

RESEARCH ARTICLE

Development of an artificial neural network model for risk assessment of skin sensitization using human cell line activation test, direct peptide reactivity assay, KeratinoSens™ and in silico structure alert parameter

Morihiro Hirota  | Takao Ashikaga | Hirokazu Kouzuki

Shiseido Global Innovation Center, Shiseido Co. Ltd., 2-2-1 Hayabuchi, Tsuzuki-ku, Yokohama-shi, Kanagawa 224-8558, Japan

Correspondence

Morihiro Hirota, Shiseido Global Innovation Center, Yokohama-shi, Kanagawa, Japan 224-8558.
Email: morihiko.hirota@to.shiseido.co.jp

Abstract

It is important to predict the potential of cosmetic ingredients to cause skin sensitization, and in accordance with the European Union cosmetic directive for the replacement of animal tests, several in vitro tests based on the adverse outcome pathway have been developed for hazard identification, such as the direct peptide reactivity assay, KeratinoSens™ and the human cell line activation test. Here, we describe the development of an artificial neural network (ANN) prediction model for skin sensitization risk assessment based on the integrated testing strategy concept, using direct peptide reactivity assay, KeratinoSens™, human cell line activation test and an in silico or structure alert parameter. We first investigated the relationship between published murine local lymph node assay EC3 values, which represent skin sensitization potency, and in vitro test results using a panel of about 134 chemicals for which all the required data were available. Predictions based on ANN analysis using combinations of parameters from all three in vitro tests showed a good correlation with local lymph node assay EC3 values. However, when the ANN model was applied to a testing set of 28 chemicals that had not been included in the training set, predicted EC3s were overestimated for some chemicals. Incorporation of an additional in silico or structure alert descriptor (obtained with TIMES-M or Toxtree software) in the ANN model improved the results. Our findings suggest that the ANN model based on the integrated testing strategy concept could be useful for evaluating the skin sensitization potential.

KEYWORDS

adverse outcome pathway, DPRA, h-CLAT, integrated testing strategy, KeratinoSens™, artificial neural network, risk assessment, skin sensitization, TIMES, Toxtree

1 | INTRODUCTION

The European Union (EU) cosmetic directive for the replacement of animal tests in safety testing of cosmetic ingredients came into effect in 2013. Therefore, previously used animal model tests for evaluating skin sensitization potential, such as the guinea pig maximization test (Magnusson & Kligman, 1969; OECD Test Guideline 406, OECD, 1992) and the murine local lymph node assay (LLNA; OECD Test Guideline 429, OECD, 2010), are no longer available for safety assessment of cosmetics in the EU and various in vitro test methods have been developed to replace them.

Skin sensitization can be divided into the induction phase and the elicitation phase. The LLNA detects cellular proliferation in the draining

lymph nodes after repeated topical application of a test compound on the ears in the induction phase of skin sensitization (Basketter et al., 2002). LLNA can be used to categorize skin sensitization potency (e.g., non-sensitizers, and weak, moderate, strong and extremely strong sensitizers). In particular, prediction of a safe level of human exposure using a quantitative risk assessment approach based on the threshold of positivity in LLNA, such as the EC3 value, was widely used in the calculation of acceptable exposure levels of fragrances for consumers by the International Fragrance Association (Api et al., 2008; Basketter et al., 2008). Many chemicals (>300) have been evaluated by LLNA (Gerberick et al., 2005; Kern et al., 2010), and this large database

obtained with the same *in vivo* method is a valuable resource for the development and validation of *in vitro* or *in silico* alternative methods for predicting skin sensitization.

Many researchers, including our group, have set out to develop *in vitro* skin sensitization tests. For example, the direct peptide reactivity assay (DPRA) and LC/MC-based peptide reactivity assay (Gerberick et al., 2007; Natsch & Gfeller, 2008; OECD, 2015a) reflect binding between protein and haptens, which is necessary for antigen presentation from antigen-presenting cells to T cells. The SH test, which measures changes of cell-surface thiols induced by haptens, reflects not only activation of intracellular signal transduction, but also antigen production, such as hapten–protein complex formation (Hirota et al., 2013; Suzuki, Hirota, Hagino, Itagaki, & Aiba, 2009). The antioxidant response element (ARE) assay (Natsch & Emter, 2008), LuSens (Urbisch et al., 2015) and KeratinoSens™ (Emter, Ellis, & Natsch, 2010) are based on stimulation of ARE-dependent gene activity in a recombinant cell line, and they detect activation of the Keap1-Nrf2-ARE regulatory pathway as an emerging innate toxic response (Emter et al., 2010; OECD, 2015b). On the other hand, the human cell line activation test (h-CLAT) (Ashikaga et al., 2006; OECD, 2016a; Sakaguchi et al., 2006), U-SENS (Piroird et al., 2015), interleukin-8 Luc assay (Kimura et al., 2015) and Genomic Allergen Rapid Detection using MUTZ-3 cells (Johansson et al., 2014) are based on activation of antigen-presenting cells, including dendritic cells, in response to hapten treatment (Aiba & Tagami, 1998; Aiba, Terunuma, Manome, & Tagami, 1997; Boislevé, Kerdine-Romer, Rougier-Larzat, & Pallardy, 2004).

However, it is widely recognized that a single *in vitro* test is insufficient to replace animal testing and that integration of results from different *in vitro* tests, as well as *in silico* methods, is needed for reliable prediction of the skin sensitization potential of chemicals (Jaworska, Harol, Kern, & Gerberick, 2011; Jowsey, Basketter, Westmoreland, & Kimber, 2006; Natsch, Emter, & Ellis, 2009). Therefore, prediction models based on an integrated testing strategy approach have been developed, using combinations of *in vitro* sensitization assays and *in silico* modeling to integrate information on various biological, biochemical and chemical factors that impact on the allergenicity of chemicals and the acquisition of skin sensitization (Jowsey et al., 2006; OECD, 2016b). Several prediction models based on this concept have been reported (Jaworska et al., 2011; Jaworska, Dancik, Kern, Gerberick, & Natsch, 2013; Nukada, Miyazawa, Saitou, Sakaguchi, & Nishiyama, 2013; Urbisch et al., 2015). However, although the models are based on a similar concept, the prediction targets were not the same, as some focused on hazard identification and others on skin sensitization potency. In the latter case, prediction of classification GHS 1A/1B, skin sensitization potency category and skin sensitization threshold, such as LLNA EC3 value, is the target. Moreover, a variety of methods has been used to combine the inputs. Nukada et al. reported a score-based prediction system combining several *in vitro* assays and a commercial *in silico* model, and Jaworska et al. developed Bayesian network (BN)-based prediction models (Jaworska et al., 2011; Jaworska et al., 2013; Nukada et al., 2013). BN networks have a data-based learning function, and it was demonstrated that LLNA potency category (strong/moderate/weak sensitizer/non-sensitizer) could be predicted with a BN model consisting of *in vitro* tests such as DPRA and KeratinoSens™ and *in*

silico parameters such as a commercial prediction model (TIMES-M) and epidermal bioavailability (Jaworska et al., 2011; Jaworska et al., 2013; Jaworska et al., 2015).

On the other hand, we have focused on artificial neural network (ANN) analysis, which is a non-linear statistical data-modeling tool that is usually used to model complex relationships between inputs and outputs. We first used ANN analysis to integrate data from the SH test and h-CLAT and found that this model, named as iSENS ver. 1 (integrated *in vitro* sensitization tests by ANN, version 1), could directly predict the LLNA thresholds (Hirota et al., 2013). We subsequently demonstrated that ARE assay is useful as an additional descriptor for improvement of the prediction model (Hirota et al., 2015; Tsujita-Inoue et al., 2014). Furthermore, in collaborative studies with the Japan Cosmetic Industry Association, we developed a large h-CLAT and DPRA database consisting of over 130 chemicals (Takenouchi et al., 2015). On the other hand, the developers of DPRA and KeratinoSens™ reported a data set of 145 chemicals, including U-937 assay, which detects augmentation of CD86 expression on test chemical-treated U-937 cells (Urbisch et al., 2015). Finally, h-CLAT, DPRA and KeratinoSens™ were validated and accepted as OECD guidelines. Recently it was reported that a BN model including the results of h-CLAT, DPRA and KeratinoSens™ showed good predictive performance for the LLNA potency category (Jaworska et al., 2015).

