Original Article

Enhancing between-facility reproducibility of the SH test as an *in vitro* skin sensitization test by the improved test method

Noriyasu Imai^{1,4}, Midori Takeyoshi¹, Sakiko Aizawa¹, Mika Tsurumaki¹,
Masaharu Kurosawa¹, Akemi Toyoda², Maki Sugiyama², Kaoru Kasahara²,
Morihiro Hirota³ and Shinichi Ogata⁴

¹Safety and Analytical Research Laboratories, KOSÉ Corporation, 48-18 Sakae-cho, Kita-ku, Tokyo 114-0005, Japan

²Frontier Research Center, POLA Chemical Industries, Inc.,

560 Kashio-cho, Totsuka-ku, Yokohama, Kanagawa 244-0812, Japan

³Shiseido Global Innovation Center, 1-2-11

Takashima, Nishi-ku, Yokohama, Kanagawa 220-0011, Japan

⁴Graduate School of Environment and Information Sciences, Yokohama National University,
79-7 Tokiwadai, Hodogaya-ku, Yokohama, Kanagawa 240-8501, Japan

(Received October 16, 2020; Accepted February 15, 2021)

ABSTRACT — There has been an increased demand to eliminate animal experiments and to replace the experiments with alternative tests for assessing the safety of cosmetics. The SH test is an *in vitro* skin sensitization test that evaluates the protein binding abilities of a test substance. Skin sensitization must be evaluated by multiple test methods. The SH test uses the same cell line and measuring instruments as the human Cell-Line Activation Test (h-CLAT), which is one of the test methods used to evaluate different key events and is listed in the OECD test guidelines. There are cost advantages to usher the SH test into facilities that are already running the h-CLAT. The SH test is conducted only at a facility that has developed the SH test because studies on the between-facility reproducibility and validity have not been performed. Therefore, to verify the transferability of the SH test and the between-facilities reproducibility, we evaluated the reproducibility of the SH test results at three facilities, including the development facility. After an initial round of testing, the protocol was refined as follows to improve reproducibility among the three facilities: i) determine the optimum pH range, ii) change the maximum applicable concentration of water-soluble substances, and iii) define the appropriate dispersion conditions for evaluating hydrophobic substances. These refinements markedly enhanced the between-facility reproducibility (from 76.0% to 96.0%) for the 25 substances evaluated in this study. This study confirmed that the SH test is an effective skin sensitization test method with high technical transferability and between-facility reproducibility.

Key words: SH test, Skin sensitization, Alternative test method to animal experiments,
Technical transferability, Between-laboratory reproducibility

INTRODUCTION

There has been an increased demand to reduce or eliminate animal experimentation in various fields including toxicology. The European Union (EU) has banned the use of animal experimentation to assess the safety of cosmetics and cosmetic ingredients. EU has also prohibited the sale of products containing ingredients which were evalu-

ated by animal experimentation (Council of the European Union & European Parliament, 2003). Similar regulations have been established in other countries and regions; therefore, the global standards for evaluation of product safety are rapidly changing to exclude animal experimentation.

Under these circumstances, testing methods to replace animal experimentation have developed at a rapid pace.

At present, many of the testing methods assessing toxicity and not relying on animal experimentation have been adopted in the test guidelines (TG) of the Organization for Economic Co-operation and Development (OECD). Substituting *in vivo* tests with a single *in vitro* or *in chemico* alternative testing method, however, is challenging. Integrated testing strategies are required to evaluate the systemic toxicity and complex biologic reactions of various compounds (Adler *et al.*, 2011), especially for testing which requires repeated administration. The current mainstream approach is to use a framework that combines test methods for evaluating key events (KE) in the development of toxicity along the Adverse Outcome Pathway (AOP) in the context of the Integrated Approaches for Testing and Assessment (IATA) (OECD, 2016c).

The IATA, combining *in chemico* and *in vitro* testing methods that can evaluate KE based on the AOP, has been proposed to evaluate skin sensitization to chemicals (OECD, 2012a, 2012b). According to the OECD guidelines, there are four KE in the AOP for skin sensitization: KE1, covalent binding to skin proteins; KE2, epidermal keratinocyte inflammatory response; KE3, activation (maturation) and migration of Langerhans cell and dermal dendritic cells; and KE4, activation/proliferation of allergen specific T-cells. Of these, alternative test methods corresponding to KE1, KE2, and KE3 have been developed, and OECD has issued the test guidelines for each method.

Currently, two test methods are available to evaluate KE1: The Direct Peptide Reactivity Assay (DPRA) and the Amino acid Derivative Reactivity Assay (ADRA). These tests evaluate whether or not a test substance has the ability to bind to a peptide that is important for the expression of skin sensitization by assessing the reactivity of the test substance to a synthetic peptide (EURL-ECVAM, 2013; OECD, 2019b, 2019c).

Two other tests have also been developed to evaluate KE2. The KeratinoSens™ test and the LuSens test were recently adopted by the OECD. Both test methods evaluate the inflammatory response of keratinocytes by assessing the induction of gene groups controlled by the antioxidant reactive region (ARE) during the induction of skin sensitization. Cells containing a reporter gene are used to evaluate whether an ARE gene group has been induced on the basis of the activation of the Keap1-Nrf2-ARE pathway by a skin sensitizer. (OECD, 2018a; Natsch, 2010; EURL-ECVAM, 2014; Ramirez *et al.*, 2014; ESAC, 2016).

KE3 is currently evaluated using methods such as the human cell line activation test (h-CLAT), the U937 cell line activation test (U-SENS™), and the interleukin-8

reporter gene assay (IL-8 Luc assay). Dendritic cells activate during the induction phase of skin sensitization, migrate to lymph nodes, and present antigens to naïve T-cells. At this time, the expression of molecules that maintain adhesion to T cells such as CD54 and CD86 is enhanced on the surface of activated dendritic cells and the expression of inflammatory cytokines, such as IL-1 β , and chemokines, such as IL-8, are induced (Aiba *et al.*, 2003; dos Santos *et al.*, 2009). These three test methods evaluate the activation of dendritic cells using the expression of cell surface markers or IL-8 as indicators (OECD, 2018b).

The current OECD test guidelines do not include any test methods to evaluate the biological response corresponding to KE4, which is the activation/proliferation of allergen-specific T cells.

Although many test methods have been developed, the non-animal-based test methods which are currently available only evaluate a part of the AOP; therefore they cannot completely replace animal-based test methods. It is necessary to combine multiple test methods along with the AOP to replace animal-based tests and various combinations have been proposed (OECD, 2016a). Even if an appropriate combination of test methods is identified, each test often requires expensive equipment and the use of different cells and reagents. The tests also require intensive labor and familiarity with a number of test methods. Thus, large financial and human labor costs are involved in preparing a location with the capabilities of performing all of the tests for KE1, KE2, and KE3.

