreases. There is a disruption of the cell membrane. Viability is often assessed this way. But SLR is suppressed much earlier than disruption of the membrane.
Basketter	Is that part of the problem with detergents? Do they disrupt the membrane?
Aiba	I don't think so. We are looking at why they disrupt the SLR activity.
Hoffmann	So these numbers are not indicative of cytotoxicity?
Basketter	Yes, that is my understanding.
Lim	I was surprised that DNCB isn't cytotoxic at that level.
Aiba	There is a time lapse after treatment with DNCB before the cells die. If we wait for 24 hours, yes, the cells die.
Basketter	The dose at which different cell types die will vary.
Hoffman	How many concentrations are called for in the standard protocol
Aiba	11
Hoffman	How do you judge solubility?
Aiba	Naked eye
Hoffman	Is there is specific regulation in Japan that this assay will be used for?
Kojima	Japanese regulators look at TG. We are starting discussions to accept

	combinations with in vitro testing.
Basketter	For REACH, there will probably be a lot of "last resort" animal testing performed. There is pressure to accept in vitro results, but this is not mandatory. With regard to skin sensitization, there is enthusiasm, but also a reluctance to accept negative data. So I expect Japanese regulators feel the same way.
Tsutsui	On Slide 43, I.I. is necessary for evaluation. The number of test substances tested with and without I.I. criterion differs. How does that affect the results?
Basketter	Would the results change if all the same chemicals were tested?
Aiba	We can redo that if required.
Takeyoshi	Why does concentration range change after first experiment?
Aiba	In the first experiment, we could roughly understand the concentrations. And change the highest concentration of chemicals. On Page 156, you can see that we understood from the first experiment a better point from which to dilute the test chemicals.
Tsutsui	4-Nitrobenzylbromide induced IL-8 activity, but not DNCB or Paraphenylenediamine. If you use a different housekeeping gene, would this phenomenon occur?
Aiba	We haven't looked at other housekeeping genes. If we chose a different one, the results might be different. We concentrated on selecting a good housekeeping gene for the results we wanted.
Basketter	I have had a similar experience. Your concern is that there is a fall in IL-8, but the housekeeping gene falls faster.
Aiba	This is why we changed from 6 to 16 hour incubation.
Basketter	Most increase occurs after 8 or 10 hours rather than 6. DNCB at 6 hours doesn't change but at 12 hours has changed a lot. Were there any other time course experiments?
Aiba	No.
Lim	THP1 is very unstable and need to reorder cells. How did you maintain cells.
Aiba	The cells we used were very stable and maintained their character for at least three months.
Eskes	What is the procedure for banking the cell line?
Aiba	We established the cell lines more than 3 or 4 years but did not have any problem. Keep alive in liquid nitrogen in various containers.
Lim	On slide 49, about X-VIVO. There is no mention in the protocol.
Aiba	This was a modification which is not in the protocol submitted to the OECD.
Eskes	What is X-VIVO?
Aiba	A chemically defined culture medium.
Lim	You used a 95% confidence interval. How did you achieve this?
Aiba	Four replicates
Hoffman	On the basis of normal distribution?
AIba	Yes
Takeyoshi	How do you check the cell responsiveness.
Aiba	If cobalt chloride does not satisfy the experiment positive control criteria, we didn't use the cells.
Basketter	Why cobalt chloride?
Aiba	It is a good stimulant and dissolves nicely.
Eskes	About the cell line, is there a patent on the cell line? D you provide cells to

	other laboratories?
Aiba	We are thinking of transferring the cell line to a company to maintain and distribute it. It is held now by our department at the university, so we are not so well able to manage it.
Ohmiya	We want to provide worldwide.
Tanaka	Certifying cell quality will be an important issue.
Ohmiya	We look at response to cobalt chloride for quality check now.
Takeyoshi	Will you use a banking system?
Ohmiya	Yes we will use AIST.
Tanaka	We use Japanese cell back JCRB. But we have to harmonize internationally.
Basketter	The OECD is considering how it might be able to accept patented cell lines.
Eskes	OECD is considering many things and a guidance document on the good in vitro methods, which storage of cell lines.
Aiba	In response to the question about the time course, we can see a clearer increase of activity over 16 hours when compared with shorter times.

