

Evaluating Skin Sensitization Via Soft and Hard Multivariate Modeling

International Journal of Toxicology
2020, Vol. 39(6) 547-559
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1091581820944395
journals.sagepub.com/home/ijt



Filipa A. L. S. Silva¹ , Gonçalo Brites^{2,3}, Isabel Ferreira^{2,3}, Ana Silva², Bruno Miguel Neves⁴, Jorge L. G. F. S. Costa Pereira¹, and Maria T. Cruz^{2,3}

Abstract

Allergic contact dermatitis is the most frequent manifestation of immunotoxicity in humans with a prevalence rate of 15% to 20% over general population. Skin sensitization is a complex end point that was for a long time being evaluated using animal testing. Great efforts have been made to completely substitute the use of animals and replace them by integrating data from *in vitro* and *in chemico* assays with *in silico* calculated parameters. However, it remains undefined how to make the best use of the cumulative data in such a way that information gain is maximized and accomplished with the fewest number of tests possible. In this work, 3 skin sensitization prediction models were considered: one to discriminate sensitizers from non-sensitizers, considering a 2-level scale; one according to the GHS, considering a 3-level scale; and the other to categorize potency in a 6-level scale, according to available human data. We used a data set of known human skin allergens for which *in vitro*, *in chemico*, and *in silico* descriptors were available to build classifiers based on soft and hard multivariate modeling. Model building, optimization, and refinement resulted in 100% accuracy in distinguishing between sensitizers and non-sensitizers. The same model was able to perform the characterization, in 3 and 6 levels, respectively, with 98.8 and 97.5% accuracy. Combining data from *in vitro* and *in chemico* tests with *in silico* descriptors is relatively simple to implement and some predictors are fitting the adverse outcome pathway for skin sensitization.

Keywords

skin sensitization, multivariate model, model selection, cross-validation, PLS, MOLS

Introduction

Allergic contact dermatitis (ACD) is a skin inflammatory disease caused by reactive chemicals with low molecular weight termed haptens. Impressively, there are currently more than 4,000 substances identified as skin allergens, with some of them being present on a multitude of products in consumers' households such as cosmetics, cleaning products, and fragrances.^{1,2} From a pathophysiological point of view, ACD involves the activation of both the innate and the adaptive arms of the immune system and is divided into 2 distinct phases: the initial sensitization phase and a later elicitation or challenge phase. The sensitization phase is triggered by the first contact of skin with the hapten, leading to formation of immunogenic hapten–protein complexes. These complexes are captured and processed by Langerhans cells and dermal dendritic cells (DCs), which then mature and migrate to draining lymph nodes in order to present the allergen to T lymphocytes. Then, T-cells are activated and expand into allergen-specific effector T-cells. Regarding the elicitation phase, it occurs after a second contact with the allergen, leading to a strong inflammatory response caused by the cytotoxic activity of allergen-specific effector T-cells against keratinocytes.²

ACD represents between 10% and 15% of occupational illnesses in the United States and Europe,³ and the only way to prevent symptoms is avoiding contact with allergens.⁴ Consequently, skin sensitization is a quintessentially toxicological end point to evaluate the safety of new molecules prior to their market launch. In the past, this was mostly done using animal models, like the guinea pig maximization test (GPMT) or the local lymph node assay (LLNA). However, the use of animals raised ethical and economic issues, and in 2013, their use to ensure the toxicological safety of cosmetics ingredients was

¹ Department of Chemistry, Faculty of Sciences and Technology, Coimbra Chemistry Centre, University of Coimbra, Coimbra, Portugal

² Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal

³ Faculty of Pharmacy, University of Coimbra, Health Sciences Campus, Coimbra, Portugal

⁴ Department of Medical Sciences and Institute of Biomedicine – iBiMED, University of Aveiro, Aveiro, Portugal

Corresponding Author:

Maria T. Cruz, Faculty of Pharmacy, Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal.

Email: trosete@ff.uc.pt

banished in Europe.⁵ As a result, there is a pressing need to develop new and alternative test methods to predict skin sensitization. Thankfully, the development of integrated approaches to testing and assessment and defined approaches (DA) had been fueled by the increasing mechanistic understanding of the adverse outcome pathway (AOP) for skin sensitization. Consequently, the development of *in vitro* and *in chemico* tests reflect the key events of these mechanisms: allergen covalent binding to skin proteins (KE1), keratinocyte responses, activation of DCs, and T-cell proliferation.⁶ Until now, 6 methods were approved by the Organization for Economic Co-operation and Development (OECD) for skin sensitization prediction, that contribute only for hazard identification, allowing to distinguish sensitizers from non-sensitizers. However, the prediction of potency is also of great importance for risk assessment and is imperative for the development of models with this aim. The validated models for hazard identification include the Direct Peptide Reactivity Assay (DPRA),⁷ KeratinoSens, LuSens,⁸ the human cell line activation test (h-CLAT), the U937 cell line activation test, and the interleukin-8 Luc assay.⁹ These have several disadvantages as they are costly, time-consuming, exhibit some variability, have a reduced applicability to poorly water-soluble compounds, and require the isolated substance for testing. In addition, the accuracy of these models is lower than 80%.¹⁰ On the other hand, these *in chemico* and *in vitro* assays are not designed to assess all the physicochemical and biological properties that are associated with a given chemical's capacity to cause skin sensitization. Thus, in an attempt to obtain a better performance, some of these methods were combined with each other and/or with computational methods. This strategy led to a slight increase in accuracy; however, the methods became more complex.¹¹ *In silico* methods have some advantages like their low cost, concomitantly requiring only the structure of the compound without the need of obtaining the synthesized or isolated product. Despite that, these methods are limited to organic substances with well-defined structures in order to be processed by software packages, such as TIMES-SS and E-Dragon. However, *in silico* calculated data may offer a valuable complement to the current battery of *in chemico/in vitro* tests. Indeed, there is a strong need to identify new ways to combine the data sets from different assays in order to obtain a robust and powerful approach able to identify and categorize skin sensitizers.