However, our aim is estimation of exposure levels of cosmetic ingredients that can be used safely by consumers. LLNA EC3 was reported to be useful for estimating acceptable exposure level of chemicals in humans for a skin sensitization endpoint (Api et al., 2008). Therefore, we wished to know whether the combination of the above three tests could be used to predict skin sensitization potency (LLNA EC3) using our ANN-based model.

Here, we aimed to determine whether an ANN model using h-CLAT, DPRA and KeratinoSens™ would be useful for predicting LLNA EC3 values, since many chemicals have already been evaluated with these three tests, which have been adopted as OECD guidelines.

2 | MATERIALS AND METHODS

2.1 | Selection of chemicals and sources of *in vivo* data

Chemicals for which LLNA, h-CLAT, DPRA and KeratinoSens™ data were all available were selected. As described in our previous work, the LLNA database was collected from two papers (Gerberick et al., 2005; Kern et al., 2010), which were considered to have high-quality LLNA *in vivo* data, all obtained according to the OECD 429 protocol with a clear dose–response curve or a negative result (Natsch et al., 2013). We also took LLNA EC3 reference data from a recent paper by Jaworska et al. (2015).

2.2 | Procedure of direct peptide reactivity assay

DPRA data were kindly provided by Procter & Gamble Company (Strombeek-Bever, Belgium), which was the developer of DPRA. The assays had been performed as described elsewhere (Gerberick et al.,

2007; Natsch et al., 2013). Briefly, test chemicals were incubated with peptides containing lysine (Ac-RFAAKAA-COOH) or cysteine (Ac-RFAACAACOOH) in the dark for 24 hours at 25°C, and then the peptide was quantified by reverse-phase HPLC with UV detection. The percentage peptide depletion and percentage remaining peptide were calculated for each sample by comparing the amount of non-reacted peptide in the reaction sample to the amount of peptide in the peptide-alone control samples.

$$\text{Relative fluorescence intensity (\%)} = \frac{(\text{MFI of chemical-treated cells} - \text{MFI of chemical-treated isotype control cells})}{(\text{MFI of vehicle control cells} - \text{MFI of vehicle isotype control cells})} \times 100 \quad (1)$$

The mean percentage depletion value was calculated as the average of peptide depletions of cysteine and lysine peptides (percentage cysteine depletion and percentage lysine depletion, respectively), and classification of chemicals as sensitizer or non-sensitizer was based on the OECD guideline (OECD, 2015a). In the following analysis, when the percentage remaining cysteine peptide value or the percentage remaining lysine peptide value was negative, zero (0) or unity (1), a value of 0.01 was used as a substitute value.

2.3 | Procedure of KeratinoSens™ assay

The KeratinoSens™ data were provided by Givaudan Schweiz AG, which was the developer of KeratinoSens™. The assays had been performed as described elsewhere (Emter et al., 2010; OECD, 2015b). Briefly, cells were grown for 24 hours in 96-well plates and then the medium was then replaced with medium containing the test chemical. Each compound was tested at 12 binary dilutions in the range from 0.98 to 2000 µM. Cells were incubated for 48 hours with test chemical, and then luciferase activity and cytotoxicity (MTT assay; Mossman, 1983) were determined. The gene induction was compared with that of dimethyl sulfoxide controls, and the wells with statistically significant induction over the threshold of 1.5 (i.e., 50% enhanced gene activity) were determined. The EC1.5 and EC3 values (concentration in µM for induction above these thresholds) were determined by linear extrapolation as described in the OECD guideline (OECD, 2015a). In the prediction model, chemicals with significant gene induction (>1.5-fold) at a concentration below 1000 µM and at a concentration at which the cells retain 70% viability in at least two of three repetitions are rated positive.

2.4 | Procedure of human cell line activation test

h-CLAT were from a Kao/Shiseido collaborative study (Nukada et al., 2012), from the Japan Cosmetic Industry Association (Takenouchi et al., 2015) or from a P&G/Givaudan/Kao/Shiseido collaborative study (Jaworska et al., 2015). The assays had been performed as described elsewhere (OECD, 2016a). Briefly, THP-1 cells were plated at 1×10^6 cells ml⁻¹ in a 24-well plate and treated for 24 hours with test chemicals. The test dose was determined based on the test concentration providing a cell viability of 75% (CV75) in the cytotoxicity test. Cells were washed and FcR was blocked. Staining was done with fluorescein isothiocyanate

(FITC)-conjugated antihuman CD86 antibody (clone Fun-1; BD-PharMingen, San Diego, CA, USA) or FITC-conjugated anti-human CD54 antibody (clone 6.5B5; DAKO, Glostrup, Denmark) at 4°C for 30 minutes. FITC-labeled mouse IgG1 (clone DAK-G01; DAKO) was used as an isotype control. The cells were washed, and expression of cell surface antigens was analyzed by flow cytometry. The relative fluorescence intensity, calculated according to Equation 1, was used as an indicator of CD86 and CD54 expression.

The thresholds for CD86 and CD54 expression (EC150 for CD86 and EC200 for CD54) were determined as described previously (Nukada et al., 2012). The unit of EC150 and EC200 (µg ml⁻¹) was converted into molar unit (µM). In the following analysis, when EC150 or EC200 could not be calculated because of negative results, a value of 100 000 (µM) was used as a substitute value. A minimum induction threshold (MIT) was determined as the smaller of either EC150 or EC200.

2.5 | Times

Data obtained with TIMES software (V.2.25.7) (Dimitrov et al., 2005) were provided by Procter & Gamble Company. Predictions based on the parent molecule (TIMES-P), as well as considering potential skin metabolism (TIMES-M), were investigated in the study. Prediction data were taken from a recent paper (Jaworska et al., 2015).

2.6 | Toxtree

Toxtree version 2.5.0, a quantitative structure–activity relationship (QSAR) tool for mechanism-based prediction of skin sensitization potential, was downloaded from the website of the Joint Research Center in EURL- ECVAM (<http://toxtree.sourceforge.net>). Toxtree determines whether defined skin sensitization alerts (SN2, acyl transfer agent, Michael acceptor, Schiff base formation and SNAr) are present in test chemical structure (Enoch, Madden, & Cronin, 2008). In the following analysis, the results of Toxtree were converted into the following characteristic values to incorporate them into ANN analysis (Table 1).

TABLE 1 Results of Toxtree were converted into the following characteristic values to incorporate them into ANN analysis

Toxtree alerts	Characteristic values
No alert	1
SN2	2
Acyl transfer agent	3
Schiff base formation	4
Michael acceptor	5
SNAr	6
>2 alerts	7

2.7 | Artificial neural network analysis

A typical ANN consists of an input layer, hidden layer and output layer, as previously reported (Hirota et al., 2013). In this study, we used two hidden layers, with published LLNA thresholds as the output layer. The descriptors used for the input layer were correlated with LLNA EC3s using the neural network. All calculations were performed using QwikNet Ver.2.23. In ANN learning with a configuration of 4-5-2-1, 5-5-2-1 or 6-5-2-1 for the input, first hidden, second hidden and output layers, an online back-propagation algorithm was used for network training, with values of the learning rate and momentum parameter of 0.02 and 0.7, respectively. The range of weights was set to be -5 to 5 (weight perturbation; 50%). When we constructed the prediction models, the data shown in Supporting information S1 were used as the training data set and incorporated in "Training Data" and "Testing Data" in QwikNet software. After training, we evaluated the correlation between predicted LLNA thresholds and published LLNA thresholds, and saved the weight parameter.

