HHS Public Access

Author manuscript

J Appl Toxicol. Author manuscript; available in PMC 2017 September 01.

Published in final edited form as:

J Appl Toxicol. 2016 September ; 36(9): 1150–1162. doi:10.1002/jat.3281.

Integrated Decision Strategies for Skin Sensitization Hazard

Judy Strickland[1], **Qingda Zang**[1], **Nicole Kleinstreuer**[1], **Michael Paris**[1], **David M. Lehmann**[2], **Neepa Choksi**[1], **Joanna Matheson**[3], **Abigail Jacobs**[4], **Anna Lowit**[5], **David Allen**[1], and **Warren Casey**[6]

[1]ILS, Research Triangle Park, NC 27709, USA

[2]EPA/NHEERL/EPHD/CIB, Research Triangle Park, NC 27709, USA

[3]U.S. Consumer Product Safety Commission, Bethesda, MD 20814, USA

[4]FDA/CDER, Silver Spring, MD 20993, USA

[5]EPA/OCSPP/OPP/HED, Washington, DC 20460, USA

[6]NIH/NIEHS/DNTP/NICEATM, Research Triangle Park, NC 27709, USA

Abstract

One of the top priorities of ICCVAM is the identification and evaluation of non-animal alternatives for skin sensitization testing. Although skin sensitization is a complex process, the key biological events of the process have been well characterized in an adverse outcome pathway (AOP) proposed by OECD. Accordingly, ICCVAM is working to develop integrated decision strategies based on the AOP using *in vitro*, *in chemico*, and *in silico* information. Data were compiled for 120 substances tested in the murine local lymph node assay (LLNA), direct peptide reactivity assay (DPRA), human cell line activation test (h-CLAT), and KeratinoSens assay. Data for six physicochemical properties that may affect skin penetration were also collected, and skin sensitization read-across predictions were performed using OECD QSAR Toolbox. All data were combined into a variety of potential integrated decision strategies to predict LLNA outcomes using a training set of 94 substances and an external test set of 26 substances. Fifty-four models were built using multiple combinations of machine learning approaches and predictor variables. The seven models with the highest accuracy (89–96% for the test set and 96–99% for the training set) for predicting LLNA outcomes used a support vector machine (SVM) approach with different combinations of predictor variables. The performance statistics of the SVM models were higher than any of the non-animal tests alone and higher than simple test battery approaches using these methods. These data suggest that computational approaches are promising tools to effectively integrate data sources to identify potential skin sensitizers without animal testing.

Corresponding author: warren.casey@nih.gov.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Keywords

Skin sensitization; allergic contact dermatitis; integrated decision strategy; machine learning; LLNA; DPRA; KeratinoSens; h-CLAT; support vector machine

Introduction

Skin sensitizers are substances capable of causing allergic contact dermatitis (ACD), a local skin reaction characterized by redness, swelling and itching (Murphy *et al.,* 2012). Like all allergic diseases, skin sensitization develops in two stages: induction and elicitation. In the induction phase, allergen-specific T cells are generated, a process that typically does not produce clinical symptoms. The elicitation phase and accompanying allergic response occurs when a previously sensitized individual is re-exposed to the inducing allergen. In ACD, the allergic response produces a pruritic rash. The latency period between exposure and appearance of the rash shortens with subsequent exposures (Murphy *et al.,* 2012). Sensitization may persist for a lifetime, and ACD can significantly impact quality of life, contributing to its importance as an occupational and environmental health issue (Kimber *et al.,* 2011). Ten to 15 percent of all U.S. occupational diseases result from this allergic condition, making it the second most commonly reported occupational illness (Anderson *et al.,* 2011). When considering the general population, 15–20% of individuals will be sensitized to at least one allergen during their lifetime (Bruckner *et al.,* 2000; Thyssen *et al.,* 2007). In order to minimize the occurrence of such reactions, skin sensitization testing is routinely performed on chemical products to meet various national and international regulatory requirements for chemical management.

In the United States, skin sensitization testing requirements vary across federal regulatory agencies (Birnbaum 2013). The U.S. Food and Drug Administration (FDA) accepts preclinical skin sensitization data but has no requirements for specific tests (A. Jacobs, personal communication). However, the U.S. Environmental Protection Agency (EPA) (40 CFR 158.500 ; 40 CFR 161.340), U.S. Consumer Product Safety Commission (16 CFR 1500.3), and U.S. Occupational Safety and Health Administration (OSHA) (29 CFR 1910.1200) do require skin sensitization data to support product hazard labeling and registration (16 CFR 1500.3; 29 CFR 1910.1200; 40 CFR 158.500; 40 CFR 161.340). Currently, the skin sensitization tests needed to meet the requirements of these agencies involve obligatory use of animal models (AltTox 2014).

Public opinion, advances in scientific knowledge, and recent political pressure have made the use of animals for testing unsustainable in some regions. The European Union's 7th Amendment to the Cosmetic Directive required phasing out animal testing for all cosmetics ingredients, with a complete ban in place by March of 2013, prompting the development and use of non-animal testing methods (European Union 2003). Additionally, the European Regulation on Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), which went into effect in 2007 and will be implemented in phases until 2018, discourages animal testing (EC 2006; Joint Research Centre of the European Union 2013). Depending on the amount of a substance manufactured or imported into the European

Union, REACH requires a progressive panel of toxicological tests (Williams *et al.,* 2009). Although REACH mandates that animal tests only be conducted as a last resort, there are no alternative, non-animal testing methods presently available for many of the toxicological endpoints evaluated to meet the requirements of REACH. Specifically, there are currently no stand-alone non-animal methods for identifying skin sensitizers (Natsch 2014). Based on the current REACH guidelines, more substances will be tested for skin sensitization than for any other toxicological endpoint (Roberts and Aptula 2008; Schoeters 2010). In fact, more than 3700 substances were evaluated for skin sensitization during Phase 1 of REACH, and thousands more are expected to require testing to satisfy requirements for Phase 3 (Angers-Loustau *et al.,* 2011).

Since its inception, the U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) has been dedicated to the 3Rs: replacing, reducing, and refining the use of animals for toxicity testing (Birnbaum 2013). ICCVAM is working on the identification and validation of non-animal alternatives for skin sensitization testing (NIEHS 2013) and places this effort among its top priorities. Currently, the murine local lymph node assay (LLNA) is the preferred method for evaluating the sensitization potential of chemical substances to satisfy regulatory requirements, including REACH (Anderson *et al.,* 2011; Williams *et al.,* 2009; Williams *et al.,* 2015). Compared to previously used guinea pig tests, the LLNA uses fewer animals and causes less pain and distress (Sailstad *et al.,* 2001; Williams *et al.,* 2015). Thus, this extensively used assay represents an early 3Rs alternative for detection of this toxicological endpoint (Russell and Burch 1992). Since the development of the LLNA, additional research on the mechanisms leading to the development and manifestation of skin sensitization has enabled the construction of an adverse outcome pathway (AOP; Figure 1), which will promote further progress toward decreased dependence on animal testing (Roberts and Aptula 2008; Urbisch *et al.,* 2015).

An AOP is a conceptual framework constructed from existing knowledge that relates exposure of a type of toxic substance to subsequent molecular and cellular changes that in turn result in illness or injury to an individual or population (OECD 2012a; OECD 2012b). The AOP for skin sensitization initiated by covalent binding to proteins (Figure 1) includes four key events with well-accepted biological significance: 1) binding of haptens to endogenous proteins in the skin, 2) keratinocyte activation, 3) dendritic cell activation, and 4) proliferation of antigen-specific T cells (OECD 2012b). The construction of this AOP for skin sensitization has prompted the development of several *in vitro* tests targeting different key events (reviewed in Mehling *et al.* (2012) and evaluated in, e.g., Reisinger *et al.* (2015). While some individual methods have proven particularly promising for the prediction of skin sensitization potential (Gerberick *et al.,* 2004; Natsch and Emter 2008; Nukada *et al.,* 2012), each method has its own limitations when used in isolation. Given the inherent complexity of the processes underlying skin sensitization, it is unlikely that any single non-animal test can replace animal use for hazard identification (Rovida *et al.,* 2015). A more realistic approach involves combining data from several non-animal methods using an integrated decision strategy (IDS) (MacKay *et al.,* 2013).

An IDS incorporates all of the available and pertinent information about a test substance to arrive at a conclusion regarding a potential hazard. Combining outputs from several data

sources minimizes the limitations of each individual assay, and the predictive power of the combination of methods may be increased compared to stand-alone tests (Bauch *et al.,* 2012; Guyard-Nicodeme *et al.,* 2015; Jaworska *et al.,* 2013; Natsch *et al.,* 2009; Natsch *et al.,* 2013; Nukada *et al.,* 2013; Takenouchi *et al.,* 2015; Tsujita-Inoue *et al.,* 2014; Urbisch *et al.,* 2015; van der Veen *et al.,* 2014). In the case of skin sensitization, several *in vitro* tests targeting key events in the AOP are already available (i.e., human cell line activation test (h-CLAT) (OECD 2015a); direct peptide reactivity assay (DPRA) (OECD 2015b) and KeratinoSens (OECD 2015c). In alignment with ICCVAM's commitment to advancing implementation of 3Rs-compliant methods, the objective of this study was to develop an IDS based on the skin sensitization AOP. The IDS presented here incorporates *in vitro*, *in chemico*, and *in silico* information on skin sensitization to predict skin sensitization hazard using machine learning approaches and the LLNA as the reference test.