Presentation of data analysis report

Basketter	I thought that Criterion 3 was clear cut, but in some places you say gray or undecided. What does that mean and why did you do that?
Aiba	We changed the protocol during the validation study. There is a difference in the criterion between Phases IIb and IIc or III. Also, there was a minor mistake in calculation. When data doesn't satisfy the final criterion, we show the result as gray.
Basketter	How should the PRP handle the gray or undecided data? Should we ignore it or something else? If I were a regulator, I would err on the side of safety and say positive. You have marked it gray because of small errors. Should we ignore it?
Aiba	Yes, I think you can ignore it.
Tsutsui	Why is judgment positive when there are two N and two P?
Aiba	When the first two results are positive, we judge the substance as a sensitizer. So we made decisions on how to assess discordant data.
Basketter	The developers made a reasonable decision about how to judge results. But why would you do a fourth run on a substance that gave the same result on the first three runs.
Aiba	When we could not judge by other criterion, we had to make four runs.
Tsutsui	If you have two negative results, you do a third experiment. But two positive results did not require a third experiment. Why.
Aiba	We want reduce false negatives, so we want to make sure by doing a third experiment.
Basketter	Yes, avoiding false negatives is a regulatory concern.
Hoffmann	I see no graphical representation of the data, just tables. I would like graphical representation.
Aiba	Like on page 197?
Hoffman	There are 11 concentrations, and for each, you measure value and confidence interval. And if one fulfills the criterion, it is positive?
Aiba	Yes. Prof. Omori created an Excel worksheet to automatically calculate

	results.
Eskes	Which protocol was used in Phase IIb? Is it different from the one we have?
Aiba	Yes, it was a different protocol. The major difference was the I.I. less than or equal to 0.8.
Basketter	The biological phase is the same?
Aiba	Yes.
Eskes	Should we assume detergents should be excluded from the Applicability Domain?
Aiba	Yes.
Basketter	Is it 5 hours or 6 hours vs. 16?
Aiba	Should be 6 hours.
Eskes	On pg 36 of the validation report, about study design, it says there will be evaluation of 1A, 1B, and non-sensitizers. But the results are mostly sensitizers and non-sensitizers. What is the main aim of the study?
Basketter	Page 268 has a relevance table that prompts a similar question.
Aiba	We did not subdivide into 1A or 1B. Perhaps it is a misprint.
Basketter	I thought that page 268 shows clearly defined GHS 1A, 1B, and No Category. So what you are showing is that the test can identify both 1A and 1B substances as sensitizers.
Eskes	So the study design is to distinguish sensitizers from non-sensitizers.
Basketter	The only way to evaluate the usefulness is by comparing results against in vivo data, but everyone always says their test must be used in a battery of tests.
Lim	The intra-laboratory concordance for Lab A is low and very different from the other two labs. Isn't this a problem?
Aiba	This related to Criterion 2.
Eskes	Can you tells us about the training? Were the laboratories naive?
Aiba	We had practical training and all the laboratories were naive.
Tanaka	We had trouble at first but after training, it was OK.
Aiba	This procedure requires dissolving and diluting chemicals, but after that there are no manual procedures.
Basketter	Throughout the entire testing, were there any substances other than detergents that had surprising results? Probably not.
Tsutsui	What explains the non-sensitizer result for beryllium sulphate?
Aiba	False negatives are often due to the solvent. If we use X-VIVO, we would get a positive result for beryllium sulphate.
Hoffmann	Why is X-VIVO better than bovine serum?
Aiba	Kao Group has reported similar results. Haptens might bind to the proteins in FBS.
Eskes	For how many passages is the cell line stable? Is there a maximum number of passages?
Aiba	We follow guidelines that call for use within 1.5 months.
Eskes	Do you refreeze and reuse cells?
Aiba	If we use a vial for more than 1.5 months, we culture a new vial.
Eskes	It is better to keep track of passage numbers.
Aiba	These lines are very stable so we did not define a maximum number of passages.