We focused on the SH test; an *in vitro* skin sensitization test developed by Suzuki *et al.* (2009). This test is performed using the human monocytic leukemia cell line (THP-1) and flow cytometry, similar to the h-CLAT (Ashikaga *et al.*, 2006; Sakaguchi *et al.*, 2006). The SH test has advantages because a facility that is already performing the h-CLAT can perform the SH test by only purchasing the necessary detection reagents and learning a few simple procedures. The SH test is not listed in the OECD TG, but is considered a reliable testing method to evaluate KE1 in the IATA for skin sensitization (OECD, 2016b). Skin sensitizing KE1 is based on the binding of chemicals to proteins. After entering the body, a low-molecular-weight sensitizer (hapten) binds to a lysine or cysteine residue of the carrier protein to acquire immunogenicity recognizable by dendritic cells. The DPRA and the ADRA assess the protein binding ability of the test substance using an indicator that depends on the reactivity of the test substance with a synthetic peptide containing either lysine or cysteine (Gerberick *et al.*, 2004, 2007; Fujita *et al.*, 2014). The SH test is a method of detecting

skin sensitizers on the basis of a decrease in the amount of cell surface thiols due to the binding of a test substance to the cell surface proteins (Hirota *et al.*, 2009, 2013). In addition, the SH test can also detect certain skin sensitizers that increase the number of cell surface thiols via endoplasmic reticulum stress in THP-1 cells and activation of the mitogen-activated protein kinase cascade (Hirota *et al.*, 2010). Furthermore, the use of the SH test for assessing KE1 in combination with an ARE assay for assessing KE2 (Natsch and Emter, 2008) and the h-CLAT for assessing KE3 produced good results in a risk assessment model of skin sensitization (Hirota *et al.*, 2015). In their report, the SH test performance was equivalent to that of the DPRA test as an *in vitro* test method for KE1 (Hirota *et al.*, 2015).

However, the SH test has not yet been validated. The SH test technical transferability to facilities other than the developing facility or its reproducibility between facilities have not been evaluated, either. Therefore, in this study, we verified the ability to establish the SH test at other facilities and evaluated the between-facility reproducibility.

First, two new facilities that adopted the SH test received technical guidance from the developer. Next, tests were conducted on common test substances at the three facilities and the reproducibility of the test results among the three facilities was evaluated. After the initial round of testing, it was revealed that the protocol needed improvement to enhance the reproducibility among the facilities. Therefore, the protocol was adjusted as follows; i) determine the optimum pH range, ii) change the maximum applicable concentration of water-soluble substances from 5000 µg/mL to 15000 µg/mL, and iii) define the appropriate dispersion conditions for evaluating poorly water-soluble substances. As a result of the adjustments, the between-facility reproducibility was markedly enhanced. The SH test became an *in vitro* skin sensitization test method with high technical transferability as well as between-facility reproducibility.

MATERIALS AND METHODS

Testing facilities

The testing was conducted at three facilities: the Shiseido Global Innovation Center (Facility A), the Frontier Research Center of POLA Chemical Industries, Inc. (Facility B), and the Safety and Analytical Research Laboratories of KOSÉ Corporation (Facility C). The SH test was developed at Facility A. In this report, we used old as well as new data reported by Facility A.

Test chemicals

A total of 25 chemicals were evaluated in this study, including 17 sensitizers and 8 non-sensitizers (Table 1). All of the chemicals were used in the studies of the Working Group for *In vitro* Skin Sensitization Evaluation of the Japan Cosmetic Industry Association (Hirota *et al.*, 2015). The chemicals were previously evaluated and classified with a local lymph node assay (LLNA) (Urbisch *et al.*, 2015). When selecting the test chemicals, we focused on the following: i) select a wide range of compounds in points of skin sensitization potency and *n*-octanol / water partition coefficient (log *K*_{ow}) predicted by KOWWIN™ contained in the Estimation Programs Interface (EPI) Suite provided by the United States Environmental Protection Agency (available at <https://www.epa.gov/tsca-screening-tools/epi-suite-tm-estimation-program-interface#models>) and ii) increase the proportion of chemical substances that have similar characteristics as cosmetic ingredients (non or low skin sensitization potency and high log *K*_{ow}). We purchased the test chemicals from MilliporeSigma Corporation (St. Louis, MO, USA) or FUJIFILM Wako Pure Chemical Corporation (Osaka, Japan). The details of the test chemicals are provided in Table 1. The LLNA is known to sometimes obtain false positive results with irritants such as surfactants (OECD, 2010). One example is Sodium lauryl sulfate (SLS). During a peer review of the LLNA, it was confirmed that the positive reaction for SLS in the LLNA was not a skin sensitization but a false positive due to skin irritation (ICCVAM, 1999). In this study, SLS was treated as an LLNA negative substance.

Cells and cell culture

The human monocytic cell line, THP-1, was obtained from the ATCC (Manassas, VA, USA). These cells were cultured in RPMI 1640 medium with HEPES buffer (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) or without HEPES buffer (Nissui Pharmaceutical Co., LTD, Tokyo, Japan), containing 1% antibiotic–antimycotic (Invitrogen, Thermo Fisher Scientific) and 10% fetal bovine serum (JRH Biosciences, Lenexa, KS, USA, or MilliporeSigma). Facility A and B used RPMI 1640 medium with HEPES buffer while Facility C used RPMI 1640 medium without HEPES buffer. The cells were maintained in an incubator at 37°C under an atmosphere of 5% CO₂ in air.

SH test procedure

The SH test was performed in accordance with the method of Hirota *et al.* (2013). THP-1 cells, which were maintained by subculture, were cultured at 37°C for 2 hr

Table 1. Summary of information on evaluated chemicals and results of the SH test by the three facilities.

No	Chemicals name	CAS RN	Sensitization Potency by LLNA	LogKow (KOWWIN TM)	LLNA EC3* (%)	Results of the individual facilities						Supplier
						Facility A		Facility B		Facility C		
						MAC value	Judgement	MAC value	Judgement	MAC value	Judgement	
1	Diphenylcyclopropanone	886-38-4	Extreme	3.25	0.003	463**	Positive	346.4	Positive	58	Positive	MilliporeSigma
2	2,4-Dinitrochlorobenzene	97-00-7	Extreme	2.27	0.05	79.9**	Positive	63.3	Positive	51.3	Positive	MilliporeSigma
3	1,4-Phenylenediamine	106-50-3	Strong	-0.39	0.16	67.7**	Positive	29.2	Positive	66.8	Positive	MilliporeSigma
4	Hexyl salicylate	6259-76-3	Strong	5.06	0.18	19.6	Positive	26.9	Positive	39.6	Positive	MilliporeSigma
5	2-Mercaptobenzothiazole	149-30-4	Moderate	2.86	1.7	74.1**	Positive	60.1	Positive	58.1	Positive	MilliporeSigma
6	Nickel sulfate	10101-97-0	Moderate	-	4.8	51.6**	Positive	37.2	Positive	50.5	Positive	MilliporeSigma
7	4-Chloroaniline	106-47-8	Moderate	1.72	6.5	48.3**	Positive	221.1	Positive	161.5	Positive	MilliporeSigma
8	1,2,4-Benzenetricarboxylic anhydride	552-30-7	Moderate	1.95	9.2	13.2**	Negative	10.6	Negative	8.7	Negative	MilliporeSigma
9	1-Bromohexane	111-25-1	Weak	3.63	10	20.3**	Positive	20.3	Negative	28.9	Negative	MilliporeSigma
10	Hexyl cinnamic aldehyde	101-86-0	Weak	4.82	11	29.0**	Positive	20.0	Positive	40.6	Positive	MilliporeSigma
11	Eugenol	97-53-0	Weak	2.73	13	16.2**	Positive	25.9	Positive	74.5	Positive	MilliporeSigma
12	Abietic acid	514-10-3	Weak	6.46	15	46.4	Positive	15.1	Positive	30	Positive	MilliporeSigma
13	Benzyl cinnamate	103-41-3	Weak	4.06	18.4	34.2	Positive	15.7	Positive	28.5	Positive	FUJIFILM Wako
14	N,N-Dibutylaniline	613-29-6	Weak	5.12	19.6	19.0	Positive	32.6	Positive	18.8	Positive	FUJIFILM Wako
15	Geraniol	106-24-1	Weak	3.47	26	25.5**	Positive	15.5	Positive	25.4	Positive	MilliporeSigma
16	d,l-Citronellol	106-22-9	Weak	3.56	43.5	33.3**	Positive	17.0	Positive	26.8	Positive	MilliporeSigma
17	Pyridine	110-86-1	Weak	0.65	72	26.0**	Positive	5.5	Negative	6.6	Negative	MilliporeSigma
18	Sodium lauryl sulfate	151-21-3	Weak (False Positive)	1.69	14	6.6**	Negative	4.9	Negative	6.7	Negative	MilliporeSigma
19	1-Iodohexane	638-45-9	Negative	4.05	-	15.2	Negative	18.7	Negative	35.3	Positive	FUJIFILM Wako
20	2-Hydroxypropyl methacrylate	923-26-2	Negative	0.72	-	22.7**	Positive	15.1	Positive	17.7	Positive	MilliporeSigma
21	4-Hydroxybenzoic acid	99-96-7	Negative	1.4	-	3.1**	Negative	3.3	Negative	7.9	Negative	MilliporeSigma
22	Chlorobenzene	108-90-7	Negative	2.6	-	16.0**	Positive	5.9	Negative	99.7	Positive	MilliporeSigma
23	Lactic acid	50-21-5	Negative	-0.65	-	4.4**	Negative	26.6	Negative	20.1	Positive	MilliporeSigma
24	Methyl salicylate	119-36-8	Negative	2.6	-	15.0**	Negative	43.0	Positive	19	Positive	FUJIFILM Wako
25	Propylene glycol	57-55-6	Negative	-0.78	-	4.8**	Negative	2.8	Negative	5.3	Negative	FUJIFILM Wako