Materials and Methods

Data Collection and Substance Database

We compiled a substance database by collecting publically available data for the DPRA, KeratinoSens, the h-CLAT, and the LLNA (Table 1). DPRA, KeratinoSens, and h-CLAT were selected because they had recently been evaluated and recommended by the European Union Reference Laboratory for Alternatives to Animal Testing as methods to be used for hazard classification of sensitizers in a weight of evidence approach (Joint Research Centre of the European Union 2013; Joint Research Centre of the European Union 2014; Joint Research Centre of the European Union 2015) and they were being considered for new chemical test guidelines by the Organisation for Economic Co-operation and Development (OECD). While test guidelines for DPRA (OECD 2015b) and KeratinoSens (OECD 2015c) have been finalized, the h-CLAT test guideline is still in draft form (OECD 2015a).

The majority of the LLNA data for the 120-substance dataset were collected previously by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (http://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/immunotoxicity/nonanimal/index.html#NICEATM-LLNA-Database). These data include sensitizer/nonsensitizer determinations for each substance as well as stimulation indices at the concentrations tested. The LLNA data for seven substances that were not in this database came from published literature (Table 1).

In total, we identified 122 substances tested in DPRA, KeratinoSens, h-CLAT, and the LLNA. Two metal compounds (nickel chloride and cobalt chloride) were excluded because the LLNA often produces conflicting results for metals (OECD 2010). Nickel and cobalt produce skin sensitization in humans by activating the Toll-like receptor 4 protein, which, in mice, is ineffective at binding these metals (reviewed in Schmidt and Goebeler 2015). Also, metals are outside the applicability domain of the DPRA because they do not react with proteins by covalent binding (OECD 2015b). For the remaining 120 substances, we collected data on six physicochemical properties relevant to skin exposure and penetration: octanol:water partition coefficient, water solubility, vapor pressure, melting point, boiling point, and molecular weight. These properties have been important for other models or weight-of-evidence assessments for skin sensitization potential (Jaworska *et al.,* 2013;

Jaworska *et al.,* 2011; Patlewicz *et al.,* 2014) We also performed an *in silico* prediction of skin sensitization hazard using the read-across algorithm in QSAR Toolbox v3.2 (OECD 2014).

Characterization of the Substances

Of the 120 substances, 73% (87/120) were classified as positive in the LLNA and 27% (33/120) were classified as negative. Skin sensitizers may require oxidation (prehapten) and/or metabolism (prohapten) in order to produce a skin sensitization reaction. Of the 87 LLNA sensitizers, three were prehaptens, 16 were prohaptens, and six were pre/prohaptens (i.e., require both oxidation and metabolism) (see Supplemental File 1 for the prehapten and prohapten information on each substance and the corresponding reference).

The 120 substances represent 14 product categories, as shown in Figure 2. Product category information was obtained from the following sources:

- Hazardous Substances Databank (http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB)
- Haz-Map (http://hazmap.nlm.nih.gov/)
- Household Products Database (http://hpd.nlm.nih.gov/index.htm)
- International Programme on Chemical Safety INCHEM database (http://www.inchem.org/)
- National Library of Medicine Drug Information Portal (http://druginfo.nlm.nih.gov/drugportal/drugportal.jsp?APPLICATION_NAME=drugportal)
- National Toxicology Program (http://ntp.niehs.nih.gov/)
- EPA's list of registered pesticides (A Lowit, personal communication)
- The Joint Food and Agriculture Organization of the United Nations/World Health Organization (WHO) Expert Committee on Food Additives
- The Good Scents Company (http://www.thegoodscentscompany.com/)
- Scientific literature (i.e., papers which also presented test method data)
- Chemical Book (http://www.chemicalbook.com)

Structural variety of the database was assessed using ChemoTyper v1.0 (https://chemotyper.org/), a free software developed under contract with the FDA. ChemoTyper uses 729 chemotypes, which are generic structural fragments that represent chemical features, including connected and nonconnected chemical patterns as well as atom, bond, and molecular-based properties (Yang *et al.,* 2015). The 120 substances in the database represented 192 chemotypes that had a frequency of appearance of 1 to 75 over the entire dataset (Figure 3). The most common chemotypes were bond:C=O_carbonyl_generic (75 substances), ring:aromatic_benzene (68 substances), chain:alkaneLinear_ethyl_C2(H_gt_1) (43 substances), chain:aromaticAlkane_Ph-C1_acyclic_generic (42 substances), and

chain:alkaneLinear_ethyl_C2_(connect_noZ_CN=4) (36 substances). Individual substances were characterized by 2–35 chemotypes each.

Data Inputs

DPRA—DPRA is an *in chemico* test that assesses the ability of a substance to form a hapten–protein complex, which is the molecular initiating event in the skin sensitization AOP (Figure 1) (OECD 2012a; OECD 2012b). It measures the reactivity of a test substance towards two model synthetic peptides, one containing lysine (mixed 1:50 with test substance) and the other containing cysteine (mixed 1:10 with test substance) (Gerberick *et al.,* 2004; Gerberick *et al.,* 2007; OECD 2015b). The depletion of the peptides after a 24 h incubation with a test substance is measured using high pressure liquid chromatography. The percent depletion values for the two peptides are averaged; substances are classified as sensitizers if average depletion >6.38% (OECD 2015b). Substances that co-elute with the lysine peptide may be evaluated based upon cysteine peptide depletion alone using >13.89% depletion as the threshold to classify a substance as a sensitizer (OECD 2015b). Data used from DPRA included average cysteine peptide depletion (Cys), average lysine peptide depletion (Lys), average depletion of cysteine and lysine peptides (Avg.Lys.Cys), and sensitizer/nonsensitizer outcome based on the above decision criteria.

KeratinoSens—The KeratinoSens test method assesses the ability of substances to activate cytokines and induce cytoprotective genes in keratinocytes, the second key event in the skin sensitization AOP (Figure 1) (OECD 2012a; OECD 2012b). The assay measures the activation of antioxidant response element (ARE)-dependent genes in HaCaT-derived human keratinocytes (Emter *et al.,* 2010; OECD 2015c). When a skin sensitizer (an electrophilic substance) covalently binds to proteins involved in the cytoprotective response, a subsequent protein disassociation event activates ARE-dependent genes. Activation of the ARE-dependent genes by skin sensitizers initiates transcription of a luciferase reporter gene via a constitutive promoter fused with an ARE, causing luminescence proportional to the degree of induction (OECD 2015c). Substances are considered to be sensitizers if the luciferase gene induction shows a statistically significant increase greater than 1.5-fold over control at a concentration <1000 μM, with cell viability >70%. We used sensitizer/nonsensitizer outcomes from KeratinoSens because adequate continuous data (i.e., effective concentration at 1.5-fold induction) for all substances were unavailable.

h-CLAT—h-CLAT assesses the ability of substances to activate and mobilize dendritic cells in the skin, the third key event of the skin sensitization AOP (Figure 1) (OECD 2012a; OECD 2012b). The assay is conducted by treating THP-1 cells, a human monocytic cell line that serves as a dendritic cell surrogate, with a test substance for 24 h (Ashikaga *et al.,* 2006; OECD 2015a). Changes in CD86 and CD54 cell surface marker expression caused by the test substance are then measured by flow cytometry. Substances are classified as sensitizers if the relative fluorescence intensity ≥150% of baseline for CD86 or at least 200% of baseline for CD54 at concentrations where cell viability ≥50% of control in at least two of three independent tests. We used sensitizer/nonsensitizer outcomes from h-CLAT because adequate continuous data (i.e., effective concentration at 150% induction for the CD86

marker and the effective concentration at 200% induction for the CD54 marker) were unavailable.

In Silico Read-across—QSAR Toolbox software v3.2 (OECD 2009; OECD 2014) was used to generate an *in silico* read-across prediction of whether each substance or its predicted auto-oxidation product or metabolite was a sensitizer or nonsensitizer based on *in vivo* data from structurally and mechanistically similar analogs. The *in silico* predictions cover the adverse outcome and all preceding key events because *in vivo* data (LLNA, guinea pig, and human outcomes) are used in the read-across method. The read-across protocol for QSAR Toolbox is provided as Supplemental File 2. Briefly, the Chemical Abstracts Service Registry Number for each substance was the input provided to QSAR Toolbox. We searched for protein binding alerts for each substance using all four protein binding profilers in QSAR Toolbox. For substances with no protein binding alerts, auto-oxidation products and skin metabolites were generated and then those were profiled for protein binding alerts. If the oxidation products and metabolites had no alerts, then the substance was classified as a nonsensitizer. Test substances, products, or metabolites with protein binding alerts were grouped into categories with substances of similar structural and mechanistic characteristics. The read-across prediction of skin sensitization hazard was determined using the *in vivo* skin sensitization hazard data for the substances nearest the target substance, based on log K_{ow}.

Physicochemical Properties—We collected data for octanol:water partition coefficient, water solubility, vapor pressure, molecular weight, melting point, and boiling point from the following sources, with preference given to experimental values:

- SRC, Inc. – EPI Suite™ (Experimental) (http://esc.syrres.com/interkow/EPiSuiteData.htm)
- ChemID*plus* – a TOXNET (Toxicology Data Network) Database (http://chem.sis.nlm.nih.gov/chemidplus)
- ChemSpider – EPI Suite™ (Predicted); Alfa Aesar (Experimental) (http://www.chemspider.com/)
- Hazardous Substances Databank (HSDB; http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB)
- ECHA database (http://echa.europa.eu/information-on-chemicals)

For 10 substances, values for one or more physicochemical properties could not be located. In these cases, values were imputed via quantitative structure–property relationship models built using binary molecular fingerprints and machine learning approaches as described in Zang et al. (manuscript in preparation).

Data Processing

If a substance had multiple continuous results for the DPRA, a geometric mean of those results was calculated after negative peptide depletion values were set to zero. If a substance had multiple sensitizer/nonsensitizer results for any assay, the most prevalent result was used; if there were an equal number of sensitizer and nonsensitizer results for a substance, it

was classified as a sensitizer for that assay. There were 11 substances with an equal number of sensitizer and nonsensitizer results for DPRA, five substances for KeratinoSens, eight substances for h-CLAT, and five substances for the LLNA. The final results for each substance are provided in the Supplemental File 1, along with the QSAR Toolbox read-across results.