The predictive ability of the neural network model was evaluated by means of a 10-fold cross-validation procedure (Hirota et al., 2013). The goodness-of-fit was evaluated in terms of root-mean-square (RMS) error, defined as follows:

$$\text{RMS error} = \sqrt{\sum ((\text{measured value} - \text{predicted value})^2) / \text{number of data}}$$

2.8 | Potency estimation in local lymph node assay and in the artificial neural network prediction model by classification and regression tree analysis

Sensitization potency in LLNA was classified into extreme (EC3 < 0.1%) or strong (0.1% ≤ EC3 < 1%), moderate (1% ≤ EC3 < 10%), weak (10% ≤ EC3 < 100%) and non-sensitizing in LLNA (NS). On the other hand, the skin sensitization potency predicted by ANN was classified into extreme or strong (predicted EC3 < 1%), moderate (1% ≤ predicted EC3 < 10%), weak (10% ≤ predicted LLNA EC3 < 100%) and negative (predicted EC3 ≥ 100% or negative in hazard judgment based on in vitro tests, h-CLAT, DPRA and KeratinoSens™). In addition, sensitization potency in LLNA or the ANN models was classified into actual or predicted 1A (strong sensitizer) (EC3 in LLNA or predicted LLNA threshold < 2%) and 1B (weak sensitizer) (EC3 in LLNA or predicted

LLNA EC3 ≥ 2%) according to the globally harmonized system of classification and labeling of chemicals (GHS) (United Nations, 2009).

3 | RESULTS

3.1 | Correlation between local lymph node assay data and in vitro tests (human cell line activation test, direct peptide reactivity assay, KeratinoSens™) (hazard identification: Positive/negative judgment)

The training data set of chemicals assessed using h-CLAT, DPRA and KeratinoSens™ assay is listed in Supporting information S1. The h-CLAT, DPRA and KeratinoSens™ data in Supporting information S1 were taken from previous reports (Natsch et al., 2013; Takenouchi et al., 2015). As shown in Table 2, the predictive accuracy, sensitivity (proportion of all positive chemicals that are correctly classified as positive) and specificity (proportion of all negative chemicals that are correctly classified as negative) of h-CLAT, DPRA and KeratinoSens™ vs LLNA (*n* = 134, Supporting information S1) were 76.1%, 79.8% and 67.5% (h-CLAT), 76.9%, 76.6% and 77.5% (DPRA) and 73.1%, 77.7% and 62.5% (KeratinoSens™), respectively. As for hazard judgment based on the combination of h-CLAT, DPRA KeratinoSens™, the "bottom-up 3 out of 3 criteria" (test chemical is judged as negative when three of three tests are negative) resulted in accuracy/sensitivity/specificity of 79.9%/96.8%/40.0%, while the "bottom-up 2 out of 3 criteria" (test chemical is judged as negative when two of three tests are negative) resulted in accuracy/sensitivity/specificity of 79.1%/80.9%/75.0% (Table 2). These results are broadly consistent with previous findings (Urbisch et al., 2015).

3.2 | Correlation between local lymph node assay EC3 value and descriptors derived from in vitro tests (human cell line activation test, direct peptide reactivity assay, KeratinoSens™)

The relationships between LLNA data and various in vitro test results have been analyzed in previous reports (Emter et al., 2010; Gerberick et al., 2004; Nukada et al., 2012). For the development of the ANN models, we first confirmed these relationships.

From h-CLAT, MIT (minimum induction threshold, the lower value of EC150 and EC200 [Nukada et al., 2012]), CV75 value (concentration

TABLE 2 Correlations between LLNA and each in vitro test (h-CLAT, DPRA and KeratinoSens™) and the combination (positive/negative judgment)

			h-CLAT/DPRA/KeratinoSens™									
			h-CLAT		DPRA		KeratinoSens™		bottom-up 3 out of 3			
			Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
n												
LLNA	Positive	94	75	19	72	22	73	21	91	3	76	18
	Negative	40	13	27	9	31	15	25	24	16	10	30
Accuracy (%)			76.1		76.9		73.1		79.9		79.1	
Sensitivity (%)			79.8		76.6		77.7		96.8		80.9	
Specificity (%)			67.5		77.5		62.5		40.0		75.0	

Italic numbers indicate that the same results were obtained in LLNA and in each of h-CLAT, DPRA and KeratinoSens™. In the "bottom-up 3 out of 3 criteria", the chemical was judged as negative when all three tests (h-CLAT, DPRA and KeratinoSens™) were negative. In the "bottom-up 2 out of 3 criteria", the chemical was judged as negative when two of three tests (h-CLAT, DPRA and KeratinoSens™) were negative.

giving 75% cell viability) in cytotoxicity assay and the minimum of MIT and CV75 were used as descriptors. From DPRA, the percentage remaining peptide values for cysteine and lysine peptides were used as descriptors (Jaworska et al., 2015). From KeratinoSens™, EC1.5 was used as a descriptor. These data were compared with LLNA EC3 values.

As shown in Figure 1, all descriptors were significantly correlated with LLNA EC3 values. However, the descriptors of h-CLAT showed better correlations (higher r values) than those of DPRA and KeratinoSens™.

3.3 | Artificial neural network analysis using human cell line activation test, direct peptide reactivity assay and KeratinoSens™

We next evaluated the utility of the ANN model with h-CLAT, DPRA and KeratinoSens™. When we previously evaluated DPRA (Hirota et al., 2015), the mean percentage peptide depletion of cysteine and lysine peptides was used as a parameter, as it is proposed to be a suitable indicator for positive/negative judgment (OECD, 2015a). However, in this study, the percentage remaining values of cysteine