Gray shading indicates a positive result. Log Kow: Log octanol-water partition coefficient. Log Kow was predicted by KOWWIN™ contained in the Estimation Programs Interface (EPI) Suite provided by United States Environmental Protection Agency. CAS RN: Chemical Abstract Service Registry Number. LLNA: Local Lymph Node Assay. EC3: Estimated concentration needed to produce a stimulation index of 3.

*, ** The LLNA data and Shiseido's data in this table were taken with permission from Hirota *et al.* (2015)

Improving the between-facility reproducibility and usefulness of the SH test

in a culture medium containing one of the test chemicals. Three concentrations of each test chemical were evaluated: one-third and one-ninth of the highest applied concentration, with the highest concentration set to the concentration that produced 50% cell survival (IC_{50}). Therefore, the cytotoxicity of each test chemical was evaluated in advance, and the IC_{50} concentration was determined from a dose-survival curve. Cell viability was measured using an MTT assay. After being treated with the test chemical, the cells were washed with phosphate-buffered saline (PBS). Next, Alexa Fluor 488 C5 maleimide (Invitrogen, Thermo Fisher Scientific), a fluorescent dye that selectively binds to the thiols of proteins, was dissolved in PBS and added to the cells and the cells were further cultured for 30 min at 37°C. The cells were then washed again with PBS and analyzed by flow cytometry (BD FACSCanto™ II or BD FACSCalibur™, Becton, Dickinson and Company, Franklin Lakes, NJ, USA). At the time of the analysis, dead cells were identified using 0.625 µg/mL propidium iodide. After the mean fluorescence intensity (MFI) at each test concentration was determined, the relative fluorescence intensity (RFI) was calculated by the following formula: $RFI (\% \text{ of control}) = (MFI \text{ of chemical-treated cells} / MFI \text{ of vehicle-treated control cells}) \times 100$. The RFI was not calculated when cell viability was less than 50%. The SH test was performed independently at least three times, and the mean RFI value was calculated from the RFI value for each test. For each test concentration, a positive result required that the mean RFI value was $\leq 85\%$ or $\geq 115\%$, and the difference from the control group was statistically significant as evaluated by a paired 2-tailed *t*-test with a *p*-value less 0.05. The SH test was conducted at three concentrations and was judged for each concentration based on the mean RFI value. The maximum amount of change of the mean RFI values in the SH test (MAC value) was calculated for each test substance and this is shown in the result as a guide of the amount of fluctuation in the RFI value of each substance. The MAC value is the maximum value of the fluctuation amount in the mean RFI value calculated for each test concentration by the following formula: The fluctuation amount in mean RFI = $|100 - (\text{mean RFI value})|$.

RESULTS

Confirmation of transferability and reproducibility among three facilities

Confirmation of transferability

After Facility A provided the hands-on training for conducting the SH test to Facility B and C, we evaluat-

ed potential differences in the detection sensitivity among the three facilities by having each facility independently perform the SH test using dinitrochlorobenzene (DNCB), which is used as a positive control substance for the SH test. Similar results were obtained for the dose response of DNCB among the three facilities (Fig. 1). The lowest dose of DNCB (3 µg/mL) was negative in all three facilities, and doses ≥ 4 µg/mL of DNCB were clearly positive in all three facilities. Thus, the detection limit for DNCB was 4 µg/mL in all three facilities. These results indicate that there were no major issues in the technical transferability of the SH test.

Confirmation of reproducibility among facilities

Next, we verified the reproducibility among the facilities using the 25 test substances (Table 1 and 2). Two of the weak sensitizers in the LLNA, 1-bromohexane and pyridine, were detected as positive at Facility A, but negative at Facility B and C, which means that they were not detected as sensitizers by the new facilities. Methyl salicylate, a non-sensitizer, was negative at Facility A, but positive at Facility B and C. Another non-sensitizing substance, lactic acid, was negative at Facility A and B, and positive at Facility C. In addition, chlorobenzene, another non-sensitizing substance, was positive at Facility A and C but negative at Facility B. 1,2,4-benzenetricarbox-

Fig. 1. Dose response curve of RFI value with DNCB treatment in the three facilities. The RFI value is the mean of three independent test results $\pm$ S.D. ($n = 3$). The test facilities are (Δ : Facility A, $\bullet$: Facility B, $\blacksquare$: Facility C). The dotted line shows the 85% criterion for a positive result. *: The RFI values of all three facilities were 85% or less and significant ($p < 0.05$) by *t*-test, indicating that a positive test result was obtained.

Table 2. Consistency of test results for the three facilities.

			Facility B		Facility C	
			Positive	Negative	Positive	Negative
Facility A	Positive	18	15	3	16	2
	Negative	7	1	6	3	4
Accuracy (%)			84.0		80.0	
Sensitivity (%)			83.3		88.9	
Specificity (%)			85.7		57.1	

The consistency rate between all three facilities: 76.0%

Gray shading indicates that the same judgments were obtained in Facility A and in Facility B or C.

ylidic anhydride, which was previously reported as a moderate skin sensitizer, was not detected as a sensitizer in any of the facilities, but the same results were obtained in all facilities. Even though the results were different from the LLNA data, the reproducibility between facilities was confirmed. In comparison with the results reported by Facility A, the test method developer, the concordance rates (*i.e.*, the accuracy) of the results of Facility B and C were 84.0% and 80.0%, respectively, and the concordance rate of the results among all three facilities was 76.0%. This value is below 80% and meets the criteria of Between Laboratory Reproducibility (BLR) provided in the performance standards of some OECD Guidelines (OECD, 2015a, 2015b, 2019a).