Building Predictive Models

Training and Test Sets for Predictive Modeling—The 120 substances in the database were divided into training and external test sets in the approximate proportions of 80% to 20%. All substances were first classified as sensitizer or nonsensitizer based on the LLNA result. The substances in each classification were then parsed into groups according to their structural similarity, as determined by the expert judgment of a chemist who examined the structure of each substance. Using these groupings, training and test sets were constructed to reflect both the structural heterogeneity and the positive/negative sensitizer classifications of the overall substance set; however, within these constraints each substance was assigned randomly to either the training or the test set. This process yielded a training set containing 94 substances (78% of the 120), which consisted of 68 LLNA sensitizers (72% or 68/94) and 26 LLNA nonsensitizers (28% or 26/94). The external test set consisted of the remaining 26 substances (22% of the 120), with 19 LLNA sensitizers (73% or 19/26) and 7 LLNA nonsensitizers (27% or 7/26). The training and test sets are similar to one another and to the full 120-substance set with respect to the distributions of LLNA potency, product use categories, diversity of chemical structures (chemotype frequencies), prehaptens and prohaptens, and mechanistic protein binding domains (see Supplemental File 3).

Prediction of LLNA Outcomes Using Training and Test Sets—We used the training set of 94 substances to build models for predicting LLNA outcomes using the following six machine learning approaches (see Kuhn and Johnson (2013) for details on the approaches):

- Artificial neural network (ANN)
- Naïve Bayes algorithm (NB)
- Classification and regression tree (CART)
- Linear discriminant analysis (LDA)
- Logistic regression (LR)
- Support vector machine (SVM)

Model building was implemented using the following packages in the R statistical analysis software for Windows v2.15.1 (The R Core Team 2013):

- Package nnet: for ANN
- Package MASS: for LDA and LR
- Package rpart: for CART
- Package e1071: for NB and SVM

We initially developed models using each of the six machine learning approaches and each of six variable sets based on different combinations of the 13 variables collected, yielding a total of 36 models. Table 2 defines the six variable sets. The numbers in the column headings represent Variable Sets 1 through 6, and the Xs in each column and the color coding indicate what data were included in each variable set. Once each model was trained it was used to predict LLNA outcomes for each substance in the test set. These outcomes were reported as probabilities; substances with a probability greater than 0.5 of being either a sensitizer or nonsensitizer were assigned to the respective class.

Evaluation of Model Performance and Further Optimization—Model performance was evaluated by calculating the sensitivity, specificity, and accuracy for predicting LLNA outcomes. These metrics were calculated using the following formulae:

$$Sensitivity = \frac{True\ Postives}{True\ Positives + False\ Negatives}$$

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives}$$

$$Accuracy = \frac{True\ Postives + True\ Negatives}{True\ Postives + False\ Negatives + True\ Negatives + False\ Positives}$$

SVM, the machine learning approach with the highest performance for predicting LLNA outcome from test set data, was selected for use in further optimizing the number and type of input variables, resulting in an additional 18 models with various combinations of input variables being evaluated. Performance of the machine learning models was compared with the performance of the individual non-animal methods alone and with two test battery approaches using results from those methods. Test Battery 1 classified a substance as a sensitizer if one non-animal method classified the substance as a sensitizer. Test Battery 2 classified a substance as a sensitizer if any two non-animal methods classified the substance as a sensitizer.

Prediction of LLNA Outcomes Using Leave-One-Out Cross-Validation—In order to confirm the robustness and reliability of the predictive models, we also evaluated the seven models with the highest performance for predicting LLNA classification using a leave-one-out cross-validation (LOOCV) procedure in addition to testing models with the external test set. For LOOCV, the training and test set substances were combined, and the performance of the model was evaluated against every substance in the dataset when it appears in an external test set on its own (Kuhn and Johnson 2013). Thus, 119 substances from the complete set of 120 substances were used as the training set for building the model and the remaining substance was used for testing the model. The procedure is performed 120 times with each of the 120 substances used exactly once as the external validation set. The predictive accuracy is calculated by averaging individual values over the 120 runs.

Results

Accuracy of Individual Methods and Test Batteries

The performance of the individual non-animal methods for predicting LLNA outcomes is shown in Table 3. Of the individual methods, h-CLAT had the highest sensitivity (84%), specificity (86%) and accuracy (85%) for the test set. Read-across using QSAR Toolbox also

had the highest specificity (86%). Test Battery 1, which classified a substance as a sensitizer if one non-animal method classified the substance as a sensitizer, yielded higher sensitivity than any of the individual methods, much lower specificity, and the same accuracy as the h-CLAT. Test Battery 2, which classified a substance as a sensitizer if any two non-animal methods classified the substance as a sensitizer, had higher sensitivity than the individual test methods, specificity within the range of the individual methods, and accuracy similar to h-CLAT.

Accuracy of Machine Learning Approaches

For each machine learning approach, the variable set(s) that produced the best performance of the first 36 models are shown in Table 4. Based on the accuracy for predicting LLNA outcomes, the modeling approaches ranked as follows: SVM > ANN > LR > LDA > CART = NB. Because SVM was the model with the best performance across multiple variable sets, it was used for further testing and optimization. In these subsequent analyses, DPRA results were represented by the average lysine and cysteine peptide depletion values, because this measurement was more highly correlated to LLNA outcomes than other DPRA measures (average cysteine peptide depletion, average lysine peptide depletion, and binary DPRA result) (data not shown).

Optimization of the SVM Models

An additional 18 feature combinations were examined to determine the optimal SVM approach. The variable set that included h-CLAT, read-across from QSAR Toolbox, and the six physicochemical properties (No. 7 in Table 5) achieved the highest accuracies for the test (96%) and training sets (97%) (Table 5). The variable set with only physicochemical properties (Table 5, No. 24) produced the lowest accuracy: 73% for both test and training sets. The three models that used log P instead of all six physicochemical properties had lower accuracy (compare model 7 and 17 in Table 5, models 8 and 19 in Table 4, and model 13 in Table 5 with SVM variable set 5 in Table 4).

LOOCV for SVM Models

As expected, applying LOOCV decreased the balanced accuracy for all seven SVM models (which ranged from 84% to 89% [Table 6]), but only marginally as compared to the results without LOOCV (89–99% balanced accuracy). Model 1 (KeratinoSens + h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties) achieved the highest LOOCV sensitivity (92%) while Model 7 (h-CLAT + Toolbox + 6 properties) had the highest specificity (94%). Evaluating based on accuracy, both Model 1 (DPRA + KeratinoSens + h-CLAT + Toolbox + Lys + Cys + Avg.Lys.Cys + 6 properties) and Model 10 (h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties) had the highest LOOCV value (89%). These two models also achieved balanced sensitivity and specificity with 89% and 91%, respectively, for Model 1, and 90% and 88%, respectively, for Model 10.

Misclassified Substances

Training Set—The seven SVM models with the highest accuracies misclassified a total of nine substances, two false positives and seven false negatives, in the training set (Table 7).

None of the false negatives were prehaptens (there were two prehaptens in the training set). Four of the seven models (1, 5, 7, and 11) correctly classified all prohaptens; however, three of the seven models (8, 9, and 10) misclassified one of three prohaptens (there were 12 prohaptens in the training set).

The most frequently misclassified substances were 3-phenoxypropiononitrile (false positive) and nonanoic acid (false negative), which were both misclassified by six models. The LLNA classifications for both of these substances were based on only one test. KeratinoSens was the only non-animal method that correctly classified 3-phenoxypropiononitrile as an LLNA nonsensitizer. The only available LLNA test for this compound was negative with a flat dose-response curve (Kern *et al.,* 2010). h-CLAT was the only non-animal method that correctly classified nonanoic acid as an LLNA sensitizer. Only one LLNA test was available for nonanoic acid, which is a weak LLNA sensitizer (EC3 = 35%) and a strong irritant (Montelius *et al.,* 1998). It is well-documented that false positives in the LLNA are often associated with skin irritants that are not sensitizers (Anderson *et al.,* 2011). The other seven misclassified substances were misclassified by only two or fewer models.

Test Set—The seven SVM models with the highest accuracies misclassified seven substances, two false positives and five false negatives, in the test set (Table 8). None of the false negatives were prehaptens (there were two prehaptens in the test set). Again, the three models with the highest accuracies correctly classified all prohaptens; however, two of the seven models misclassified the same two prohaptens (there were four prohaptens in the test set).

The most frequently misclassified substances in the test set were two false negatives, coumarin (misclassified by all seven models) and undecylenic acid (misclassified by three models). For coumarin, KeratinoSens was the only non-animal method that had a correct positive result. Coumarin produced equivocal LLNA results (i.e., an equal number of positive [2] and negative [2] LLNA tests) (Gerberick *et al.,* 2005; Vocanson *et al.,* 2006). To be conservative (i.e., protective of human health), our reference result was positive. While it was a weak sensitizer in the positive tests (EC3 = 29.6%), the response was attributed to contaminants in commercial products containing coumarin (Vocanson *et al.,* 2006). For undecylenic acid, both DPRA and KeratinoSens produced correct positive results. Only one LLNA test was available for undecylenic acid; it is also a weak sensitizer (EC3 =19.4%) (Kreiling *et al.,* 2008). The other five misclassified substances were misclassified by only two or fewer models.

Coumarin and undecylenic acid had structural analogs that were misclassified in the training set (see Supplemental File 1). 3,4-Dihydrocoumarin, which was in the training set, is a benzopyran that is structurally similar to coumarin. 3,4-Dihydrocoumarin was misclassified by one model. Nonanoic acid, which was in the training set, is an aliphatic carboxylic acid that is structurally similar to undecylenic acid. It was misclassified in six of seven models. The seven models with the highest accuracies misclassified two (Models 7–10) to three (Models 1, 5, and 11) of these compounds.