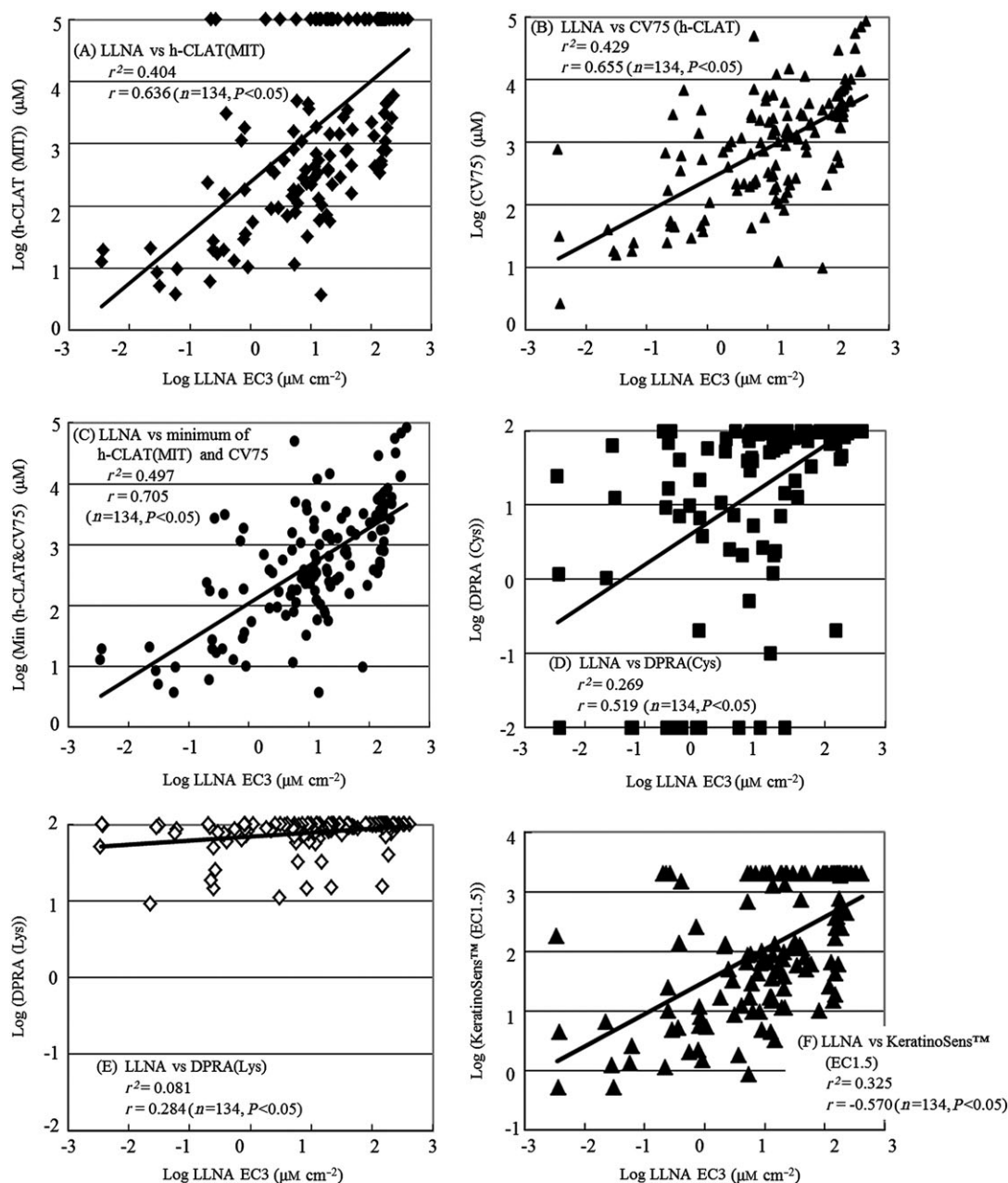


FIGURE 1 Relationships between LLNA threshold and parameters of individual in vitro tests (h-CLAT, DPRA and KeratinoSens™). Relationship between LLNA EC3 and the minimum value of thresholds for CD86 and CD54 in h-CLAT (MIT μM) (A), the CV75 (cell toxicity) in h-CLAT (B), minimum value of MIT and CV75 in h-CLAT (C), the percentage remaining cysteine peptide values (D) and the percentage remaining lysine peptide values (E) in DPRA, and EC1.5 in KeratinoSens™ (F). Data in these plots were taken from Supporting information S1. Values of correlation coefficient (r value) were 0.636 (A), 0.655 (B), 0.705 (C) and 0.519 (D), 0.284 (E) and 0.570 (F). DPRA, direct peptide reactivity assay; h-CLAT, human cell line activation test; LLNA, local lymph node assay; MIT, minimum induction threshold

and lysine peptides were used as independent descriptors, because we found that these parameters were not correlated to each other.

In the case of h-CLAT, we used MIT or the minimum of MIT and CV75 as the descriptor for h-CLAT. We previously used only the minimum of MIT and CV75 because the two parameters were highly correlated. However, this high correlation exists only for h-CLAT-positive chemicals. In this study, we examined either the minimum of MIT and CV75 or MIT alone as a descriptor of the h-CLAT result. The descriptors used in each of the ANN models were summarized in Table 3.

First, we conducted ANN analysis using the minimum of MIT and CV75 for h-CLAT, percentage remaining cysteine peptide values and percentage remaining lysine peptide values for DPRA and EC1.5 for KeratinoSens™ as descriptors (ANN model 1). The correlation coefficient (*r* value) and RMS error were 0.897 and 0.501, respectively. When we assessed this model by 10-fold cross-validation, the *r* value and RMS error were 0.874 ± 0.008 and 0.549 ± 0.015 , respectively (Table 4). Next, when ANN analysis was performed using MIT as an indicator of the threshold concentration of h-CLAT, CV75 (indicating cell toxicity), percentage remaining cysteine peptide values, percentage remaining lysine peptide values and EC1.5 as descriptors (ANN model 2), the *r* value/RMS error were 0.903/0.486 (before cross-validation) and $0.885 \pm 0.010/0.525 \pm 0.021$ (after cross-validation) (Table 4).

3.4 | Model evaluation using a testing panel of chemicals not in the training set

Next, the developed ANN models were evaluated using a testing data set of 28 chemicals (20 sensitizers and eight non-sensitizers) for which h-CLAT, DPRA and KeratinoSens™ data were reported (Jaworska et al., 2015) (summarized in Supporting information S2). When we evaluated these 28 chemicals using the ANN models 1 and 2, the EC3 values of some chemicals predicted by ANN model 1 were very much lower than the published EC3 values or the EC3 values predicted by ANN model 2 (Figure 2A,B).

3.5 | Incorporation of in silico parameter or structure alert into the artificial neural network model

An in silico model for providing structure alerts for skin sensitization has been incorporated as a part of the integrated testing strategy concept (Jaworska et al., 2013; Jowsey et al., 2006; Takenouchi et al., 2015). Therefore, we investigated the effect of adding in silico input in our ANN analysis. TIMES-M was reported to improve a BN prediction model (Jaworska et al., 2011). Furthermore, we examined Toxtree, which is structure alert program released by the Joint Research Center in EURL-ECVAM. The combinations of descriptors examined in ANN analysis are shown in Table 3. The *r* value, *r*² value

TABLE 3 Descriptors used in each of the ANN models

Descriptors		ANN models					
		ANN model 1 (Figure 2A)	ANN model 2 (Figure 2B)	ANN model 3	ANN model 4 (Figure 3A)	ANN model 5	ANN model 6 (Figure 3B)
h-CLAT	MIT (μM)		x		x		x
Cell toxicity	CV75 (μM)		x		x		x
Minimum of h-CLAT and cell toxicity	Min (MIT/CV75) (μM)	x		x		x	
DPRA (% remaining peptide values)	Cys	x	x	x	x	x	x
	Lys	x	x	x	x	x	x
KeratinoSens™	EC1.5 (μM)	x	x	x	x	x	x
In silico/structure alert	TIMES-M			x	x		
	Toxtree					x	x
	Number of descriptors	4	5	5	6	5	6

Descriptors marked with check-mark "x" were used in the ANN models. Min (MIT/CV75) indicates the minimum of MIT and CV75.

TABLE 4 *r* values, *r*² values and RMS errors of each of the ANN models

ANN models	<i>r</i> values		<i>r</i> values		RMS error	
	Before cross-validation	After cross-validation	Before cross-validation	After cross-validation	Before cross-validation	After cross-validation
ANN model 1 (Figure 2A)	0.897	0.874 ± 0.008	0.805	0.764 ± 0.013	0.501	0.549 ± 0.015
ANN model 2 (Figure 2B)	0.903	0.885 ± 0.010	0.816	0.784 ± 0.018	0.486	0.525 ± 0.021
ANN model 3	0.921	0.896 ± 0.012	0.848	0.802 ± 0.021	0.442	0.503 ± 0.027
ANN model 4 (Figure 3A)	0.935	0.911 ± 0.004	0.874	0.829 ± 0.008	0.402	0.467 ± 0.010
ANN model 5	0.900	0.866 ± 0.003	0.809	0.750 ± 0.005	0.494	0.565 ± 0.006
ANN model 6 (Figure 3B)	0.926	0.897 ± 0.005	0.857	0.805 ± 0.009	0.429	0.500 ± 0.012

r values, *r*² values and RMS errors of the ANN models before and after 10-fold cross-validation are listed.

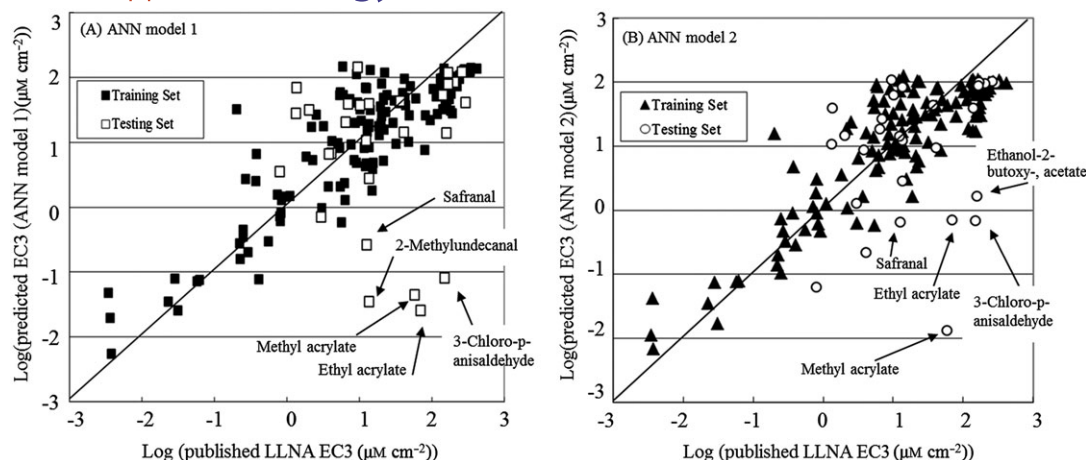


FIGURE 2 Relationship between published and predicted EC3 values of 28 chemicals not included in the training set (ANN models 1 and 2). Relationship between published and predicted EC3 values of chemicals not in the training set is shown for ANN model 1 (open squares) or model 2 (open circles), merged into the scatter plot of published and predicted EC3 values of the training set chemicals derived from the result of 10-fold cross-validation ($n = 3$) of ANN model 1 or 2, respectively. ANN, artificial neural network; LLNA, local lymph node assay

and RMS error (corresponding to prediction performance) were similar for all models (Table 4). However, when MIT in h-CLAT and CV75 were used as independent descriptors, and when the result of TIMES-M or Toxtree was incorporated (ANN models 4 and 6), the correlation with published LLNA EC3 was slightly better (Table 4). When we performed model evaluation using unlearned chemicals, we found that ANN models 4 and 6 did not show the anomalously low values seen with ANN models 1 and 2 (Figure 3A,B). On the other hand, ANN models 3 and 5 did not show marked improvements in this regard, compared to ANN models 4 and 6 (data not shown).