For 5 of the 6 substances that did not produce reproducible results among the facilities (1-bromohexane, pyridine, 1-iodohexane, lactic acid, and methyl salicylate), the results of Facility A were in agreement with the results of the LLNA, whereas the results of Facility C did not agree with the LLNA results. For the remaining substance (chlorobenzene), the SH test results at Facility A and C differed from the LLNA results, but the results at Facility B were consistent with the LLNA results. Based on Table 1, Facility A differed from the LLNA for three substances (1,2,4-benzenetricarboxylic anhydride, 2-hydroxypropyl methacrylate, and chlorobenzene). Facility B differed from the LLNA for five substances (1,2,4-benzenetricarboxylic anhydride, 1-bromohexane, pyridine, 2-hydroxypropyl methacrylate, and methyl salicylate). Facility C differed from the LLNA for eight substances (1,2,4-benzenetricarboxylic anhydride, 1-bromohexane, pyridine, 1-iodohexane, 2-hydroxypropyl methacrylate, chlorobenzene, lactic acid, and methyl salicylate). Regarding the predictive performance for LLNA results at each test facility (Table 3), the probability of correctly predicting a positive substance as positive (*i.e.*, the sensitivity) was 94.1% for Facility A and 82.4% for both Facility B and C. The probability of predicting a negative substance as neg-

ative (*i.e.*, the specificity) was 75.0% for Facility A and B and remarkably lower (37.5%) for Facility C. The accuracies were 88.0%, 80.0%, and 68.0% for Facility A, B, and C, respectively. The between-facility reproducibility was only slightly below the OECD BLR criterion of 80%, but there was a large difference in the prediction performance of each facility. Therefore, to improve the test accuracy and further improve reproducibility to meet the BLR criteria, the causes of the differences in reproducibility and prediction performance as well as potential improvements of the original standard operating procedure (SOP) were studied as described below.

SH test protocol refinements

Optimization of the pH range during cell treatment

For lactic acid, a non-skin sensitizer, Facility A and B produced negative results while Facility C produced a positive result. In addition, the MAC value was small (4.4) in Facility A, but large in both Facility B (26.4), which had a negative test result, and Facility C (20.1), which had a positive test result. Although the RFI value in Facility B for lactic acid changed in relation to the control treatment, the test was considered to be negative because the change was not significantly different from the control treatment. Because the two facilities with negative lactic acid results (Facility A and B) used HEPES buffer in the cell medium while the facility with positive results (Facility C) used medium without HEPES buffer, we hypothesized that differences in the pH of the medium affected the SH test results. Therefore, we evaluated the effect of the pH of the treatment medium on the SH test (Fig. 2). Lactic acid (3500 µg/mL, the IC₅₀ value) was added to the medium (pH 3.8) and then NaOH was added to prepare a treatment medium with a pH ranging from 4.0 to 10.0 in 1-pH value increments. The RFI value clearly decreased in the unadjusted and the pH 4.0 medium compared to the higher pH adjusted medium (Fig. 2). Although the SH test result could be correctly evaluated as negative in medium

Table 3. Prediction performance against LLNA results from each facilities.

			Facility A		Facility B		Facility C	
			Positive	Negative	Positive	Negative	Positive	Negative
LLNA	Positive	17	16	1	14	3	14	3
	Negative	8	2	6	2	6	5	3
Accuracy (%)			88.0		80.0		68.0	
Sensitivity (%)			94.1		82.4		82.4	
Specificity (%)			75.0		75.0		37.5	

Gray shading indicates a prediction consistent with the LLNA results.

with a pH ranging from 5.0 to 10.0, there was a tendency for the tests of the pH 10 medium to become positive due to the mean RFI value. Thus, we determined that it was necessary to maintain the pH of the treatment medium near neutral to obtain good predictability and reproducibility of the SH test. Adjusting the pH changed the lactic acid results at Facility C from a false positive to negative. Furthermore, the amount of fluctuation in the RFI values was reduced and stable negative results were obtained (refer to lactic acid's data in Table 4).

Maximum applicable concentration

We then investigated why pyridine, which is a weak sensitizer, produced false negative results at Facility B and C. Both of the new facilities had low MAC values and little fluctuation of the RFI values. This suggests that the exposure level to the test substance was insufficient. Therefore, we examined the maximum applicable concentration of the test substance. In the conventional SOP, the maximum applicable concentration of water-soluble substances for the SH test is 5000 µg/mL. By applying 15000 µg/mL, we succeeded in detecting pyridine as positive (Fig. 3-1). When 10000 µg/mL was applied, the mean RFI value exceeded 115% of the cut-off line, but the change was not statistically significant by *t*-test and was therefore not determined as a positive result.

On the other hand, Fig. 3-2 shows the results of the evaluation of 4 non-sensitizing substances (2-propanol, glycerol, ethanol, and propylene glycol) that are commonly used as solvents. Increasing the maximum applicable concentration of each of these substances to 15000 µg/mL did not produce a false positive. Therefore, changing the maximum applicable concentration of the water-soluble test substances for the SH test to 15000 µg/mL improved the detection sensitivity without increasing the false positive rate.

As a result of changing the maximum applicable concentration of pyridine to 15000 µg/mL, the result of Facility B turned from false negative to positive, but Facility C did not (refer to pyridine's data in Table 4). Compared to

the conventional SOP, the MAC values for both Facility B and C were clearly increased, suggesting that the cells are reacting to the sensitizer. However, in Facility C the amount of change was not statistically significant, and the result was not determined as positive.

Of the 25 substances examined in this study, the water-soluble substances that required re-testing after changing the maximum concentration were pyridine and propylene glycol. For the other water-soluble substances, it was not necessary to change the maximum concentration due to cytotoxicity and solubility.

Dispersibility of poorly water-soluble substances

Some of the substances with low between-facility reproducibility had high log Kow and low water solubility. The predicted log Kow, determined using KOW-WIN™ (<https://www.epa.gov/tsca-screening-tools/>)

Fig. 2. Effect of pH on RFI value with lactic acid treatment in the SH test. RFI values are mean $\pm$ S.D. ($n = 3$) of three independent tests. The dotted line shows the cut-off line for a positive result with an RFI value of $\leq 85\%$. Lactic acid, which is a non-sensitizing substance, was added at 3500 µg/mL, and pH of the medium was adjusted to each pH value with NaOH.

Fig. 3-1. Dose response when evaluating pyridine in the SH test. The maximum applicable concentration of the original standard operating procedure of the SH test is 5000 $\mu\text{g/mL}$. * indicates that the mean RFI value was in the positive region of the SH test and was significant ($p < 0.05$) by *t*-test. The gray area shows the negative region of the mean RFI value in the SH test ($85 < \text{RFI} < 115$). The mean RFI values (bars) and cell viability (lines) indicate the mean $\pm$ S.D. of three independent test results.

Fig. 3-2. The SH test results for each solvent when changing the maximum applicable concentration. The maximum applicable concentration of the original standard operating procedure of the SH test is 5000 $\mu\text{g/mL}$. * indicates that the mean RFI value was in the positive region of the SH test and was significant ($p < 0.05$) by *t*-test. The gray area shows the negative region of the mean RFI value in the SH test ($85 < \text{RFI} < 115$). Each bar represents the mean RFI value $\pm$ S.D.. In the test, three tests were independently repeated 7 times. □: 2-propanol, ■: ethanol, ▨: glycerol, ▤: propylene glycol.

epi-suitetm-estimation-program-interface), was 3.63 for 1-bromohexane, 4.05 for 1-iodohexane, and 2.6 for both chlorobenzene and methyl salicylate. The SH test incubation time is 2 hr shorter than general testing methods using cells. The evaluation can be performed within the time when the dispersed state can be maintained even if a

substance with low water solubility is not uniformly dissolved. As shown in Table 1, substances, such as hexyl salicylate, hexyl cinnamic aldehyde, abietic acid, benzyl cinnamate, and *N,N*-dibutylaniline, with a predicted log *K*_{ow} of 4 to 6 were able to be evaluated. The difference in the dispersion state was considered to increase the

Improving the between-facility reproducibility and usefulness of the SH test

Table 4. Results of the SH test with improved methods at the tree facilities.