Basketter	Why did you specify 1.5 months.
Aiba	We checked to see how long we could keep a cell line. We ensured that they could keep the cell lines for 3 months, and 1.5 months was to ensure they would be active.
Tanaka	We need to ensure the stability of the cell line.
Eskes	How was the between- and within laboratory variation shown on page 82 calculated? Why is the between-laboratory reproducibility higher than the within-laboratory reproducibility?
Aiba	For benzyl cinnamate, there were 9 positives at three laboratories. For 2,4-dichloronitrobenzene, we relied in the dominant result.
Eskes	The statistical reports also shows something different.
Lim	You used 1.4 as a cut off. Is there a biological significance?
Aiba	We used 1.4 because it produces a better results, after that we used other criteria in with the 95% confidence interval. But unfortunately it did not produce better results.
Takeyoshi	Does this require specialized reagent. Is it commercially available?
Ohmiya	Yes. And same for equipment.
Hoffmann	Data template was designed by Dr. Omori and sent to the laboratories, where the data was input and then returned to Dr. Omori? Can we get an example of the Excel file?
Hoffmann	What is the general picture with IL-8? Does the dose response vary?
Aiba	Not so much.
Basketter	Other than detergents, were there any non-compatible test chemicals?
Aiba	No.

The Table below was prepared during the PRP meeting as a point of comparison for the IL8 Luc assay. It is provided solely for information and was not intended for inclusion in the main review document.

	DPRA		Keratinosens		h-CLAT	
	VS	Extra	VS	Extra	VS	Extra
Accuracy	79% (n=24)	80% (n=157)	90% (n=21)	77% (155/201)	76% (n=24)	84% (n=101)
Sensitivity	71%	80% (88/109)	87%	78% (71/91)	81%	88%
Specificity	92%	77% (37/48)	100%	76% (84/110)	66%	72%

Draft Peer Review Panel Evaluation of the IL-8 Luc Assay for skin sensitisation

Ver.1.0

Japanese Center for the Validation of Alternative Methods
National Institute of Health Sciences
Tokyo, Japan

11 March 2015

DRAFT RESPONSE OF THE IL8 Luc ASSAY PEER REVIEW PANEL

The peer review panel (PRP) met on Friday March 6th and the morning of Saturday March 7th, 2015.

The PRP agreed to complete the review by addressing 14 evaluation criteria. To this was added any other considerations that did not fit within these specific responses.

Evaluation Criterion 1: A rationale for the test method should be available, including a description of the human health effect, a clear statement of scientific need, and regulatory application.

- The validation report covers well the need for the assay in the current regulatory context.
- Although a clear rationale was given, it would be helpful to update the introduction to recognize that some in vitro methods for skin sensitization have been validated and adopted into OECD guidelines. This has led to a working group at OECD considering the integrated approaches to testing and assessment (IATA).
- In addition, a clearer perspective on which key event in the AOP is the target of this model is required. The Peer Review Panel (PRP) was of the opinion that the method presented an alternative endpoint for the assessment of dendritic cell (DC) activation (Key Event 3) compared to other assays, e.g., the human cell line activation test (hCLAT).
- The human adverse health effect arising from exposure to skin-sensitizing chemicals, allergic contact dermatitis, is inadequately described in the background. This should be briefly addressed.
- The PRP agreed that, when used on its own, the regulatory application of this assay could contribute to hazard identification (sensitizer/non-sensitizer). However, follow-up work suggests the assay could contribute also to Globally Harmonised System (GHS) sub-categorization when used in integrated approaches such as IATA.

Evaluation Criterion 2: The toxicological mechanisms and the relationship between the test method endpoint(s) with the biological effect as well as the toxicity of interest should be addressed, describing limitations of the test method.

- The PRP agreed that the mechanistic basis of the method and how it related to the skin-sensitization endpoint was well described in the report.
- It would be helpful to have a paragraph describing the test method limitations in greater detail, e.g., substances presenting problems related to solubility, cytotoxicity, and/or metabolism.
- Additional rationale, information, or references that demonstrate the stability and suitability of the selected housekeeping gene is desirable.