3.6 | Evaluation of predictive performance of artificial neural network models 1, 2, 4 and 6 based on classification and regression tree analysis

The LLNA EC3 values of chemicals predicted by the ANN models based on h-CLAT (MIT) and CV75/DPRA/KeratinoSens™/cell toxicity

(ANN model 1), h-CLAT/DPRA/KeratinoSens™/cell toxicity (ANN model 2), h-CLAT/DPRA/KeratinoSens™/cell toxicity/TIMES-M (ANN model 4) and h-CLAT/DPRA/KeratinoSens™/cell toxicity/Toxtree (ANN model 6) are summarized in Supporting information 3. These predicted EC3 values are the average values of 10-fold cross-validation ($n = 3$) performed in each ANN model. To judge a negative result of in vitro tests, corresponding to not sensitizing (NS) in LLNA, the “bottom-up 3 out of 3 criteria” hazard judgment model shown in Table 2 was applied. As shown in Table 5, the accuracy of prediction of potency class in four categories was 56.0% with ANN model 1 (Table 5A), while with ANN models 2, 4 and 6, accuracy was 53.0%, 67.9% and 59.0%, respectively (Table 5B–D).

Supporting information S4 shows the results of classification and regression tree (CART) analyses using chemicals that were not included in the training set. As shown in Table 6, the accuracy of prediction of potency class in four categories was 32.1% (Table 6A, condition 1) with ANN model 1, while with ANN models 2, 4 and 6, accuracy was 28.6%

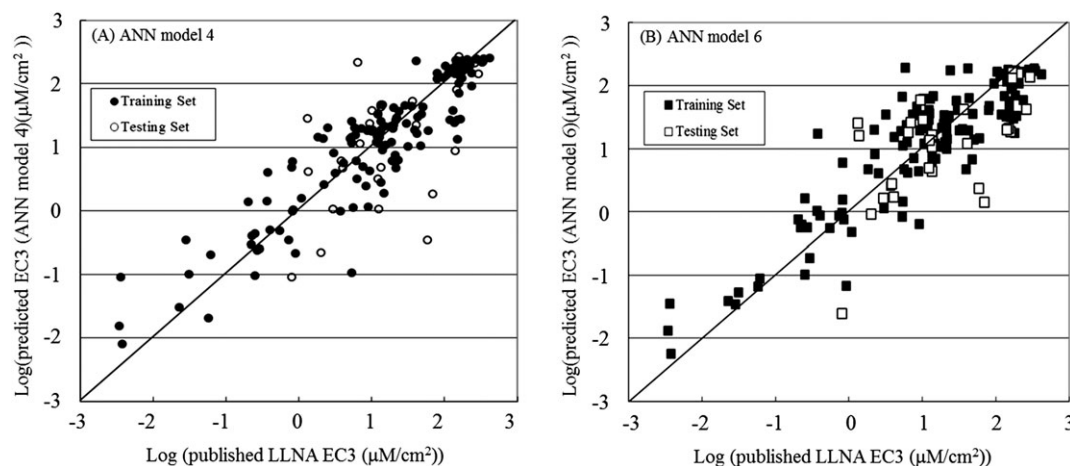


FIGURE 3 Relationship between published and predicted EC3 values of 28 chemicals not included in the training set (ANN models 4 and 6). Relationship between published and predicted EC3 values of the testing set chemicals not in the training set are shown in ANN model 4 (open circle dots) or model 6 (open square dots), merged into the scatter plot of published and predicted EC3 values of the training set chemicals derived from the result of 10-fold cross-validation ($n = 3$) of ANN model 4 or 6, respectively. ANN, artificial neural network; LLNA, local lymph node assay

TABLE 5 Correlations between LLNA EC3 classified into four categories and predicted LLNA classification obtained with ANN models 1, 2, 4 and 6 ($n = 134$)

	Potency category	LLNA classification (in vivo)			
		Extreme or strong (26)	Moderate (37)	Weak (31)	NS (40)
(A) ANN model 1 (Min(h-CLAT&cell toxicity)/ DPRA/ KeratinoSens™)	Extreme or strong	19	3	0	0
	Moderate	6	17	7	0
	Weak	1	15	23	24
	Negative, predicted EC3 > 100%	0	2	1	16
	Accuracy (%)	56.0			
(B) ANN model 2 (h-CLAT/ DPRA/ KeratinoSens™/ cell toxicity)	Overpredicted (%)	25.4			
	Underpredicted (%)	18.7			
	Extreme or strong	21	3	0	0
	Moderate	4	15	10	0
	Weak	1	17	19	24
(C) ANN model 4 (h-CLAT/ DPRA/ KeratinoSens/cell toxicity™/TIMES-M)	Negative, predicted EC3 > 100%	0	2	2	16
	Accuracy (%)	53.0			
	Overpredicted (%)	27.6			
	Underpredicted (%)	19.4			
	Extreme or strong	20	4	0	0
(D) ANN model 6 (h-CLAT/DPRA/ KeratinoSens/cell toxicity™/ Toxtree)	Moderate	6	25	9	1
	Weak	0	6	21	14
	Negative, predicted EC3 > 100%	0	2	1	25
	Accuracy (%)	67.9			
	Overpredicted (%)	20.9			
(E) ANN model 6 (h-CLAT/DPRA/ KeratinoSens/cell toxicity™/ Toxtree)	Underpredicted (%)	11.2			
	Extreme or strong	22	4	0	0
	Moderate	3	18	10	1
	Weak	1	13	20	20
	Negative, predicted EC3 > 100%	0	2	1	19
	Accuracy (%)	59.0			
	Overpredicted (%)	26.1			
	Underpredicted (%)	14.9			

NS, non-sensitizing in LLNA.

Classification based on published and predicted EC3 values was performed as described in Material and methods. Italic numbers indicate that the same results were obtained in LLNA classification (extreme or string, moderate, weak and NS) and predicted classification (strong, moderate, weak and negative) with ANN models 1, 2, 4 and 6. Calculated EC3 values in each ANN model are shown in Supporting information S3.

(Table 6B, condition 1), 46.4% (Table 6C, condition 1) and 39.3% (Table 6D, condition 1), respectively. On the other hand, when we considered chemicals for which the ratio of predicted/published EC3 lay within the arbitrary range of square root of 10 (3.16) and 1/10 (0.316), the accuracy of prediction of potency class in four categories was 42.9% (Table 6A, condition 2), 50.0% (Table 6B, condition 2), 71.4% (Table 6C, condition 2) and 64.3% (Table 6D, condition 2) in ANN models 1, 2, 4 and 6, respectively. As shown in Table 7, the accuracy of prediction of the GHS category was 89.6% (result of 10-fold cross-validation, $n = 134$)/64.3% (outer validation, $n = 28$) (Table 7A) with ANN model 1, while it was 91.8%/64.3% (Table 7B), 86.6%/78.6% (Table 7C) and 90.3%/82.1% (Table 7D) in ANN models 2, 4 and 6, respectively.

4 | DISCUSSION

In this work, we focused on ANN modeling to predict LLNA EC3 values using combinations of h-CLAT, DPRA and KeratinoSens™, which have been adopted as OECD test guidelines, with a database of >130 chemicals shared by the developers of these in vitro tests. First, we re-examined the accuracy of h-CLAT, DPRA and KeratinoSens™ and their combination from the viewpoint of hazard judgment

(positive/negative judgment). The results were similar to those previously reported, and indicated that our data set was not biased (Takenouchi et al., 2015; Urbisch et al., 2015).