No	Chemicals name	Sensitization Potency by LLNA	LLNA EC3* (%)	Results of each facilities						Remarks
				Facility A		Facility B		Facility C		
				MAC value	Judgement	MAC value	Judgement	MAC value	Judgement	
1	Diphenylcyclopropenone	Extreme	0.003	448.9	Positive	346.4	Positive	59.2	Positive	
2	2,4-Dinitrochlorobenzene	Extreme	0.05	82.0	Positive	63.3	Positive	51.3	Positive	
3	1,4-Phenylenediamine	Strong	0.16	67.7**	Positive	29.2	Positive	66.8	Positive	
4	Hexyl salicylate	Strong	0.18	16.8	Positive	26.9	Positive	39.6	Positive	
5	2-Mercaptobenzothiazole	Moderate	1.7	74.1**	Positive	60.1	Positive	46.8	Positive	
6	Nickel sulfate	Moderate	4.8	51.6**	Positive	37.2	Positive	48.4	Positive	
7	4-Chloroaniline	Moderate	6.5	48.3**	Positive	221.1	Positive	17	Positive	
8	1,2,4-Benzenetricarboxylic anhydride	Moderate	9.2	4.8	Negative	10.6	Negative	8.7	Negative	
9	1-Bromohexane	Weak	10	20.3**	Positive	19.3	Positive	17.9	Positive	3
10	Hexyl cinnamic aldehyde	Weak	11	29.0**	Positive	20.0	Positive	40.6	Positive	
11	Eugenol	Weak	13	19.0	Positive	25.9	Positive	54.7	Positive	
12	Abietic acid	Weak	15	46.4	Positive	15.1	Positive	16.3	Positive	
13	Benzyl cinnamate	Weak	18.4	34.2	Positive	15.7	Positive	28.5	Positive	
14	N,N-Dibutylaniline	Weak	19.6	19.0	Positive	32.6	Positive	16.1	Positive	
15	Geraniol	Weak	26	25.5**	Positive	15.5	Positive	25.4	Positive	
16	d,l-Citronellol	Weak	43.5	35.1	Positive	17.0	Positive	19.3	Positive	
17	Pyridine	Weak	72	55.3	Positive	36.6	Positive	21.4	Negative	2
18	Sodium lauryl sulfate	Weak (False Positive)	14	6.8	Negative	4.9	Negative	6.7	Negative	
19	1-Iodohehexane	Negative	-	15.2	Negative	18.7	Negative	39.5	Negative	3
20	2-Hydroxypropyl methacrylate	Negative	-	22.7**	Positive	15.1	Positive	17.7	Positive	
21	4-Hydroxybenzoic acid	Negative	-	4.8	Negative	3.3	Negative	2	Negative	
22	Chlorobenzene	Negative	-	16.0**	Positive	69.0	Positive	99.7	Positive	3
23	Lactic acid	Negative	-	5.8	Negative	12.4	Negative	9.4	Negative	1
24	Methyl salicylate	Negative	-	24.1	Positive	43.0	Positive	19	Positive	3
25	Propylene glycol	Negative	-	3.5	Negative	2.8	Negative	8.2	Negative	

Gray shading indicates a positive result. LLNA: Local Lymph Node Assay. EC3: Estimated concentration needed to produce a stimulation index of 3. The factors that enhanced between-facility reproducibility are explained in the remarks column as follows. 1: the optimization of pH, 2: the increase in the maximum applicable concentration, and 3: maintaining a homogenized dispersion state.

*, ** The LLNA data and Shiseido's data in this table were taken with permission from Hirota *et al.* (2015)

variability of the test results. Therefore, in the SH tests for substances with poor water solubility, the dispersion criterion of test substance in the medium was set to create a dispersion state in which the test substance would not quickly coalesce during the 2 hr incubation period. In Facility B and C where the false-negative result had been obtained to 1-bromohexane, the improved dispersion state reduced variability of the result and made a statistically significant difference compared to the control group, and the result turned to the positive. On the other hand, for non-sensitizing substances, there were some cases where the result changed to a false positive due to the improvement of dispersibility. In Facility A, the methyl salicylate result was negative and the MAC value was 15 on the borderline. This result was changed to a false positive because the MAC value became larger and the variability

ty was reduced due to the improved dispersion state and associated improved exposure condition in this retest. The results of chlorobenzene at Facility B were also affected by changes in exposure conditions associated with improved dispersibility. Initially, although the MAC value was 5.9, which was almost unchanged, the improved dispersion state increased the MAC value to 69.0 at a lower concentration than before. In Facility C, 1-iodohexane results changed from false positives to negatives because variability in results could not be completely ruled out and no significant difference was obtained (Table 4).

Effect of improved test conditions on between-facility reproducibility

After adjusting the pH, increasing the maximum applicable concentration of water-soluble substances, and

Table 5. Consistency of test results with the improved protocol at the three facilities.

			Facility B		Facility C	
			Positive	Negative	Positive	Negative
Facility A	Positive	19	19	0	18	1
	Negative	6	0	6	0	6
Accuracy (%)			100.0		96.0	
Sensitivity (%)			100.0		94.7	
Specificity (%)			100.0		100.0	
The consistency rate between all three laboratories : 96.0%						

Gray shading indicates that the same judgments were obtained in Facility A and in Facility B or C.

Table 6. Prediction performance with improved protocol against LLNA for each facilities.

			Facility A		Facility B		Facility C	
			Positive	Negative	Positive	Negative	Positive	Negative
LLNA	Positive	17	16	1	16	1	15	2
	Negative	8	3	5	3	5	3	5
Accuracy (%)			84.0		84.0		80.0	
Sensitivity (%)			94.1		94.1		88.2	
Specificity (%)			62.5		62.5		62.5	

Gray shading indicates a prediction consistent with the LLNA results.

improving the dispersion of the poorly water-soluble substances, we again tested the various test substances in the three facilities (Table 4).

Adjusting the pH of lactic acid changed the positive result in Facility C to a negative result, matching those results of facility A and B. Increasing the maximum concentration of pyridine changed the test result from negative to positive in Facility B, similar to Facility A. By increasing the dispersion, 1-bromohexane, chlorobenzene, and methyl salicylate were positive at all facilities, and 1-iodohexane was negative at all facilities. Comparisons of the SH test results between Facility A and Facility B and C are summarized in Table 5. Facility B gave the same results as Facility A for all substances, and Facility C gave the same results as Facility A, except for pyridine. Reproducibility at the three facilities was 96.0%, and the reproducibility between the facilities was greatly increased by improving the test conditions.

The predictability to the LLNA results are summarized in Table 6. The sensitivity was 94.1%, 94.1%, and 88.2% for Facility A, B, and C, respectively. The specificity was 62.5% at all facilities, and the accuracy was 84.0% for Facility A and B, and 80.0% for Facility C. Compared with results obtained before improving the test conditions, the accuracy of Facility A and the specificity of Facility A and B slightly decreased, but the accuracy and sensitivity of Facility B and C improved, and the specificity of Facility C was remarkably improved.