Discussion

Fostering the evaluation and promotion of alternative test methods for regulatory use in skin sensitization hazard assessment has long been one of ICCVAM's top priorities (NIEHS 2013). ICCVAM is committed toward continued work in this area and believes that development of non-animal testing strategies for the identification of skin sensitizers is an achievable near-term goal. Although development of skin sensitization is a complex process, the key biological events have been documented and agreed upon in the AOP for substances that produce skin sensitization through covalent binding to proteins (OECD 2012a; OECD 2012b). We compiled a database of chemicals with test data from validated skin sensitization tests (LLNA, DPRA, KeratinoSens, and h-CLAT), *in silico* read-across predictions that considered auto-oxidation and skin metabolism products and *in vivo* skin sensitization hazard, and physicochemical parameters relevant to skin penetration. We then created and evaluated machine learning methods to integrate the non-animal data to predict skin sensitization hazard.

Our study confirmed that an integrated approach to skin sensitization testing is required to accurately identify these hazards, as a single non-animal method cannot recapitulate the complexity of the multi-step physiological process that occurs *in vivo* (Rovida *et al.,* 2015). For the test set of 26 substances used in this study, the highest accuracy for the prediction of LLNA outcomes for any single non-animal method alone was 85% (Table 3). The best performing simple test battery (e.g., Test Battery 1 with accuracy = 85%) did not improve upon the accuracy of the individual non-animal methods. However, the seven best performing machine learning models greatly improved upon the individual methods and test batteries with accuracies of 89–96% for the test set and 96–99% for the training set (Tables 4 and 5). The LOOCV, which avoids any bias introduced during the selection of test and training sets, yielded accuracies of 84–89% for these models. Due to the removal of this bias, the LOOCV accuracies are more likely to reflect the accuracy of these models when they are applied to additional external datasets.

Multiple models using different combinations of non-animal data exhibited high accuracy in hazard classification predictions. This raises the potential for flexibility in the choice of data inputs among the various test methods and physiochemical properties evaluated. This could be particularly important to laboratories or groups constrained by available resources. In fact, one of the seven highest performing models, Model 7, used only one *in vitro* assay. The performance of the top seven models is similar enough that investigators could select from two *in vitro* or *in chemico* methods to use, based on their experience with the methods: DPRA and KeratinoSens (Model 8), DPRA and h-CLAT (Model 10) or h-CLAT and KeratinoSens (Model 11). However, based on the results with the current dataset, the SVM model with h-CLAT as the only *in vitro* method (Model 7) or the models with all of the *in vitro* methods (Model 1 and Model 5) were best at correctly classifying prohaptens.

The advantages to integrating data from these non-animal methods to determine skin sensitization hazard is that the limitations of each individual method can be overcome. For example, DPRA has no metabolic capacity and thus is not expected to correctly classify prohaptens (OECD 2015b). KeratinoSens (OECD 2015c) and h-CLAT (OECD 2015a) can

classify some but not all prohaptens correctly. However, four of the highest performing SVM models correctly classified all 16 prohaptens in the training and test sets. The capacity to correctly predict prohaptens may have been aided by the inclusion of the *in silico* read-across input, which evaluated auto-oxidation products and skin metabolites if no protein binding alerts were identified in the parent compound. Although DPRA has consistently classified prehaptens correctly (OECD 2015b), KeratinoSens (OECD 2015c) and h-CLAT (OECD 2015a) have not. The seven best performing SVM models however, correctly classified the three prehaptens and six pre/prohaptens as sensitizers.

A number of uncomplicated test batteries (Bauch *et al.,* 2012; Natsch *et al.,* 2009; Natsch *et al.,* 2013; Nukada *et al.,* 2013; Urbisch *et al.,* 2015) and testing strategies (Bauch *et al.,* 2012; Nukada *et al.,* 2013; Takenouchi *et al.,* 2015) have been developed to predict LLNA skin sensitization hazard without using animals. These batteries and strategies are all biologically based models that use some combination of non-animal tests that are mechanistically relevant to the AOP for skin sensitization. The models provide good performance (79–96% accuracy), but they have not yet been evaluated on external data sets.

There are a limited number of published machine learning approaches to predict LLNA skin sensitization hazard (Hirota *et al.,* 2015; Jaworska *et al.,* 2013; Jaworska *et al.,* 2011; Luechtefeld *et al.,* 2015; Tsujita-Inoue *et al.,* 2014). Bayesian networks (Jaworska *et al.,* 2013; Jaworska *et al.,* 2011; Pirone *et al.,* 2014), artificial neural networks (Hirota *et al.,* 2015; Tsujita-Inoue *et al.,* 2014) and hidden Markov models (Luechtefeld *et al.,* 2015) have mainly been applied to the prediction of potency. The Bayesian network models, which integrate a variety of data (KeratinoSens, U937 activation assay, skin bioavailability, DPRA, log octanol:water partition coefficient, and an *in silico* prediction from TIMES SS), some of which were not included in our modeling effort (U937 activation assay, skin bioavailability, and TIMES SS), are arguably the most developed and well-tested machine learning models. The Bayesian network accuracies for predicting LLNA hazard for the test sets and training sets (91–95%) (Jaworska *et al.,* 2013; Jaworska *et al.,* 2011; Pirone *et al.,* 2014) are comparable to the best SVM models (89–99%; Tables 4 and 5) from our work, although the performance of the two models cannot be rigorously compared because they do not use assay data from exactly the same substances.

One of the limitations of the SVM models as presented is that they predict skin sensitization hazard but not potency. Potency information would be needed for risk assessment (i.e., to define the maximum concentration of a substance that is unlikely to produce skin sensitization). However, there are regulatory classification and labeling applications that require only hazard assessment. EPA (40 CFR 158.500 ; 40 CFR 161.340), OSHA (29 CFR 1910.1200) (Appendix A), and the European Chemicals Agency (for REACH) (ECHA 2015) use skin sensitization data for labeling to warn consumers and workers of the hazards associated with handling and use of potential skin sensitizers. OSHA requirements, which are consistent with the Globally Harmonized System of Classification and Labeling of Chemicals (UN 2013), require potency classification only if the skin sensitization data are adequate to characterize potency (29 CFR 1910.1200). For hazard classification, however, the seven best SVM models developed here have an advantage over the published Bayesian network models (Jaworska *et al.,* 2013; Jaworska *et al.,* 2011; Pirone *et al.,* 2014) in that

there are seven models to choose from that require a different combination of resources, all of which are publicly available and without licensure requirements. The R code for the models can be obtained by contacting one of the authors (Judy Strickland). A laboratory can choose the model that best fits their resources and expertise. Additionally, all the test information needed for the SVM models comes from internationally-accepted (or nearly accepted in the case of h-CLAT) OECD test guidelines (OECD 2015a; OECD 2015b; OECD 2015c) or freely available software supported by OECD (OECD 2014). Physicochemical property data can also be obtained from publically available sources.

The integrated decision strategies developed for this effort suggest that computational approaches are promising tools to effectively integrate data sources to identify potential skin sensitizers without testing animals. ICCVAM's future efforts in sensitization modeling will be directed at testing these models with additional substances and adapting the models for use with formulations or unknown mixtures. ICCVAM also plans to evaluate the use of machine learning approaches to predict skin sensitization hazard for humans, the species of interest. In addition, models to predict skin sensitization potency will be constructed and evaluated to more completely inform classification and risk assessment applications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Drs. D. Germolec, B.A. Merrick, R. Luebke, and M. Ward for their thoughtful critical review of this manuscript. This project was funded in whole or in part with federal funds from the NIEHS, NIH under contract HHSN273201500010C to ILS in support of NICEATM.

Abbreviations

ICCVAM	Interagency Coordinating Committee for the Validation of Alternative Methods
OECD	Organisation for Economic Co-operation and Development
EPA	U.S. Environmental Protection Agency
NIH	U.S. National Institutes of Health
NIEHS	U.S. National Institute of Environmental Health Sciences
FDA	U.S. Food and Drug Administration
LLNA	murine local lymph node assay
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
AOP	adverse outcome pathway
IDS	integrated decision strategy
DPRA	direct peptide reactivity assay

h-CLAT	human cell line activation test
SVM	support vector machine
LOOCV	leave-one-out cross-validation

References

16 CFR 1500.3. Title 16- Commercial Practices. Consumer Product Safety Commission; 2014.

29 CFR 1910.1200. Title 29- Labor. Occupational Safety and Health Administration; 2012. Occupational Safety and Health Standards.

40 CFR 158.500. Title 40- Protection of Environment. Environmental Protection Agency; 2012.

40 CFR 161.340. Title 40- Protection of Environment. Environmental Protection Agency; 2012.

AltTox. [November 4, 2015] Toxicity Endpoints & Tests: Skin Sensitization. 2014. http://alttox.org/mapp/toxicity-endpoints-tests/skin-sensitization/

Anderson SE, Siegel PD, Meade BJ. The LLNA: A Brief Review of Recent Advances and Limitations. J Allergy. 2011; 2011:424203.