Next, we re-evaluated the correlations among descriptors derived from in vitro tests and LLNA EC3 value. Each of the descriptors of these in vitro tests was significantly correlated with LLNA EC3. Interestingly, h-CLAT-related descriptors (MIT and minimum of MIT and CV75) were more highly correlated than descriptors of DPRA and KeratinoSens™, suggesting that h-CLAT-related descriptors might be more effective in the ANN model (Hirota et al., 2015). The correlation of reactivity of lysine peptide in DPRA showed the smallest r value. However, the percentage remaining lysine peptide might be important for certain skin sensitizers that have been reported to bind free amino groups in proteins, such as phthalic anhydride and trimellitic anhydride (Gerberick et al., 2004).

In contrast to previous work, which used a data set of about 60 chemicals (Hirota et al., 2015; Tsujita-Inoue et al., 2014), the present study used a larger data set of over 130 chemicals, and we considered that additional descriptors might improve the predictive performance. Thus, in DPRA, the results for cysteine peptide and lysine peptide were used as independent descriptors. It is noteworthy that ANN model 1 gave a good predictive performance before and after 10-fold cross-validation (Table 4). In this model, MIT of h-CLAT and CV75

TABLE 6 Correlations between LLNA EC3 classified into four categories and predicted LLNA classification obtained with ANN models 1, 2, 4 and 6 for chemicals not in the training set (n = 28)

Potency category	LLNA classification (in vivo)				Ns (8)
	Extreme or strong	Moderate	Weak	Condition 2	
(A) ANN model 1 (min(h-CLAT& cell toxicity)/DPRA/ KeratinoSens™)	Extreme or strong 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	0	2 (0)	3	1
	Moderate 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1 (0)	3 -	2 (1)	1
	Weak 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	2	5 (0)	2 -	2 (2)
	Negative, predicted EC3 > 100% 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	0	0 -	0 (0)	4 -
	Accuracy (%)	Condition 1	32.1	Condition 2	42.9
	Overpredicted (%)		39.3		28.6
	Underpredicted (%)		28.6		28.6
	Extreme or strong 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1	3 (1)	2	1
	Moderate 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1 (0)	2 -	3 (2)	1
	Weak 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1	5 (2)	1 -	2 (1)
(B) ANN model 2 (h-CLAT/DPRA/ KeratinoSens™/ cell toxicity)	Negative, predicted EC3 > 100% 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	0	0 -	1 (0)	4 -
	Accuracy (%)	Condition 1	28.6	Condition 2	50.0
	Overpredicted (%)		42.9		28.6
	Underpredicted (%)		28.6		21.4
	Extreme or strong 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1	3 (1)	2	0
	Moderate 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1 (1)	4 -	1 (1)	1
	Weak 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1	2 (1)	4 -	3 (3)
	Negative, predicted EC3 > 100% 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	0	1 -	0 (0)	4 -
	Accuracy (%)	Condition 1	46.4	Condition 2	71.4
	Overpredicted (%)		35.7		17.9
	Underpredicted (%)		17.9		10.7
(C) ANN model 4 (h-CLAT/DPRA/ KeratinoSens™/ cell toxicity/TIMES-M)	Extreme or strong 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1	3 (1)	2	0
	Moderate 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1 (1)	4 -	1 (1)	1
	Weak 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1	2 (1)	4 -	3 (3)
	Negative, predicted EC3 > 100% 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	0	1 -	0 (0)	4 -
	Accuracy (%)	Condition 1	46.4	Condition 2	71.4
	Overpredicted (%)		35.7		17.9
	Underpredicted (%)		17.9		10.7
	Extreme or strong 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1	3 (1)	2	0
	Moderate 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1 (1)	4 -	1 (1)	1
	Weak 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1	2 (1)	4 -	3 (3)
	Negative, predicted EC3 > 100% 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	0	1 -	0 (0)	4 -

(Continues)

TABLE 6 (Continued)

	Potency category	LLNA classification (in vivo)			
		Extreme or strong (3)	Moderate (10)	Weak (7)	Ns (8)
(D) ANN model 6 (h-CLAT/DPRA/ KeratinoSens™/ cell toxicity/Toxtree)	Extreme or strong 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1	2 (2)	2	0
	Moderate 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1 (0)	4 –	3 (2)	0
	Weak 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	1 –	4 (2)	2 –	4 (1)
	Negative, predicted EC3 > 100% 0.316 × EC3 (pub) ≤ predicted EC3 ≤ 3.16 × EC3 (pub)	0 –	0 –	0 (0)	4 –
	Accuracy (%)	Condition 1	39.3	Condition 2	64.3
	Overpredicted (%)		39.3		21.4
	Underpredicted (%)		21.4		14.3

NS, non-sensitizing in LLNA; pub, published. EC3(pub); published EC3 value.

Classification based on published and predicted EC3 values was performed as described in Materials and methods. Italic numbers indicate that the same results were obtained in LLNA classification (extreme or strong, moderate, weak and NS) and predicted classification (strong, moderate, weak and negative) with ANN models 1, 2, 4, and 6. Bold numbers indicate chemicals for which the ratio of predicted/published EC3 was within the range of the square root of 10 (3.16) and 1/10 (0.316), even though the results in LLNA classification and predicted classification were different. Calculated EC3 data in each ANN model are shown in Supporting information S4.

were combined as a descriptor (minimum of MIT and CV75) because of the high mutual correlation between the two parameters (Hirota et al., 2013). However, in validation using 28 chemicals not in the training set, this model overestimated LLNA EC3 for several chemicals (3-chloro-p-anisaldehyde [in vivo LLNA; NS], ethyl acrylate [in vivo LLNA EC3 28%], methyl acrylate [in vivo LLNA EC3 20%], 2-methylundecanal [in vivo LLNA EC3 10%] and safranal [in vivo LLNA EC3 7.5%]) (Figure 2A). We thought this might be because the number of parameters (four descriptors) was insufficient, and indeed, when MIT of h-CLAT and CV75 were used as independent descriptors (ANN model 2, 5 descriptors), the overestimation was partly eliminated (Figure 2B). Furthermore, when we included the result of TMES-M or Toxtree, the *r* values and RMS errors were better than in other ANN models, as seen with ANN model 4 (descriptors; h-CLAT/DPRA/KeratinoSens™/cell toxicity (CV75)/TIMES-M) and ANN model 6 (descriptors; h-CLAT/DPRA/KeratinoSens™/cell toxicity (CV75)/Toxtree). These results were also reflected in the CART analyses of ANN models. Overall, these results indicate that addition of novel descriptors of chemical structure information contributed to the improvement of *r* value and RMS error of the ANN models.

In the skin sensitization adverse outcome pathway (OECD, 2012), the initial molecular event is associated with electrophilic character. In connection with this, DPRA detects hapten-protein binding, while in silico models such as TIMES-M or structure alert-detecting software such as Toxtree might also be focused on the molecular initial event in terms of electrophilicity.

On the other hand, in the outer validation using 28 chemicals not in the training set, the predictive performance of ANN models was different from that in a 10-fold cross-validation. The correlations between published EC3 values and predicted EC3 values derived from ANN models 4 and 6 in the scatter plot analyses were better than the cases of ANN models 1 and 2. However, the accuracy in CART analysis was not better than the result of cross-validation. In general, the aim of CART analysis is to classify skin sensitization potency (e.g., strong, moderate, weak). On the other hand, there were cases where the EC3 values of chemicals were similar, but the chemicals were classified into different skin sensitization categories. Furthermore, the EC3 values of some sensitizers in the 28 outer validation samples were close to 10% (cut-off value determining weak or moderate sensitizers) or 1% (cut-off value determining moderate or strong sensitizers). Therefore, we took into account the ratio of published and predicted EC3s in CART analysis. When this ratio was in the arbitrary range of the square root of 10 (3.16) and 1/10 (0.316), the accuracy of prediction of potency class in the four categories was improved in ANN models 1, 2, 4 and 6 (Table 6, condition 2). For ANN models 4 and 6, the accuracy of CART analysis considering predicted/published LLNA EC3 ratio was over 60%, and the predictive performance of classification GHS 1A/1B was about 80% in both data sets (Table 7). Overall, it seems clear that additional descriptors derived from in silico or structure alert software were useful to improve the predictive performance of the ANN model as regards LLNA EC3. On the other hand, it is also important to consider hazard judgment. In this connection, the bottom up 3 out of 3 judgment was reported to show high sensitivity (less false negatives), while the bottom up 2 out of 3 judgment was better correlated with human skin sensitization (Urbisch et al., 2015).