DISCUSSION

To avoid animal testing, a number of alternative skin sensitization test methods have been developed. The SH test examined in this study is a method that evaluates the protein binding of sensitizers, which is an important KE for establishing skin sensitization (Hirota *et al.*, 2010; OECD, 2016a). Therefore, we examined the BLR of the SH test, which has not yet been evaluated. The developer, Facility A, provided hands-on training to Facility B and C, and we first compared the detection sensitivity to the positive control substance DNCB among the three facilities. In all three facilities, the limit of detection of DNCB was 4 µg/mL, and the dose-effect curves were similar among the facilities. These results confirmed that the newly introduced technologies had sufficient detection sensitivity in the two new facilities and that it was possible to easily transfer the technique between laboratories.

In the SH test development at Facility A, the positive control was set at 4 µg/mL DNCB as one of the acceptance criteria. It should be noted that the positive control is close to the detection limit, but this was considered to be an appropriate index for confirming the sensitivity of the test method. In addition, confirming the detection sensitivity for DNCB, including the dose-responsiveness, can guide new facilities in the technical acquisition of the SH test. Although the principle of the test method is different between the SH test and the h-CLAT, they both use the

same cells and DNCB as a positive control. As described in OECD TG442E (OECD, 2018b), one of the criteria for the technical mastering of the h-CLAT is that the range of the lower limit concentration that can detect DNCB as positive is 0.5 to 10 µg/mL. In other words, if DNCB of 10 µg/mL or less can be detected, it is determined that the h-CLAT has reached a certain technical level. Although the SH test and the h-CLAT use the same cells, the treatment conditions and principles of the tests differ, making direct comparison difficult. However, the DNCB detection sensitivity of the SH test (its detection limit was 4 µg/mL) was not so different from the h-CLAT. Therefore, it was considered that the detection sensitivity of the SH test was sufficient for a skin sensitization test.

To confirm the reproducibility between facilities, the SH test was performed on 25 chemical substances and the results were compared with the data previously reported by Facility A. The concordance rate of the results of all three facilities was 76.0%, which is close to the 80% criteria for BLR in the *in vitro* test method performance standard of the OECD test guidelines (OECD, 2015a, 2015b, 2019a). Relative to the LLNA data, however, the sensitivity of Facility A was very high (94.1%), while that of Facility B and C was only 82.4%. In addition, there were many false positives in Facility C, with an accuracy of 68.0% and a specificity of less than 40%. The accuracy and specificity are significantly lower than those of Facility A, which were both 75.0%. Even though the SH test tends to have higher sensitivity and lower specificity than other test methods (Hirota *et al.*, 2015), a specificity less than 40% is too low.

Therefore, we investigated the cause of the decrease in specificity and improved the protocol. Lactic acid is a typical non-sensitizing substance and is used as a negative control in the h-CLAT. Although the reported data from Facility A for lactic acid was negative and the MAC value was small, the result was negative with a large MAC value in Facility B and positive with a large MAC value in Facility C. Adjusting the pH of the medium during cell treatment to near neutral decreased the MAC value and stable negative results were obtained among the three facilities.

The two facilities (Facility A and B) that were initially negative for lactic acid used a HEPES buffer-containing medium, while Facility C, which was positive, used a HEPES buffer-free medium. Thus, we hypothesized that the SH test results might be influenced by pH. We determined that a low pH gave false positives for lactic acid; therefore, it was necessary to adjust the pH of the culture solution during cell treatment to be within the range of 5.0 to 9.0 (7.0 ± 2.0). Regarding the use of HEPES buff-

er-containing medium, the MAC value was large, even in Facility B, and the results from Facility B showed the same changes as in Facility C due to the pH adjustment. In addition, adjusting the pH was more important and the use of HEPES buffer-containing medium was set as a recommended condition rather than an indispensable condition for the SH test.

After the SH test was reported by Suzuki *et al.* (2009), the maximum applicable concentration of poorly water-soluble components was increased from 1000 µg/mL to 2500 µg/mL by Hirota *et al.* (2013). To correctly evaluate pyridine, we further increased the maximum applicable concentration of water-soluble components from 5000 µg/mL to 15000 µg/mL. In Facility B and C, the MAC value was small when evaluating 5000 µg/mL pyridine and there was almost no change in the cell surface thiols. Because the same phenomenon was observed at the two facilities, it was not considered to be due to the technical or experimental environment of the facilities. When the maximum applicable concentration was increased to 15000 µg/mL pyridine, the MAC value greatly increased. Although a positive result was obtained in Facility B, a positive result was not obtained in Facility C because the change was not statistically significant. The RFI values of each independent testing fluctuated sufficiently, but the fluctuation was not significant and the test result was not determined as a positive (data not shown). For the current SH test to produce a positive result, the mean RFI value obtained by each independent test must be $\leq 85\%$ or $\geq 115\%$ and the fluctuation must be statistically significant. The variability among each individual test must be sufficiently small to produce a significant difference relative to the solvent control group. With this method, even if all the test results are positive, it is difficult to make a positive judgment when the variability of the result is large across each test.

In the study of dispersibility, the concordance rate of the results was improved by making the dispersal state between facilities similar. We considered specifying the dispersion method, such as the mixing device, ultrasonication conditions, and processing time. When considering the possibility that the SH test will be introduced to many facilities in the future, however, the processing conditions will depend on the equipment used in each test facility and it may not be appropriate to attempt to set specific dispersion conditions. In addition, the appropriate mixing conditions will also depend on the substance being evaluated. Thus, we did not define these conditions in the SOP.

The exposure time of the test substance to the cells in the SH test was 2 hr, which, is relatively short for a test method using cultured cells. In addition, the test sys-

tem was not greatly affected even if the test substance was not completely dissolved in the medium. Therefore, we confirmed that complete dissolution of the test substance was not an essential condition during the test, and that the test can be performed on substances that are not in a uniformly dispersed state. Thus, if the dispersion state can be maintained only for 2 hr, which is the exposure time required for the SH test, reproducible results can be obtained. For 1-bromohexane, 1-iodohexane, chlorobenzene, and methyl salicylate, which had a large log *K*_{ow} and low consistency at the three facilities, a stirring method that could maintain a good dispersion state for two hours was set, and the test was conducted again. By improving the dispersion conditions, the reproducibility between facilities for these substances was increased. In the case of chlorobenzene, while cytotoxicity was observed more than when the MAC value of 5.9 was obtained, the result was considered to be the effect of improved exposure of the test substance on the cells. Similar levels of cytotoxicity were observed at the other two facilities, suggesting that the test conditions were homogenized. Although it was possible to turn false negatives to positives by improving the dispersion state, there were also some cases where negatives were turned to false positives such as the case of methyl salicylate and chlorobenzene. However, in most cases the variability in test results was reduced and the reproducibility between facilities could be enhanced by improving and homogenizing the dispersion state. This greatly contributed to the purpose behind this study. As shown in the remarks column of Table 4, among all the test conditions examined in this study, the improvement of dispersion state contributed most to the enhance of reproducibility of the 25 substances. The improvement of dispersion state also contributed to the predictability of the LLNA results. It is ideal to maintain a uniformly dispersed state but our results indicated that even if the test substance slightly precipitates in the medium, an appropriate evaluation can be made as long as it can be readily redispersed by gently shaking the test tube.

By paying attention to the dispersion state, the reproducibility rates between the facilities were improved, but the variation in the RFI value between tests was relatively large because the substances were not homogeneously dissolved. In such conditions, it is difficult to make a positive judgment and the results might be negative. When evaluating poorly water-soluble substances, it is necessary to pay attention to the variability in the RFI of each test and to judge whether the test result is appropriate or not, but the problem may be solved by improving the procedures and conditions. We will consider this issue in a

future study.