Angers-Loustau A, Tosti L, Casati S. The regulatory use of the Local Lymph Node Assay for the notification of new chemicals in Europe. Regul Toxicol Pharmacol. 2011; 60:300–307. [PubMed: 21539884]

Ashikaga T, Sakaguchi H, Sono S, Kosaka N, Ishikawa M, Nukada Y, Miyazawa M, Ito Y, Nishiyama N, Itagaki H. A comparative evaluation of in vitro skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). Altern Lab Anim. 2010; 38:275–284. [PubMed: 20822320]

Ashikaga T, Yoshida Y, Hirota M, Yoneyama K, Itagaki H, Sakaguchi H, Miyazawa M, Ito Y, Suzuki H, Toyoda H. 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. 2006; 20:767–773. [PubMed: 16311011]

Ball N, Cagen S, Carrillo JC, Certa H, Eigler D, Emter R, Faulhammer F, Garcia C, Graham C, Haux C, Kolle SN, Kreiling R, Natsch A, Mehling A. Evaluating the sensitization potential of surfactants: Integrating data from the local lymph node assay, guinea pig maximization test, and in vitro methods in a weight-of-evidence approach. Regul Toxicol Pharmacol. 2011; 60:389–400. [PubMed: 21645576]

Basketter DA, Gerberick GF, Kimber I, Loveless SE. The local lymph node assay: a viable alternative to currently accepted skin sensitization tests. Food Chem Toxicol. 1996; 34:985–997. [PubMed: 9012774]

Basketter, DA.; Kimber, I. Predictive tests for irritants and allergens and their use in quantitative risk assessment. In: Frosch, P.; Menné, T.; Lepoittevin, J-P., editors. Contact Dermatitis. Springer Verlag; Heidelberg: 2006. p. 179-188.

Bauch C, Kolle SN, Fabian E, Pachel C, Ramirez T, Wiench B, Wruck CJ, Ravenzwaay BV, Landsiedel R. Intralaboratory validation of four in vitro assays for the prediction of the skin sensitizing potential of chemicals. Toxicol In Vitro. 2011; 25:1162–1168. [PubMed: 21669280]

Bauch C, Kolle SN, Ramirez T, Eltze T, Fabian E, Mehling A, Teubner W, van Ravenzwaay B, Landsiedel R. Putting the parts together: combining in vitro methods to test for skin sensitizing potentials. Regul Toxicol Pharmacol. 2012; 63:489–504. [PubMed: 22659254]

Birnbaum LS. 15 years out: reinventing ICCVAM. Environ Health Perspect. 2013; 121:a40. [PubMed: 23380598]

Bruckner AL, Weston WL, Morelli JG. Does sensitization to contact allergens begin in infancy? Pediatrics. 2000; 105:e3. [PubMed: 10617740]

Delaine T, Niklasson IB, Emter R, Luthman K, Karlberg AT, Natsch A. Structure--activity relationship between the in vivo skin sensitizing potency of analogues of phenyl glycidyl ether and the induction of Nrf2-dependent luciferase activity in the KeratinoSens in vitro assay. Chem Res Toxicol. 2011; 24:1312–1318. [PubMed: 21751775]

EC. Regulation (EC) No. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No. 793/93 and Commission Regulation (EC) No. 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. OJ L. 2006:396.

ECHA. Chapter R.7a: Endpoint Specific Guidance (Version 4.0). European Chemicals Agency (ECHA); Helsinki: 2015. Guidance on Information Requirements and Chemical Safety Assessment.

Emter R, Ellis G, Natsch A. Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. Toxicol Appl Pharmacol. 2010; 245:281–290. [PubMed: 20307559]

Estrada E, Patlewicz G, Chamberlain M, Basketter D, Larbey S. Computer-aided knowledge generation for understanding skin sensitization mechanisms: the TOPS-MODE approach. Chem Res Toxicol. 2003; 16:1226–1235. [PubMed: 14565764]

European Union. 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. OJ L. 2003; 66:26–35.

Gerberick GF, Ryan CA, Kern PS, Schlatter H, Dearman RJ, Kimber I, Patlewicz GY, Basketter DA. Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. Dermatitis. 2005; 16:157–202. [PubMed: 16536334]

Gerberick GF, Troutman JA, Foertsch LM, Vassallo JD, Quijano M, Dobson RLM, Goebel C, Lepoittevin JP. Investigation of peptide reactivity of pro-hapten skin sensitizers using a peroxidase-peroxide oxidation system. Toxicol Sci. 2009; 112:164–174. [PubMed: 19748994]

Gerberick GF, Vassallo JD, Bailey RE, Chaney JG, Morrall SW, Lepoittevin JP. Development of a peptide reactivity assay for screening contact allergens. Toxicol Sci. 2004; 81:332–343. [PubMed: 15254333]

Gerberick GF, Vassallo JD, Foertsch LM, Price BB, Chaney JG, Lepoittevin JP. Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach. Toxicol Sci. 2007; 97:417–427. [PubMed: 17400584]

Guyard-Nicodeme M, Gerault E, Platteel M, Peschard O, Veron W, Mondon P, Pascal S, Feuilloley MG. Development of a multiparametric in vitro model of skin sensitization. J Appl Toxicol. 2015; 35:48–58. [PubMed: 24496914]

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, Miyazawa M. Evaluation of combinations of in vitro sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization. J Appl Toxicol. 2015; 35:1333–1347. [PubMed: 25824844]

Jaworska J, Dancik Y, Kern P, Gerberick F, Natsch A. Bayesian integrated testing strategy to assess skin sensitization potency: From theory to practice. J Appl Toxicol. 2013; 33:1353–1364. [PubMed: 23670904]

Jaworska J, Harol A, Kern PS, Frank Gerberick G. Integrating non-animal test information into an adaptive testing strategy - Skin sensitization proof of concept case. ALTEX. 2011; 28:211–225. [PubMed: 21993957]

Joint Research Centre of the European Union. EURL ECVAM Recommendation on the Direct Peptide Reactivity Assay (DPRA) for Skin Sensitisation Testing. Publications Office of the European Union; Luxembourg: 2013.

Joint Research Centre of the European Union. EURL ECVAM Recommendation on the KeratinoSens™ assay for skin sensitisation testing. Publications Office of the European Union; Luxembourg: 2014.

Joint Research Centre of the European Union. EURL ECVAM Recommendation on the human Cell Line Activation Test (h-CLAT) for skin sensitisation testing. Publications Office of the European Union; Luxembourg: 2015.

Kern PS, Gerberick GF, Ryan CA, Kimber I, Aptula A, Basketter DA. Local lymph node data for the evaluation of skin sensitization alternatives: A second compilation. Dermatitis. 2010; 21:8–32. [PubMed: 20137736]

Kimber I, Basketter DA, Gerberick GF, Ryan CA, Dearman RJ. Chemical allergy: Translating biology into hazard characterization. Toxicol Sci. 2011; 120:S238–S268. [PubMed: 21097995]

Kreiling R, Hollnagel HM, Hareng L, Eigler D, Lee MS, Griem P, Dreeflen B, Kleber M, Albrecht A, Garcia C, Wendel A. Comparison of the skin sensitizing potential of unsaturated compounds as assessed by the murine local lymph node assay (LLNA) and the guinea pig maximization test (GPMT). Food Chem Toxicol. 2008; 46:1896–1904. [PubMed: 18343554]

Kuhn, M.; Johnson, K. Applied predictive modeling. Springer; New York: 2013.

Luechtefeld T, Maertens A, McKim JM, Hartung T, Kleensang A, Sa-Rocha V. Probabilistic hazard assessment for skin sensitization potency by dose-response modeling using feature elimination instead of quantitative structure-activity relationships. J Appl Toxicol. 2015; 35:1361–1371. [PubMed: 26046447]

MacKay C, Davies M, Summerfield V, Maxwell G. From pathways to people: applying the adverse outcome pathway (AOP) for skin sensitization to risk assessment. ALTEX. 2013; 30:473–486. [PubMed: 24173169]

Mehling A, Eriksson T, Eltze T, Kolle S, Ramirez T, Teubner W, van Ravenzwaay B, Landsiedel R. Non-animal test methods for predicting skin sensitization potentials. Arch Toxicol. 2012; 86:1273–1295. [PubMed: 22707154]

Montelius J, Wahlkvist H, Boman A, Wahlberg JE. Murine local lymph node assay for predictive testing of allergenicity: Two irritants caused significant proliferation. Acta Derm Venereol. 1998; 78:433–437. [PubMed: 9833042]

Murphy, K.; Travers, P.; Walport, M.; Janeway, C. Janeway's immunobiology. Garland Science; New York: 2012.

Natsch, A. Reducing, Refining and Replacing the Use of Animals in Toxicity Testing. Royal Society of Chemistry; 2014. Integrated Approaches to Safety Testing: General Principles and Skin Sensitization as a Test Case; p. 265-288.

Natsch A, Emter R. Skin sensitizers induce antioxidant response element dependent genes: Application to the in vitro testing of the sensitization potential of chemicals. Toxicol Sci. 2008; 102:110–119. [PubMed: 17932397]

Natsch A, Emter R, Ellis G. 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. Toxicol Sci. 2009; 107:106–121. [PubMed: 18832184]

Natsch A, Haupt T. Utility of rat liver S9 fractions to study skin-sensitizing prohaptens in a modified KeratinoSens assay. Toxicol Sci. 2013; 135:356–368. [PubMed: 23872582]

Natsch A, Ryan CA, Foertsch L, Emter R, Jaworska J, Gerberick F, Kern P. A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. J Appl Toxicol. 2013; 33:1337–1352. [PubMed: 23576290]

NIEHS. Request for Information on Alternative Skin Sensitization Test Methods and Testing Strategies and for Comment on ICCVAM's Proposed Activities. Fed Regist. 2013; 78:68076–68077.

Nukada Y, Ashikaga T, Miyazawa M, Hirota M, Sakaguchi H, Sasa H, Nishiyama N. Prediction of skin sensitization potency of chemicals by human Cell Line Activation Test (h-CLAT) and an attempt at classifying skin sensitization potency. Toxicol In Vitro. 2012; 26:1150–1160. [PubMed: 22796097]

Nukada Y, Ashikaga T, Sakaguchi H, Sono S, Mugita N, Hirota M, Miyazawa M, Ito Y, Sasa H, Nishiyama N. Predictive performance for human skin sensitizing potential of the human cell line activation test (h-CLAT). Contact Dermatitis. 2011; 65:343–353. [PubMed: 21767275]

Nukada Y, Miyazawa M, Kazutoshi S, Sakaguchi H, Nishiyama N. Data integration of non-animal tests for the development of a test battery to predict the skin sensitizing potential and potency of chemicals. Toxicol In Vitro. 2013; 27:609–618. [PubMed: 23149339]

OECD. Guidance Document for Using the OECD (Q)SAR Application Toolbox to Develop Chemical Categories According to the OECD Guidance on Grouping of Chemicals. OECD Publishing; Paris: 2009. OECD Series on Testing and Assessment No. 102.