TABLE 7 Correlations between GHS categories based on LLNA and ANN models 1, 2, 4 and 6

	Potency category	LLNA GHS classification (in vivo)			
		DB; Supporting information S3		DB; Supporting information S4	
		1A (34)	1B (100)	1A (6)	1B (22)
(A) ANN model 1 (min (h-CLAT& cell toxicity)/DPRA/ KeratinoSens™)	1A	25	5	1	5
	1B	9	95	5	17
	Accuracy (%)	89.6		64.3	
	Overpredicted (%)	3.7		17.9	
	Underpredicted (%)	6.7		17.9	
(B) ANN model 2 (h-CLAT/DPRA/ KeratinoSens™/cell toxicity)	1A	26	3	2	6
	1B	8	97	4	16
	Accuracy (%)	91.8		64.3	
	Overpredicted (%)	2.2		21.4	
	Underpredicted (%)	6.0		14.3	
(C) ANN model 4 (h-CLAT/DPRA/ KeratinoSens™/cell toxicity/TIMES-M)	1A	24	8	3	3
	1B	10	92	3	19
	Accuracy (%)	86.6		78.6	
	Overpredicted (%)	6.0		10.7	
	Underpredicted (%)	7.5		10.7	
(D) ANN model 6 (h-CLAT/DPRA/ KeratinoSens™/cell toxicity/Toxtree)	1A	25	4	4	3
	1B	9	96	2	19
	Accuracy (%)	90.3		82.1	
	Overpredicted (%)	3.0		10.7	
	Underpredicted (%)	6.7		7.1	

Italic numbers indicate that the same results were obtained in LLNA classification and predicted classification in ANN models 1, 2, 4 and 6 (Supporting information S3 and S4).

However, there was little difference between the two criteria in the accuracy of judgment or the predictive performance of our ANN models (data not shown).

The adverse outcome pathway for skin sensitization includes protein-binding, dendritic cell activation, keratinocyte signaling and T-cell proliferation (OECD, 2012). As our ANN model does not include an assay reflecting T-cell proliferation, it is interesting to speculate whether incorporation of T-cell proliferation step as an additional descriptor might improve our model. However, a T-cell proliferation assay is not a high-throughput assay like h-CLAT, DPRA and KeratinoSens™, because it requires the use of human peripheral blood (Richter et al., 2013; Vocanson et al., 2014). Therefore, we have not included it at the present stage.

Recently, a BN model including h-CLAT, DPRA and KeratinoSens™ was developed for predicting LLNA potency category (Jaworska et al., 2015); its predictive performance was 96% (GHS potency classification) and 89% (potency [four classes]). Though the predictive performance of our ANN model for the skin sensitization category was weaker than that of the published BN model, the ANN model in this study can directly predict EC3 value. In principle, both ANN and BN can be used as prediction models for both EC3 and skin sensitization category, and EC3 prediction using BN was recently investigated (unpublished data). Our model and the published BN model are constructed from almost the same data set, except for bioavailability related parameters. Interestingly, though, we preliminarily confirmed that bioavailability-related parameters did not improve *r* values and RMS error in our model, compared with the use of TIMES-M or Toxtree (data not shown). Therefore, a combination of ANN and BN models might be useful for comprehensive assessment of cosmetic ingredients. We previously demonstrated that the combination of an

ANN-based QSAR model and ANN-based in vitro model was useful for decreasing the number of underestimated results (Tsujita-Inoue et al., 2015).

As a next step to improve the predictive performance and robustness of our ANN model, it would be desirable to increase the size of the learning data set. As the number of chemicals evaluated using LLNA is over 300 (Gerberick et al., 2005; Kern et al., 2010), further work should be done to examine more of these chemicals in the three in vitro tests, h-CLAT, DPRA and KeratinoSens™. Nevertheless, we believe the present findings will be helpful for further developing non-animal safety assessment tools for evaluating skin sensitization potency of chemicals based on in silico (QSAR and Read Across) and in vitro evaluations.

ACKNOWLEDGEMENTS

We thank Drs. Joanna Jaworska, Petra Kern, Cindy A. Ryan and Frank Gerberick from Procter and Gamble Company for sharing DPRA and TIMES-M data, Dr. Andreas Natsch from Givaudan Schweiz AG for sharing KeratinoSens™ data, Drs. Masaaki Miyazawa and Hitoshi Sakaguchi from Kao Corporation for generating new h-CLAT data in collaboration with Shiseido Co., and Richard Steele for a critical review of the manuscript.

CONFLICT OF INTEREST

None of the authors have any conflicts of interest associated with this study.