By improving the testing conditions, reproducibility rates among the three facilities was increased from 76.0% to 96.0%. For the predictability to the LLNA results, Facility C showed a marked improvement in the specificity from 37.5% to 62.5%, and sensitivity and accuracy also improved from 82.4% to 88.2% and 68.0% to 80.0%, respectively. Facility B showed an improvement in the sensitivity from 82.4% to 94.1% and the accuracy from 80.0% to 84.0%, but the specificity was slightly decreased from 75.0% to 62.5%. Facility A showed slight decrease in the accuracy from 88.0% to 84.0% and specificity from 75.0% to 62.5%. This is because, as a result of improving and homogenizing the dispersion state, exposure to the test substance was also improved and homogenized, reproducibility between facilities was improved, and false negative results turned positive, but some negative results turned also false positive. According to the report by Hirota *et al.* (2015), the predictability of the SH test for the LLNA results of 73 substances is 96.2% for sensitivity, 52.4% for specificity, and 83.6% for accuracy. Compared with these earlier reports, the predicted performance of the refined SH test was almost the same as that of the conventional SH test, and it is considered that the new test conditions resulted in better between-facility reproducibility. In addition, the performance standard for other *in vitro* skin sensitization test methods requires a criterion of 80% to be cleared for sensitivity, specificity, and accuracy (OECD, 2015a, 2015b, 2019a). Although the improved SH test did not meet the standard for specificity, it did meet the standard for sensitivity and accuracy.

Of the seven LLNA-negative substances used in this study, three substances (2-hydroxypropyl methacrylate, chlorobenzene, and methyl salicylate) were determined to be positive in the new SH test. These substances, however, also give false positives in other test methods. Specifically, 2-hydroxypropyl methacrylate was determined to be positive in the DPRA and the Keratinosens™ test methods, and chlorobenzene was determined to be positive in the h-CLAT (Hirota *et al.*, 2013).

Regarding the false-negative findings, only 1,2,4-benzenetricarboxylic anhydride in this chemical set was not detected as positive in all three facilities. According to previous reports, this substance was detected as positive in the h-CLAT, but produced a false negative in the Keratinosens™ test. In the DPRA, Hirota *et al.* (2013) reported a positive result while Urbisch *et al.* (2015) reported a false negative result. These test methods, like the SH test, are performed in an aqueous solution, and it is reportedly difficult to evaluate skin sensitizing prop-

Improving the between-facility reproducibility and usefulness of the SH test

erties of acid anhydrides under these conditions (Narita *et al.*, 2017, 2018). In addition, since 1,2,4-benzenetricarboxylic anhydride binds to the lysine group instead of the cysteine group during protein binding (Urbisch *et al.*, 2015), it is reasonable that it cannot be detected by the SH test.

This study revealed that the SH test still has some issues with the operational efficiency of the test and protocol clarity. These issues are currently being evaluated and will be reported in the near future.

In conclusion, based on the results of this study, the SOP of the SH test was improved as follows: i) the maximum applicable concentration of water-soluble substances was changed from the conventional level of 5000 µg/mL to 15000 µg/mL, ii) the pH was set to pH 7.0 ± 2 during the treatment of the test substance, and iii) the dispersion state was maintained for 2 hr in order to evaluate a poorly water-soluble substance. This study suggested that the SH test with the improved protocols is a reliable skin sensitization test with high technical transferability and between-facility reproducibility.

Conflict of interest---- The authors declare that there is no conflict of interest.

REFERENCES

- Adler, S., Basketter, D., Creton, S., Pelkonen, O., van Benthem, J., Zuang, V., Andersen, K. E., Angers-Loustau, A., Aptula, A., Bal-Price, A., Benfenati, E., Bernauer, U., Bessems, J., Bois, F.Y., Boobis, A., Brandon, E., Bremer, S., Broschard, T., Casati, S., Coecke, S., Corvi, R., Cronin, M., Daston, G., Dekant, W., Felter, S., Grignard, E., Gundert-Remy, U., Heinonen, T., Kimber, I., Kleinjans, J., Komulainen, H., Kreiling, R., Kreysa, J., Batista, Leite, S., Loizou, G., Maxwell, G., Mazzatorta, P., Munn, S., Pfuhler, S., Phrakonkham, P., Piersma, A., Poth, A., Prieto, P., Repetto, G., Rogiers, V., Schoeters, G., Schwarz, M., Serafimova, R., Tähti, H., Testai, E., van Delft, J., van Loveren, H., Vinken, M., Worth, A. and Zaldivar, J.M. (2011): Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. *Arch. Toxicol.*, **85**, 367-485.
- Aiba, S., Manome, H., Nakagawa, S., Mollah, Z.U., Mizuashi, M., Ohtani, T., Yoshino, Y. and Tagami, H. (2003): p38 Mitogen-activated protein kinase and extracellular signal-regulated kinases play distinct roles in the activation of dendritic cells by two representative haptens, NiCl₂ and 2,4-dinitrochlorobenzene. *J. Invest. Dermatol.*, **120**, 390-399.
- Ashikaga, T., Yoshida, Y., Hirota, M., Yoneyama, K., Itagaki, H., Sakaguchi, H., Miyazawa, M., Ito, Y., Suzuki, H. and Toyoda, H. (2006): Development of an *in vitro* skin sensitization test using human cell lines: the human Cell Line Activation Test (h-CLAT). I. Optimization of the h-CLAT protocol. *Toxicol. In Vitro*, **20**, 767-773.
- Council of the European Union and European Parliament. (2003): Directive 2003/15/EC of the European Parliament and of the council of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products. *Official Journal*, **L66**, 26-35.
- dos Santos, G.G., Reinders, J., Ouwehand, K., Rustemeyer, T., Scheper, R.J. and Gibbs, S. (2009): Progress on the development of human *in vitro* dendritic cell based assays for assessment of the sensitizing potential of a compound. *Toxicol. Appl. Pharmacol.*, **236**, 372-382.
- ESAC. (2016): ESAC opinion on the BASF-coordinated performance standards-based validation of the LuSens test method for skin sensitisation testing, Publications Office of the European Union.
- EURL-ECVAM. (2013): EC EURL-ECVAM recommendation on the Direct Peptide Reactivity Assay (DPRA) for the skin sensitisation testing. Publications Office of the European Union.
- EURL-ECVAM. (2014): EURL-ECVAM recommendation on the KeratinoSens™ assay for skin sensitisation testing. Publications Office of the European Union.
- Fujita, M., Yamamoto, Y., Tahara, H., Kasahara, T., Jimbo, Y. and Hioki, T. (2014): Development of a prediction method for skin sensitization using novel cysteine and lysine derivatives. *J. Pharmacol. Toxicol. Methods*, **70**, 94-105.
- Gerberick, G.F., Vassallo, J.D., Bailey, R.E., Chaney, J.G., Morrall, S.W. and Lepoittevin, J.P. (2004): Development of a peptide reactivity assay for screening contact allergens. *Toxicol. Sci.*, **81**, 332-343.
- Gerberick, G.F., Vassallo, J.D., Foertsch, L.M., Price, B.B., Chaney, J.G. and Lepoittevin, J.P. (2007): Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach. *Toxicol. Sci.*, **97**, 417-427.
- Hirota, M., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., Fujishiro, M., Kyotani, D., Kato, Y., Kasahara, T., Fujita, M., Toyoda, A., Sekiya, D., Watanabe, S., Seto, H., Takenouchi, O., Ashikaga, T. and Miyazawa, M. (2015): Evaluation of combinations of *in vitro* sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization. *J. Appl. Toxicol.*, **35**, 1333-1347.
- Hirota, M., Kouzuki, H., Ashikaga, T., Sono, S., Tsujita, K., Sasa, H. and Aiba, S. (2013): Artificial neural network analysis of data from multiple *in vitro* assays for prediction of skin sensitization potency of chemicals. *Toxicol. In Vitro*, **27**, 1233-1246.
- Hirota, M., Motoyama, A., Suzuki, M., Yanagi, M., Kitagaki, M., Kouzuki, H., Hagino, S., Itagaki, H., Sasa, H., Kagatani, S. and Aiba, S. (2010): Changes of cell-surface thiols and intracellular signaling in human monocytic cell line THP-1 treated with diphenylcyclopropanone. *J. Toxicol. Sci.*, **35**, 871-879.
- Hirota, M., Suzuki, M., Hagino, S., Kagatani, S., Sasaki, Y., Aiba, S. and Itagaki, H. (2009): Modification of cell-surface thiols elicits activation of human monocytic cell line THP-1: possible involvement in effect of haptens 2,4-dinitrochlorobenzene and nickel sulfate. *J. Toxicol. Sci.*, **34**, 139-150.
- ICCVAM. (1999): The murine local lymph node assay: A test method for assessing the allergic contact dermatitis potential of chemicals/compounds, The results of an independent peer review evaluation coordinated by the interagency coordinating committee on the validation of alternative methods (ICCVAM) and the national toxicology program interagency center for the evaluation of alternative toxicological methods (NICEATM), NIH Publication No. 99-4494, Research Triangle Park, NC: National Institute of Environmental Health Sciences.
- Narita, K., Ishii, Y., Vo, P.T., Nakagawa, F., Ogata, S., Yamashita, K., Kojima, H. and Itagaki, H. (2018): Improvement of human