OECD. OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. OECD Publishing; Paris: 2010. Test No. 429. Skin Sensitisation: Local Lymph Node Assay.

OECD. Part 2: Use of the AOP to Develop Chemical Categories and Integrated Assessment and Testing Approaches. OECD Publishing; Paris: 2012a. OECD Series on Testing and Assessment No. 168. The Adverse Outcome Pathway for Skin Sensitisation Initated by Covalent Binding to Proteins.

OECD. Part 1: Scientific Assessment. OECD Publishing; Paris: 2012b. OECD Series on Testing and Assessment No. 168. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins.

OECD. [November 24, 2014] The OECD QSAR Toolbox. 2014. http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm

OECD. [August 12, 2015] Draft Proposal for a New Test Guideline. *In Vitro* Skin Sensitisation: human Cell Line Activation Test (h-CLAT). 2015a. http://www.oecd.org/env/ehs/testing/Draft-Proposal-for-a-new-Test-Guideline-on-in-vitro-skin-sensitisation-h-CLAT.pdf

OECD. OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. OECD Publsihing; Paris: 2015b. Test No. 442C. *In Chemico* Skin Sensitization: Direct Peptide Reactivity Assay (DPRA).

OECD. OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. OECD Publishing; Paris: 2015c. Test No. 442D. I*n Vitro* Skin Sensitisation: ARE-Nrf2 Luciferase Test Method.

Patlewicz G, Kuseva C, Kesova A, Popova I, Zhechev T, Pavlov T, Roberts DW, Mekenyan O. Towards AOP application--implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitization. Regul Toxicol Pharmacol. 2014; 69:529–545. [PubMed: 24928565]

Pirone JR, Smith M, Kleinstreuer NC, Burns TA, Strickland J, Dancik Y, Morris R, Rinckel LA, Casey W, Jaworska JS. Open source software implementation of an integrated testing strategy for skin sensitization potency based on a Bayesian network. ALTEX. 2014; 31:336–340. [PubMed: 24687303]

Reisinger K, Hoffmann S, Alepee N, Ashikaga T, Barroso J, Elcombe C, Gellatly N, Galbiati V, Gibbs S, Groux H, Hibatallah J, Keller D, Kern P, Klaric M, Kolle S, Kuehnl J, Lambrechts N, Lindstedt M, Millet M, Martinozzi-Teissier S, Natsch A, Petersohn D, Pike I, Sakaguchi H, Schepky A, Tailhardat M, Templier M, van Vliet E, Maxwell G. Systematic evaluation of non-animal test methods for skin sensitisation safety assessment. Toxicol In Vitro. 2015; 29:259–270. [PubMed: 25448812]

Roberts DW, Aptula AO. Determinants of skin sensitisation potential. J Appl Toxicol. 2008; 28:377–387. [PubMed: 17703504]

Rovida C, Alepee N, Api AM, Basketter DA, Bois FY, Caloni F, Corsini E, Daneshian M, Eskes C, Ezendam J, Fuchs H, Hayden P, Hegele-Hartung C, Hoffmann S, Hubesch B, Jacobs MN, Jaworska J, Kleensang A, Kleinstreuer N, Lalko J, Landsiedel R, Lebreux F, Luechtefeld T, Locatelli M, Mehling A, Natsch A, Pitchford JW, Prater D, Prieto P, Schepky A, Schuurmann G, Smirnova L, Toole C, van Vliet E, Weisensee D, Hartung T. Integrated Testing Strategies (ITS) for safety assessment. ALTEX. 2015; 32:25–40. [PubMed: 25413849]

Russell, WMS.; Burch, RL. The principles of humane experimental technique. Universities Federation for Animal Welfare; Potters Bar, Herts, England: 1992.

Sailstad DM, Hattan D, Hill RN, Stokes WS. ICCVAM evaluation of the murine local lymph node assay: I. The ICCVAM review process. Regul Toxicol Pharmacol. 2001; 34:249–257. [PubMed: 11754529]

Sakaguchi H, Ryan C, Ovigne JM, Schroeder KR, Ashikaga T. Predicting skin sensitization potential and inter-laboratory reproducibility of a human Cell Line Activation Test (h-CLAT) in the European Cosmetics Association (COLIPA) ring trials. Toxicol In Vitro. 2010; 24:1810–1820. [PubMed: 20510347]

Schmidt M, Goebeler M. Immunology of metal allergies. Journal of the German Society of Dermatology. 2015; 13:653–660. [PubMed: 26053629]

Schoeters G. The REACH perspective: toward a new concept of toxicity testing. J Toxicol Environ Health B Crit Rev. 2010; 13:232–241. [PubMed: 20574899]

Smith, CK.; Hotchkiss, SAM. Allergic Contact Dermatitis. Chemical and Metabolic Mechanisms. Taylor and Francis; London and New York: 2001.

Takenouchi O, 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, Hirota M, Ashikaga T, Miyazawa M. 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. J Appl Toxicol. 2015; doi: 10.1002/jat.3127

Takenouchi O, Miyazawa M, Saito K, Ashikaga T, Sakaguchi H. Predictive performance of the human Cell Line Activation Test (h-CLAT) for lipophilic chemicals with high octanol-water partition coefficients. J Toxicol Sci. 2013; 38:599–609. [PubMed: 23824015]

The R Core Team. R: A Language and Environment for Statistical Computing. Reference Index. R Foundation for Statistical Computing; 2013.

Thyssen JP, Linneberg A, Menne T, Johansen JD. The epidemiology of contact allergy in the general population--prevalence and main findings. Contact Dermatitis. 2007; 57:287–299. [PubMed: 17937743]

Tsujita-Inoue K, Hirota M, Ashikaga T, Atobe T, Kouzuki H, Aiba S. Skin sensitization risk assessment model using artificial neural network analysis of data from multiple in vitro assays. Toxicol In Vitro. 2014; 28:626–639. [PubMed: 24444449]

UN. Globally Harmonised Sysem of Classification and Labelling of Chemicals (GHS). United Nations; New York: 2013.

Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, Jaworska J, Kern PS, Gerberick F, Natsch A, Emter R, Ashikaga T, Miyazawa M, Sakaguchi H. Assessing skin sensitization hazard in mice and men using non-animal test methods. Regul Toxicol Pharmacol. 2015; 71:337–351. [PubMed: 25541156]

van der Veen JW, Rorije E, Emter R, Natsch A, van Loveren H, Ezendam J. Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals. Regul Toxicol Pharmacol. 2014; 69:371–379. [PubMed: 24813372]

Van Och FMM, Slob W, De Jong WH, Vandebriel RJ, Van Loveren H. A quantitative method for assessing the sensitizing potency of low molecular weight chemicals using a local lymph node assay: employment of a regression method that includes determination of the uncertainty margins. Toxicology. 2000; 146:49–59. [PubMed: 10773362]

Vocanson M, Goujon C, Chabeau G, Castelain M, Valeyrie M, Floc'h F, Maliverney C, Gard A, Nicolas JF. The skin allergenic properties of chemicals may depend on contaminants - evidence from studies on coumarin. Int Arch Allergy Immunol. 2006; 140:231–238. [PubMed: 16685137]

Williams ES, Panko J, Paustenbach DJ. The European Union's REACH regulation: a review of its history and requirements. Crit Rev Toxicol. 2009; 39:553–575. [PubMed: 19650717]

Williams WC, Copeland C, Boykin E, Quell SJ, Lehmann DM. Development and utilization of an ex vivo bromodeoxyuridine local lymph node assay protocol for assessing potential chemical sensitizers. J Appl Toxicol. 2015; 35:29–40. [PubMed: 24532485]

Yang C, Tarkhov A, Marusczyk J, Bienfait B, Gasteiger J, Kleinoeder T, Magdziarz T, Sacher O, Schwab CH, Schwoebel J, Terfloth L, Arvidson K, Richard A, Worth A, Rathman J. New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. J Chem Inf Model. 2015; 55:510–528. [PubMed: 25647539]