ORCID

Morihiko Hirota  <http://orcid.org/0000-0002-4110-5015>

REFERENCES

- Aiba, S., & Tagami, H. (1998). Dendritic cell activation induced by various stimuli, e.g. exposure to microorganisms, their products, cytokines, and simple chemicals as well as adhesion to extracellular matrix. *Journal of Dermatological Science*, 20, 1–13.
- Aiba, S., Terunuma, A., Manome, H., & Tagami, H. (1997). Dendritic cells differently respond to haptens and irritants by their production of cytokines and expression of co-stimulatory molecules. *European Journal of Immunology*, 27, 3031–3038.
- Api, A. M., Basketter, D. A., Cadby, P. A., Cano, M. F., Ellis, G., Gerberick, G. F., ... Safford, R. (2008). Dermal sensitization quantitative risk assessment (QRA) for fragrance ingredients. *Regulatory Toxicology and Pharmacology*, 52, 3–23.
- Ashikaga, T., Yoshida, Y., Hirota, M., Yoneyama, K., Itagaki, H., Sakaguchi, H., ... 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. *Toxicology In Vitro*, 20, 767–773.
- Basketter, D. A., Clapp, C. J., Safford, B. J., Jowsey, I. R., McNamee, P., Ryan, C. A., & Gerberick, G. F. (2008). Preservatives and skin sensitization quantitative risk assessment. *Dermatitis*, 19, 20–27.
- Basketter, D. A., Evans, P., Fielder, R. J., Gerberick, G. F., Dearman, R. J., & Kimber, I. (2002). Local lymph node assay – Validation, conduct and use in practice. *Food and Chemical Toxicology*, 40, 593–598.
- Boislevé, F., Kerdine-Romer, S., Rougier-Larzat, N., & Pallardy, M. (2004). Nickel and DNCB induce CCR7 expression on human dendritic cells through different signalling pathways: Role of TNF- α and MAPK. *Journal of Investigative Dermatology*, 123, 494–502.
- Dimitrov, S. D., Low, L. K., Patlewicz, G. Y., Kern, P. S., Dimitrova, G. D., Comber, M. H., ... Mekenyan, O. G. (2005). Skin sensitization: Modeling based on skin metabolism simulation and formation of protein conjugates. *International Journal of Toxicology*, 24, 189–204.
- Emter, R., Ellis, G., & Natsch, A. (2010). Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. *Toxicology and Applied Pharmacology*, 245, 281–290.
- Enoch, S. J., Madden, J. C., & Cronin, M. T. (2008). Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR & QSAR in Environmental Research*, 19, 555–578.
- Gerberick, G. F., Ryan, C. A., Kern, P. S., Schlatter, H., Dearman, R. J., Kimber, I., ... Basketter, D. A. (2005). Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis*, 16, 157–202.
- Gerberick, G. F., Vassallo, J. D., Bailey, R. E., Chaney, J. E., Morrall, S. W., & Lepoittevin, J. P. (2004). Development of a peptide reactivity assay for screening contact allergens. *Toxicological Sciences*, 81, 332–343.
- Gerberick, G. F., Vassallo, J. D., Foertsch, L. M., Price, B. B., Chaney, J. G., & Lepoittevin, J. P. (2007). Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. *Toxicological Sciences*, 97, 417–427.
- Hirota, M., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., Fujishiro, M., ... Miyazawa, M. (2015). Evaluation of combinations of in vitro sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization. *Journal of Applied Toxicology*, 35, 1333–1347.
- Hirota, M., Kouzuki, H., Ashikaga, T., Sono, S., Tsujita, K., Sasa, H., & Aiba, S. (2013). Artificial neural network analysis of data from multiple in vitro assays for prediction of skin sensitization potency of chemicals. *Toxicology In Vitro*, 27, 1233–1246.
- Jaworska, J., Dancik, Y., Kern, P. S., Gerberick, G. F., & Natsch, A. (2013). Bayesian integrated testing strategy to assess skin sensitization potency: From theory to practice. *Journal of Applied Toxicology*, 33, 1353–1364.
- Jaworska, J., Harol, A., Kern, P. S., & Gerberick, G. F. (2011). Integrating non-animal test information into an adaptive testing strategy – Skin sensitization proof of concept case. *ALTEX*, 28, 211–225.
- Jaworska, J., Natsch, A., Ryan, C., Strickland, J., Ashikaga, T., & Miyazawa, M. (2015). Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: A decision support system for quantitative weight of evidence and adaptive testing strategy. *Archives of Toxicology*, 89, 2355–2383.
- Johansson, H., Rydnert, F., Kuhn, J., Schepky, A., Borrebaek, C., & Lindstedt, M. (2014). Genomic allergen rapid detection in-house validation – A proof of concept. *Toxicological Sciences*, 139, 362–370.
- Jowsey, I. R., Basketter, D. A., Westmoreland, C., & Kimber, I. (2006). A future approach to measuring relative skin sensitizing potency: A proposal. *Journal of Applied Toxicology*, 26, 341–350.
- Kern, P. S., Gerberick, G. F., Ryan, C. A., Kimber, I., Aptula, A., & Basketter, D. A. (2010). Local lymph node data for the evaluation of skin sensitization alternatives: A second compilation. *Dermatitis*, 21, 8–32.
- Kimura, Y., Fujimura, C., Ito, Y., Takahashi, T., Nakajima, Y., Ohmiya, Y., & Aiba, A. (2015). Optimization of the IL-8 Luc assay as an in vitro test for skin sensitization. *Toxicology In Vitro*, 29, 1816–1830.
- Magnusson, B., & Kligman, A. M. (1969). The identification of contact allergens by animal assay. The guinea pig maximization test. *Journal of Investigative Dermatology*, 52, 268–276.
- Mossmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *Journal of Immunological Methods*, 65, 55–63.
- Natsch, A., & Emter, R. (2008). Skin sensitizers induce antioxidant response element dependent genes: Application to the in vitro testing of the sensitization potential of chemicals. *Toxicological Sciences*, 102, 110–119.
- Natsch, A., Emter, R., & Ellis, G. (2009). Filling the concept with data: Integrating data from different in vitro and in silico assays on skin sensitizers to explore the battery approach for animal-free skin sensitization testing. *Toxicological Sciences*, 107, 106–121.
- Natsch, A., & Gfeller, H. (2008). LC-MS-based characterization of the peptide reactivity of chemicals to improve the in vitro prediction of the skin sensitization potential. *Toxicological Sciences*, 106, 464–478.
- Natsch, A., Ryan, C. A., Foertsch, L., Emter, R., Jaworska, J., Gerberick, F., & Kern, P. (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *Journal of Applied Toxicology*, 33, 1337–1352.
- Nukada, Y., Ashikaga, T., Miyazawa, M., Hirota, M., Sakaguchi, H., Sasa, H., & Nishiyama, N. (2012). Prediction of skin sensitization potency of chemicals by human cell line activation test (h-CLAT) and an attempt at classifying skin sensitization potency. *Toxicology In Vitro*, 26, 1150–1160.
- Nukada, Y., Miyazawa, M., Saitou, K., Sakaguchi, H., & Nishiyama, N. (2013). Data integration of non-animal tests for the development of a test battery to predict the skin sensitizing potential and potency of chemicals. *Toxicology In Vitro*, 27, 609–618.
- OECD (1992). *OECD Guideline for testing chemical, Section 4: Health Effects. Guideline no. 406. skin sensitization*. Paris: OECD Publishing.
- OECD (2010). *OECD Guideline for testing chemicals, Section 4: Health Effects. Guideline no. 429. skin sensitization: the Local lymph node assay*. Paris: OECD Publishing.
- OECD (2012). *The adverse outcome pathway for skin sensitisation initiated by covalent binding for proteins. Part 1: Scientific evidence. OECD Environment, Health and Safety Publications Series on Testing and Assessment 168*. Paris: OECD Publishing.
- OECD (2015a). *OECD Guideline for testing chemicals, Section 4: Health Effects. Guideline no. 442C. In Chemico Skin Sensitization: Direct Peptide Reactivity Assay (DPRA)*. Paris: OECD Publishing.
- OECD (2015b). *OECD Guideline for testing chemicals, Section 4: Health Effects. Guideline no. 442D. In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method*. Paris: OECD Publishing.
- OECD (2016a). *OECD Guideline for testing chemicals, Section 4: Health Effects. Guideline no. 442E. In Vitro Skin Sensitisation: human Cell Line Activation Test (h-CLAT)*. Paris: OECD Publishing.

- OECD (2016b). 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. Series on Testing & Assessment No. 256. [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO\(2016\)29&doclanguage=Den](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2016)29&doclanguage=Den)
- Piroird, C., Ovigine, J. M., Rousset, F., Martinozzi-Teissier, S., Gomes, C., Cotovio, J., & Alepee, N. (2015). The myeloid U937 skin sensitization test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization. *Toxicology In Vitro*, 29, 901–916.
- Richter, A., Schmucker, S. S., Esser, P. R., Traska, V., Weber, V., Dietz, L., ... Martin, S. F. (2013). Human T cell priming assay (hTCPA) for the identification of contact allergens based on naive T cells and DC-IFN- γ and TNF- α readout. *Toxicology In Vitro*, 27, 1180–1185.
- Sakaguchi, H., Ashikaga, T., Miyazawa, M., Yoshida, Y., Ito, Y., Yoneyama, K., ... 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. *Toxicology In Vitro*, 20, 774–784.
- Suzuki, M., Hirota, M., Hagino, S., Itagaki, H., & Aiba, S. (2009). Evaluation of changes of cell-surface thiols as a new biomarker for in vitro sensitization test. *Toxicology In Vitro*, 23, 687–696.
- Takenouchi, O., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., Fujishiro, M., ... Miyazawa, M. (2015). Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. *Journal of Applied Toxicology*, 35, 1318–1332.
- Tsujita-Inoue, K., Atobe, T., Hirota, M., Ashikaga, T., & Kouzuki, H. (2015). In silico risk assessment for skin sensitization using artificial neural network analysis. *Journal of Toxicological Sciences*, 40, 193–209.
- Tsujita-Inoue, K., Hirota, M., Ashikaga, T., Atobe, T., Kouzuki, H., & Aiba, S. (2014). Skin sensitization risk assessment model using artificial neural network analysis of data from multiple in vitro assays. *Toxicology In Vitro*, 28, 626–639.
- United Nations (2009). *Globally Harmonized System of Classification and Labeling of Chemicals (GHS)* ((3rd revised ed.). New York: United Nations.
- Urbisch, D., Mehling, A., Guth, K., Ramirez, T., Honarvar, N., Kolle, S., ... Sakaguchi, H. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regulatory Toxicology and Pharmacology*, 71, 337–351.
- Vocanson, M., Achachi, A., Mutez, V., Cluzel-Tailhardat, M., Varlet, B. L., Rozières, A., ... Nicolas, J. F. (2014). Human T cell priming assay: Depletion of peripheral blood lymphocytes in CD25⁺ cells improves the in vitro detection of weak allergen-specific T cells. *EXS*, 104, 89–100.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Hirota M, Ashikaga T, Kouzuki H. Development of an artificial neural network model for risk assessment of skin sensitization using human cell line activation test, direct peptide reactivity assay, KeratinoSens™ and in silico structure alert parameter. *J Appl Toxicol*. 2018;38: 514–526. <https://doi.org/10.1002/jat.3558>