- cell line activation test (h-CLAT) using short-time exposure methods for prevention of false-negative results. *J. Toxicol. Sci.*, **43**, 229-240.
- Narita, K., Vo, P.T., Yamamoto, K., Kojima, H. and Itagaki, H. (2017): Preventing false-negatives in the *in vitro* skin sensitization testing of acid anhydrides using interleukin-8 release assays. *Toxicol. In Vitro*, **42**, 69-75.
- Natsch, A. (2010): The Nrf2-Keap1-ARE toxicity pathway as a cellular sensor for skin sensitizers--functional relevance and a hypothesis on innate reactions to skin sensitizers. *Toxicol. Sci.*, **113**, 284-292.
- Natsch, A. and Emter, R. (2008): Skin sensitizers induce antioxidant response element dependent genes: application to the *in vitro* testing of the sensitization potential of chemicals. *Toxicol. Sci.*, **102**, 110-119.
- OECD. (2010): Test Guideline No. 429. Skin sensitisation: local lymph node assay. In: OECD Guidelines for the Testing of Chemicals), Organisation for Economic Cooperation and Development, Paris.
- OECD. (2012a): The adverse outcome pathway for skin sensitisation initiated by covalent binding to proteins. Part 1: Scientific evidence. In: Series on Testing & Assessment), Organisation for Economic Cooperation and Development, Paris.
- OECD. (2012b): The adverse outcome pathway for skin sensitisation initiated by covalent binding to proteins. Part 2: Use of the AOP to develop chemical categories and integrated assessment and testing approaches. In: Series on Testing & Assessment), Organisation for Economic Cooperation and Development, Paris.
- OECD. (2015a): Performance standards for assessment of proposed similar or modified *in vitro* skin sensitisation ARE-NRF2 luciferase test methods. In: OECD Environment, Health and Safety Publications: Series on Testing & Assessment), Organisation for Economic Cooperation and Development, Paris.
- OECD. (2015b): Performance standards for the assessment of proposed similar or modified *in vitro* reconstructed human epidermis (RhE) test methods for skin irritation testing as described in TG 439. In: OECD Environment, Health and Safety Publications: Series on Testing & Assessment), Organisation for Economic Cooperation and Development, Paris.
- OECD. (2016a): Annex I: Case studies to the guidance document on the reporting of defined approaches and individual information sources to be used within integrated approaches to testing and assessment (IATA) for skin sensitisation. In: OECD Environment, Health and Safety Publications: Series on Testing & Assessment), Organisation for Economic Cooperation and Development, Paris.
- OECD. (2016b): Guidance document for the reporting of defined approaches and individual information sources to be used within integrated approaches to testing and assessment (IATA) for skin sensitisation. In: OECD Environment, Health and Safety Publications: Series on Testing & Assessment), Organisation for Economic Cooperation and Development, Paris.
- OECD. (2016c): Guidance document for the use of adverse outcome pathways in developing integrated approaches to testing and assessment (IATA). In: OECD Environment, Health and Safety Publications: Series on Testing & Assessment), Organisation for Economic Cooperation and Development, Paris.
- OECD. (2018a): Test Guideline No. 442D. *In vitro* skin sensitisation assays addressing the key event on keratinocyte activation. In: OECD Guidelines for the Testing of Chemicals), Organisation for Economic Cooperation and Development, Paris.
- OECD. (2018b): Test Guideline No. 442E. *In vitro* skin sensitisation assays addressing the key event on activation of dendritic cells on the adverse outcome pathway for skin sensitisation. In: OECD Guidelines for the Testing of Chemicals), Organisation for Economic Cooperation and Development, Paris.
- OECD. (2019a): Performance standards for the assessment of proposed similar or modified *in vitro* skin sensitisation DPRA and ADRA test methods. In: OECD Environment, Health and Safety Publications: Series on Testing & Assessment), Organisation for Economic Cooperation and Development, Paris.
- OECD. (2019b): Report of the peer review panel on the validation of the amino acid derivative reactivity assay (ADRA). In: Series on Testing & Assessment), Organisation for Economic Cooperation and Development, Paris.
- OECD. (2019c): Test Guideline No. 442C. Key-event-based test guideline for *In chemico* skin sensitisation assays addressing the adverse outcome pathway key event on covalent binding to proteins. In: OECD Guidelines for the Testing of Chemicals), Organisation for Economic Cooperation and Development, Paris.
- Ramirez, T., Mehling, A., Kolle, S.N., Wruck, C.J., Teubner, W., Eltze, T., Aumann, A., Urbisch, D., van Ravenzwaay, B. and Landsiedel, R. (2014): LuSens: a keratinocyte based ARE reporter gene assay for use in integrated testing strategies for skin sensitization hazard identification. *Toxicol. In Vitro*, **28**, 1482-1497.
- Sakaguchi, H., Ashikaga, T., Miyazawa, M., Yoshida, Y., Ito, Y., Yoneyama, K., Hirota, M., Itagaki, H., Toyoda, H. and Suzuki, H. (2006): Development of an *in vitro* skin sensitization test using human cell lines; human Cell Line Activation Test (h-CLAT). II. An inter-laboratory study of the h-CLAT. *Toxicol. In Vitro*, **20**, 774-784.
- Suzuki, M., Hirota, M., Hagino, S., Itagaki, H. and Aiba, S. (2009): Evaluation of changes of cell-surface thiols as a new biomarker for *in vitro* sensitization test. *Toxicol. In Vitro*, **23**, 687-696.
- Urbisch, D., Mehling, A., Guth, K., Ramirez, T., Honarvar, N., Kolle, S., Landsiedel, R., Jaworska, J., Kern, P.S., Gerberick, F., Natsch, A., Emter, R., Ashikaga, T., Miyazawa, M. and Sakaguchi, H. (2015): Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul. Toxicol. Pharmacol.*, **71**, 337-351.