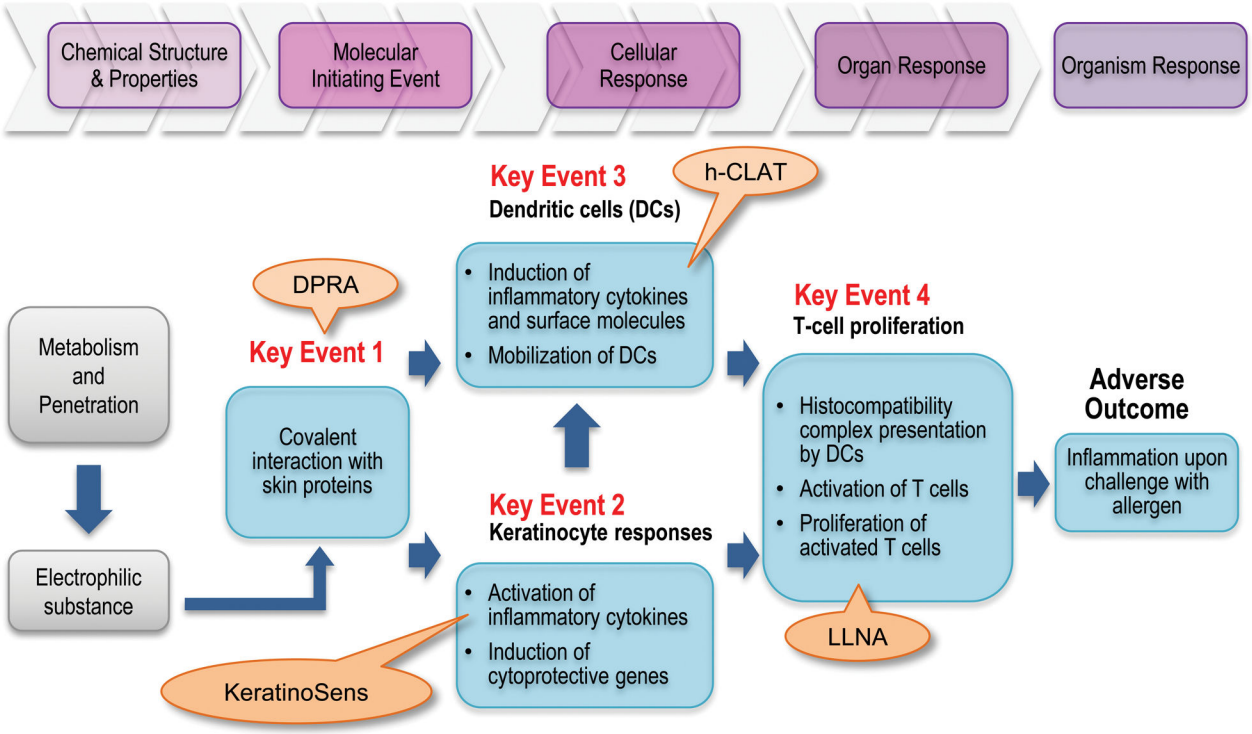

Figure 1. Adverse Outcome Pathway for Skin Sensitization Initiated by Covalent Binding to Proteins

Abbreviations: DPRA = direct peptide reactivity assay; h-CLAT = human cell line activation test; LLNA = murine local lymph node assay.

Note: Although KeratinoSens, h-CLAT, and LLNA are aligned with single key events, these assays also recapitulate the prior key events.

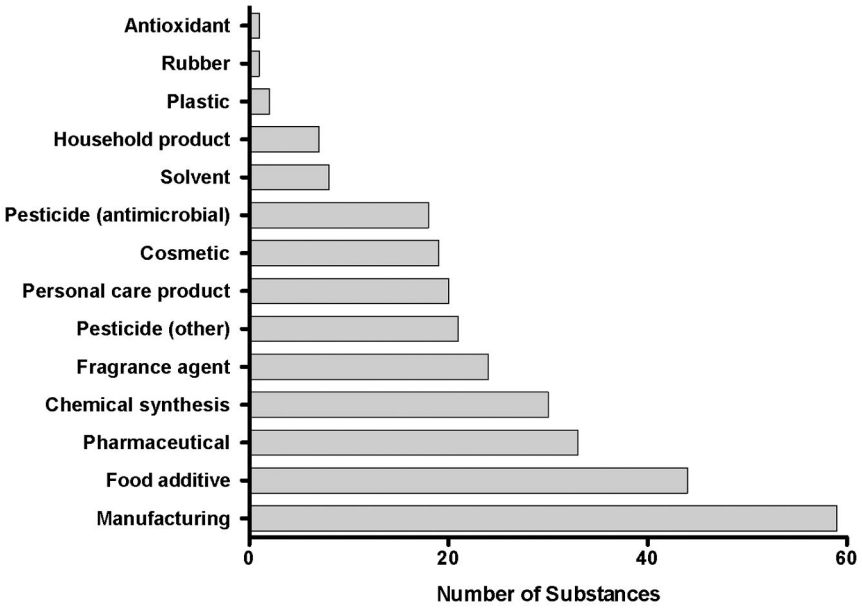

Figure 2. Product Uses for 120 Substances

Total number of substances exceeds 120 because most substances were associated with more than one product use.

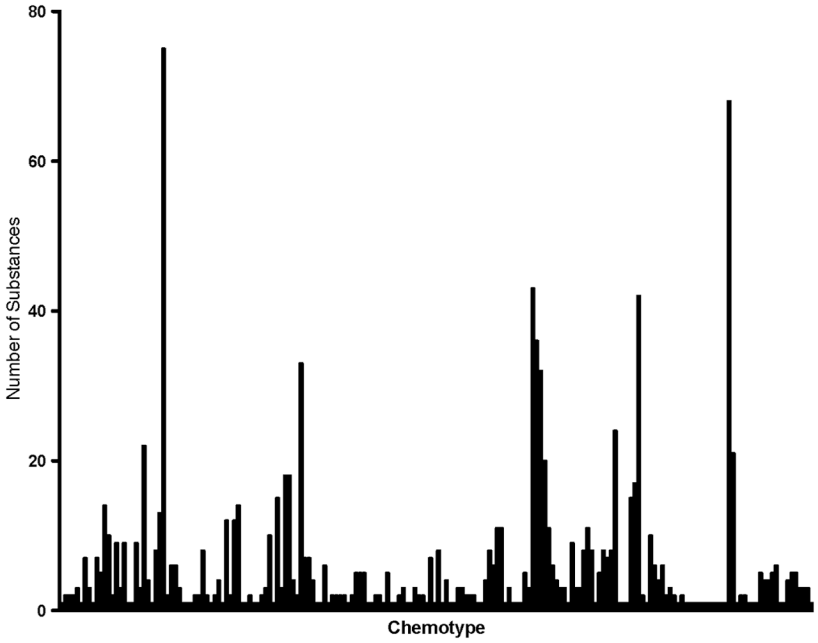

Figure 3. Frequency of Appearance of 192 Chemotypes in the 120 Substance Set
Bars show the number of substances with each of 192 chemotypes.

Table 1

Data Sources

Test Method	Reference
DPRA	Bauch *et al.* (2011)
	Bauch *et al.* (2012)
	Gerberick *et al.* (2004)
	Gerberick *et al.* (2007)
	Jaworska *et al.* (2011)
	Jaworska *et al.* (2013)
	Joint Research Centre of the European Union (2013)
	Natsch *et al.* (2013)
	Nukada *et al.* (2013)
KeratinoSens	Ball *et al.* (2011)
	Bauch *et al.* (2011)
	Bauch *et al.* (2012)
	Natsch *et al.* (2011)
	Emter *et al.* (2010)
	Joint Research Centre of the European Union (2014)
	Natsch *et al.* (2013)
h-CLAT	Ashikaga *et al.* (2010)
	Bauch *et al.* (2011)
	Bauch *et al.* (2012)
	Nukada *et al.* (2011)
	Nukada *et al.* (2012)
	Nukada *et al.* (2013)
	Sakaguchi *et al.* (2010)
	Takenouchi *et al.* (2013)
LLNA	NICEATM LLNA database
	Basketter *et al.* (1996) and Estrada *et al.* (2003) (xylene)
	Basketter and Kimber (2006) (diphenylcyclopropenone, maleic anhydride, and propyl gallate)
	Montelius *et al.* (1998) (nonanoic acid)
	Smith and Hotchkiss (2001) (2,4,6-trinitrobenzensulfonic acid)
	Van Och *et al.* (2000) (phthalic anhydride)

Abbreviations: DPRA = direct peptide reactivity assay; h-CLAT = human cell line activation test; LLNA = murine local lymph node assay; NICEATM = National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods.

Table 2

Six Variable Sets Used to Build Models for Predicting LLNA Outcome

Variable	Variable Set[a]					
	1	2	3	4	5	6
DPRA [binary]	X		X	X		
KeratinoSens [binary]	X		X	X	X	
h-CLAT [binary]	X		X	X	X	
Toolbox [binary]	X		X	X	X	X
Lys [continuous]	X	X		X		
Cys [continuous]	X	X		X		
Avg.Lys.Cys [continuous]	X	X		X	X	X
Log P [continuous]	X	X		X	X	X
Log S [continuous]	X	X		X	X	X
Log VP [continuous]	X	X			X	X
Melting Point [continuous]	X	X			X	X
Boiling Point [continuous]	X	X			X	X
Molecular Weight [continuous]	X	X			X	X

Abbreviations: Avg.Lys.Cys = average depletion for lysine and cysteine peptides; Cys = average depletion of cysteine peptide; DPRA = direct peptide reactivity assay; h-CLAT = human cell line activation test; LLNA = murine local lymph node assay; Log P = log octanol:water partition coefficient; Log S = log water solubility; Log VP = log vapor pressure; Lys = average depletion of lysine peptide; Toolbox = read-across using QSAR Toolbox.

[a]The Xs denote the input variables included in each variable set.

Table 3

Performance of Individual Methods and Simple Test Battery Approaches for Predicting LLNA Outcomes for Training and Test Sets

Method	Data Set[a]	Sensitivity (%)	Specificity (%)	Accuracy (%)
DPRA	Training	85	69	81
	Test	74	71	73
	All	83	70	79
KeratinoSens	Training	79	65	74
	Test	63	57	62
	All	76	64	73
h-CLAT	Training	83	58	75
	Test	84	86	85
	All	84	64	78
Toolbox	Training	78	73	75
	Test	74	86	77
	All	77	76	77
Test Battery 1 (≥ 1 positive)	Training	97	27	78
	Test	100	43	85
	All	98	30	79
Test Battery 2 (≥ 2 positive)	Training	91	62	83
	Test	90	71	84
	All	91	64	83

Abbreviations: DPRA = direct peptide reactivity assay; h-CLAT = human cell line activation test; LLNA = murine local lymph node assay; Toolbox = read-across using QSAR Toolbox.

[a]The training set of 94 substances contains 68 LLNA sensitizers and 26 LLNA nonsensitizers. The test set of 26 substances contains 19 LLNA sensitizers and 7 LLNA nonsensitizers. The entire set (All) contains 120 substances; 87 sensitizers and 33 nonsensitizers.

Table 4

Performance of Machine Learning Methods Predicting LLNA Outcomes for Training and Test Sets

Approach[a]	Variable Set[b]	Data Set[c]	Sensitivity (%)	Specificity (%)	Accuracy (%)
SVM	**1, 5**	**Training**	**99**	**96**	**98**
		Test	**90**	**100**	**92**
ANN	4	Training	93	89	93
		Test	90	86	89
LR	1	Training	93	85	90
		Test	84	100	89
LDA	1	Training	93	85	90
		Test	84	86	85
CART	1,2,4,5,6	Training	87	89	87
		Test	74	86	77
NB	6	Training	87	89	87
		Test	74	86	77

Abbreviations: ANN = artificial neural network; NB = naïve Bayes algorithm; CART = classification and regression tree; LDA = linear discriminant analysis; LLNA = murine local lymph node assay; LR = logistic regression; SVM = support vector machine.

[a]Performance statistics for the best performing variable sets for each machine learning approach are shown. Bolded text shows the best performing machine learning approach and variable sets.

[b]Predictor variables for each variable set are shown in Table 2. Multiple variable sets indicate equal performance.

[c]The training set of 94 substances contains 68 LLNA sensitizers and 26 LLNA nonsensitizers. The test set of 26 substances contains 19 LLNA sensitizers and 7 LLNA nonsensitizers.

J Appl Toxicol. Author manuscript; available in PMC 2017 September 01.

Table 5

Classification Results for SVM Models with 18 Additional Variable Combinations

No[a].	Variable Set	Data Set[b]	Sensitivity (%)	Specificity (%)	Accuracy (%)
7	h-CLAT + Toolbox + 6 properties	Training	97	97	97
		Test	95	100	96
8	KeratinoSens + Toolbox + Avg.Lys.Cys + 6 properties	Training	99	100	99
		Test	84	100	89
9	KeratinoSens + h-CLAT + Avg.Lys.Cys + 6 properties	Training	97	92	96
		Test	90	86	89
10	h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties	Training	96	96	96
		Test	84	100	89
11	KeratinoSens + h-CLAT + Toolbox + 6 properties	Training	96	96	96
		Test	90	86	89
12	h-CLAT + KeratinoSens + 6 properties	Training	94	89	93
		Test	90	86	89
13	h-CLAT + Avg.Lys.Cys + KeratinoSens + Toolbox + Log P	Training	91	96	93
		Test	90	86	89
14	h-CLAT + Avg.Lys.Cys + 6 properties	Training	96	92	95
		Test	84	86	85
15	Avg.Lys.Cys + Toolbox + 6 properties	Training	91	100	94
		Test	79	100	85
16	h-CLAT + 6 properties	Training	87	89	87
		Test	90	86	89
17	h-CLAT + Toolbox + Log P	Training	81	92	84
		Test	84	100	89
18	Avg.Lys.Cys + KeratinoSens + 6 properties	Training	93	96	94
		Test	74	86	77
19	Avg.Lys.Cys + KeratinoSens + Toolbox + Log P	Training	88	92	89
		Test	79	86	81

No[a].	Variable Set	Data Set[b]	Sensitivity (%)	Specificity (%)	Accuracy (%)
20	Avg.Lys.Cys + 6 properties	Training	85	100	89
		Test	74	100	81
21	Toolbox + 6 properties	Training	90	81	87
		Test	84	71	81
22	KeratinoSens + Toolbox + 6 properties	Training	91	85	89
		Test	74	86	77
23	KeratinoSens + 6 properties	Training	79	89	82
		Test	74	86	77
24	6 properties only	Training	68	89	73
		Test	73	71	73

Abbreviations: 6 properties = molecular weight, log octanol:water partition coefficient, log water solubility, log vapor pressure, melting point, and boiling point; Avg.Lys.Cys = average depletion for lysine and cysteine peptides from the direct peptide reactivity assay; h-CLAT = human cell line activation test; log P = log octanol:water partition coefficient; No. = number; SVM = support variable machine; Toolbox = read-across using QSAR Toolbox.

[a]Models are listed in descending order of the average accuracy of the test and training sets.

[b]The training set of 94 substances contains 68 LLNA sensitizers and 26 LLNA nonsensitizers. The test set of 26 substances contains 19 LLNA sensitizers and 7 LLNA nonsensitizers.

Table 6

LOOCV Results for Seven Highest Performing SVM Models

No.	Model (Accuracy[a])	Sensitivity (%)	Specificity (%)	Accuracy (%)
1	DPRA + KeratinoSens + h-CLAT + Toolbox + Lys + Cys + Avg.Lys.Cys + 6 properties (95%)	89	91	89
5	KeratinoSens + h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties (95%)	92	79	88
7	h-CLAT + Toolbox + 6 properties (97%)	85	94	88
8	KeratinoSens + Toolbox + Avg.Lys.Cys + 6 properties (94%)	84	91	86
9	KeratinoSens + h-CLAT + Avg.Lys.Cys + 6 properties (92%)	89	73	84
10	h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties (92%)	90	88	89
11	KeratinoSens + h-CLAT + Toolbox + 6 properties (92%)	89	79	86

Abbreviations: Avg.Lys.Cys = average depletion for lysine and cysteine; Cys = average % cysteine; DPRA = direct peptide reactivity assay; h-CLAT = human cell line activation test; LOOCV = leave-one-out cross-validation; Lys = average % lysine depletion; Toolbox = read-across using QSAR Toolbox; SVM = support vector machine.

[a]Average accuracy of the training and test sets for predicting the reference LLNA outcomes.

Table 7

Misclassified Substances for the Seven SVM Models with the Highest Accuracy – Training Set[a]

Model No./Variables[b]	3-Phenoxypropiononitrile	2-Acetylcyclohexanone	Pyridine[c]	Nonanoic acid	3,4-Dihydrocoumarin[c]	Benzylidene acetone	Xylene	2-Hydroxyethyl acrylate	Eugenol[c]
7) h-CLAT + Toolbox + 6 properties (97%)	POS	NEG	POS	NEG	POS	NEG	POS	POS	POS
1) DPRA + KeratinoSens + h-CLAT + Toolbox + Lys + Cys + Avg.Lys.Cys + 6 properties (95%)	POS	NEG	POS	NEG	POS	POS	POS	POS	POS
5) KeratinoSens + h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties (95%)	POS	NEG	POS	NEG	POS	POS	POS	POS	POS
8) KeratinoSens + Toolbox + Avg.Lys.Cys + 6 properties (94%)	NEG	NEG	POS	POS	NEG	POS	POS	POS	POS
9) KeratinoSens + h-CLAT + Avg.Lys.Cys + 6 properties (92%)	POS	POS	POS	NEG	POS	POS	POS	POS	NEG
10) h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties (92%)	POS	NEG	NEG	NEG	POS	POS	NEG	POS	POS
11) KeratinoSens + h-CLAT + Toolbox + 6 properties (92%)	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS

Abbreviations: 6 properties = molecular weight, log octanol:water partition coefficient, log water solubility, log vapor pressure, melting point, boiling point; Avg.Lys.Cys = average depletion for lysine and cysteine peptides from the DPRA; Cys = average depletion of cysteine peptide; DPRA = direct peptide reactivity assay categorical response; h-CLAT = human cell line activation test; Lys = average depletion of lysine peptide from the DPRA; No. = number; SVM = support vector machine; Toolbox = read-across using QSAR Toolbox.

[a]Misclassifications, which are discordant from the murine local lymph node assay outcomes, are shaded in gray.

[b]Parentheses show the average accuracy of the test and training sets for the SVM models. Models are listed in descending order of accuracy.

[c]Prohapten. References: Jaworska *et al.* (2011) for pyridine; Gerberick *et al.* (2004) for 3,4-dihydrocoumarin; and Natsch and Haupt (2013) and Gerberick *et al.* (2009) for eugenol.

Table 8

Misclassified Substances for the Seven SVM Models with the Highest Accuracy – Test Set[a]

Model No./Variables[b]	Benzoic acid	Tartaric acid	Resorcinol[c]	Undecylenic acid	3-Aminophenol[c]	Linalool	Coumarin
7) h-CLAT + Toolbox + 6 properties (97%)	NEG	NEG	POS	POS	POS	POS	NEG
1) DPRA + KeratinoSens + h-CLAT + Toolbox + Lys + Cys + Avg.Lys.Cys + 6 properties (95%)	NEG	NEG	POS	NEG	POS	POS	NEG
5) KeratinoSens + h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties (95%)	NEG	NEG	POS	NEG	POS	POS	NEG
8) KeratinoSens + Toolbox + Avg.Lys.Cys + 6 properties (94%)	NEG	NEG	NEG	POS	NEG	POS	NEG
9) KeratinoSens + h-CLAT + Avg.Lys.Cys + 6 properties (92%)	POS	NEG	POS	POS	POS	NEG	NEG
10) h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties (92%)	NEG	NEG	NEG	POS	NEG	POS	NEG
11) KeratinoSens + h-CLAT + Toolbox + 6 properties (92%)	NEG	POS	POS	NEG	POS	POS	NEG

Abbreviations: 6 properties = molecular weight, log octanol:water partition coefficient, log water solubility, log vapor pressure, melting point, boiling point; Avg.Lys.Cys = average depletion for lysine and cysteine peptides from the DPRA; Cys = average depletion of cysteine peptide; DPRA = direct peptide reactivity assay categorical response; h-CLAT = human cell line activation test; Lys = average depletion of lysine peptide from the DPRA; No. = number; SVM = support vector machine; Toolbox = read-across using QSAR Toolbox.

[a]Misclassifications, which are discordant from the murine local lymph node assay outcomes, are shaded in gray.

[b]Parentheses show the average accuracy of the test and training sets for the SVM models. Models are listed in descending order of accuracy.

[c]Prohapten. References: Kern *et al.* (2010) for 3-aminophenol and Natsch and Haupt (2013) for resorcinol.