Data sets are often multivariate systems with several independent variables (predictors) and a dependent variable (target; skin sensitization), where predictors are used to generate a decision rule able to describe the response. However, in systems with a high number of variables not all are necessarily relevant to the description of the response and can potentially be excluded. Thus, multivariate analysis is a helpful tool because it allows the condensation of the information in order to facilitate an easier representation and interpretation, condense variables, perform a data space analysis, and remove unexplained variations such as ones caused by noise. Principal component analysis (PCA) and hierarchical clustering analysis (HCA) are essentially unsupervised exploratory methods, omitting the

response from the initial data set, they provide information about the organization of samples among the multivariate system. Linear discriminant analysis (LDA) and partial least squares (PLS) are considered supervised methods since they require response values to classify and predict new situations. PCA is more often used for object representation while HCA reveals object similarities¹²; LDA is more suited for classification purposes while PLS is used for multivariate response description.^{13,14} Either LDA and PLS can describe original variables as factors (linear combinations of variables that maximize the intergroup discrimination or correlation), which will complicate the final interpretation in terms of variable impact on model response.¹⁵⁻¹⁸ For that reason, Multivariate Ordinarily Least Squares (MOLS) is the preferred algorithm to describe the responses by assuming a multivariate polynomial response, it is possible to combine all the advantages of least squares statistics and the possibility to directly relate the response with each original variable accessing all relevant dependencies. However, in order to avoid over-fitting, it is imperative to perform a careful model selection and validation to preserve a parsimonious and more robust model.¹⁹⁻²²

In view of all this, the aim of this study was to develop models for hazard identification, distinguishing sensitizers from non-sensitizers (2 level response), and to also able to ensure the categorization of chemicals according to the Globally Harmonized System of Classification and Labeling of Chemicals (GHS), considering skin sensitizers category 1A, skin sensitizers category 1B and non-sensitizers (3 level response), and 6 classes (6 level response), using experimental (*in vitro* and *in chemico* descriptors) and computational (*in silico*) information. The relevance and correlation of the information gathered in the models with the skin sensitization response, more specifically with the AOP, is also discussed.

Experimental

Data Sets

The initial data set was based on the work of Basketter and colleagues that identified 131 chemicals organized by 6 levels of potency according to the available human information.²³ This classification is based on the probability of these compounds to induce skin sensitization in humans. Category 1 compounds are the most potent ones and they are considered extreme sensitizers. Categories 2, 3, and 4 contain less potent compounds, designated strong, moderate, and weak sensitizers, respectively. Categories 5 and 6 included chemicals showing a low capacity to induce skin sensitization. Thus, categories 1 to 4 are considered sensitizers and 5 and 6 are considered non-sensitizers.²³ A 3 category classification system was also included according to the classification defined by the GHS. This system considers one category for skin sensitizers that includes 2 subcategories (1A and 1B) and a non-sensitizers category. Subcategory 1A compounds are extreme and strong skin sensitizers and subcategory 1B chemicals correspond to moderate and weak sensitizers.²⁴ Doing the parallelism with the previous classification system,

the 3 GHS categories correspond to the combination of categories 1 and 2 (subcategory 1A), categories 3 and 4 (subcategory 1B), and categories 5 and 6 (non-sensitizers).^{23,25}

For this study, compounds belonging to each category have been selected, excluding metals and compounds for which information about all descriptors was not available. Experimental (*in chemico* and *in vitro*) and computational (*in silico*) descriptors were considered and used as predictors. The *in chemico* predictors included cysteine (CysDep) and lysine (LysDep) depletion from DPRA whose values were retrieved from reference published data sets.^{10,26–28} The concentration that reduces 30% cell viability (EC30) was determined in house (*in vitro* descriptor), using the Alamar Blue assay.²⁹ THP-1 cells were exposed for 24 hours to several concentrations of the considered chemicals in a dose–response experiment and analyzed for viability by the reduction of resazurin. The EC30 values were calculated by linear regression of obtained data. A protein adduct formation descriptor obtained by the TIMES-SS package software²⁸ was used as an *in silico* predictor alongside constitutional, topological charge indices, functional group counts, topological, charge, geometrical, molecular properties, walk and path counts, information indices, and atom-centered fragments descriptors calculated through the VCLABS E-Dragon, an online free software.^{30,31} The electrophilicity index (ω) has also been included since previous studies have shown its relevance for the description of the skin sensitization response.³² To achieve this goal, the energy of highest occupied (E_{HOMO}) and lowest unoccupied (E_{LUMO}) molecular orbitals have been obtained from geometry optimization calculations, using density functional theory (B3LYP/6-31G) and the values obtained were related by the following equation 1 described by Parr and colleagues³³:

$$\omega = \frac{\left(\frac{E_{\text{HOMO}} + E_{\text{LUMO}}}{2}\right)^2}{2(E_{\text{LUMO}} - E_{\text{HOMO}})} \quad (1)$$

The explanation of the descriptors can be consulted in supporting information.

Test Cases

In order to evaluate the suitability and predictive ability of chemicals to induce skin sensitization, 3 test cases were studied based on the predictors mentioned above. The *in vitro* and *in chemico* test case included the 3 experimental descriptors above indicated (case1: $n = 81$ samples, $m = 3$ variables), the *in silico* test case was based exclusively on descriptors calculated computationally (case2: $n = 90$ samples, $m = 383$ variables), and the last test case used a combination of experimental and computational variables (*in vitro* + *in chemico* + *in silico*; case3: $n = 81$, $m = 386$). The second test case (exclusively computational) presents the advantage of not requiring any laboratory testing. These 3 data sets can be consulted in supplementary material.

Data Processing

Direct relations were explored between a pool of variables (predictors) and the skin sensitization response transduced in discrete levels of potency as a homogeneously distributed numerical vector.

By default, variables with null variance were removed and the remaining variables were standardized to null mean and unitary variance to avoid scale effects. PLS models were explored in Octave (MatLab compatible program).³⁴

Model Refinement and Selection

MOLS model refinement was performed with R-project resources,³⁵ using forward and forward-backward model based on Akaike Information Criterion (AIC) eq. (2), and Bayesian Information Criterion (BIC), eq. (3):

$$\text{AIC} = 2p - 2\log(\hat{L}) \quad (2)$$

$$\text{BIC} = \log(n)p - 2\log(\hat{L}) \quad (3)$$

where n corresponds to the number of samples and $\hat{L}$ to the Likelihood function of a given fit for the model with p parameters.^{36,37}

Several merit functions were used to evaluate each model fitting and predictive ability such as: squared Pearson multivariate correlation (R^2), adjusted squared Pearson's multivariate correlation (R^2_{adj}), mean fitting error (σ_{fit}), fitting accuracy for 6-level response description (FA6), fitting accuracy for 3-level response description (FA3), fitting accuracy for 2-level response description (FA2), leave-one-out cross validation mean prediction error (CV[LOO]), 10-fold cross validation mean prediction error (CV[10fold]), and representative 1/5th cross validation mean prediction error by systematic preservative resampling of 1/5 of samples (keeping all categories present either in calibrating and testing data sets, CV[1/5]).

Results

Case1: In Vitro and In Chemico Approach

For case1, we tried to access data structures only by using *in vitro* and *in chemico* information. This data set consists of $n = 81$ compounds (samples) with skin sensitization accessed via 3 ($m = 3$) experimental descriptors (“CysDep,” “LysDep,” and “EC30”). PCA allows for 80.2% of information recovery with $f = 2$ factors (and 100% recovery with $f = 3$ factors). Figure 1 represents the corresponding 3-dimensional scores plot.

From Figure 1, it is possible to conclude that evaluation of the potency of sensitizers according to 6 or 3 categories may not be feasible by using only these 3 *in vitro* and *in chemico* descriptors. We can also infer that “CysDep” seems to be the most discriminating variable since there is a tendency to allocate more potent skin sensitizers to higher scores, while less toxicity generally corresponding to lower ones. There is also a tendency to have all results aligned in a single line parallel to “CysDep” axis, which also reveals that this is the most relevant variable from this pool

of experimental descriptors in terms of skin sensitization hazard. In the same way, “LysDep” and “EC30” seem to be the less discriminating variables that do not allow clear distinction in 2, 3, or 6 categories. The dispersion of the objects over these axes is lower, which indicates a lower contribution of these variables in the description of the system.

Previous work demonstrated that using Ward method as an association criterion for Euclidean distance similarity could easily evidence data structure.¹⁷ By performing HCA with Ward’s method on standardized variables of case1 data set, we obtained the dendrogram pictured in Figure 2.

Two distinct groups can be easily identified (group 1: on the left; group 2: on the right). Group 1 contains 66.7 and group 2 only 33.3% of the total number of samples. In a first attempt, this looks very promising since we have 56 sensitizers (69.1%) and 25 non-sensitizers (30.9%) in this data set. However, a closer inspection into these 2 groups revealed that the first one consists of similar percentages of sensitizers and non-sensitizers, but the second one mostly consists of sensitizers. In addition, the less populated group, group 2, includes most of categories 1 (extreme skin sensitizers) and 2 (strong skin sensitizers)

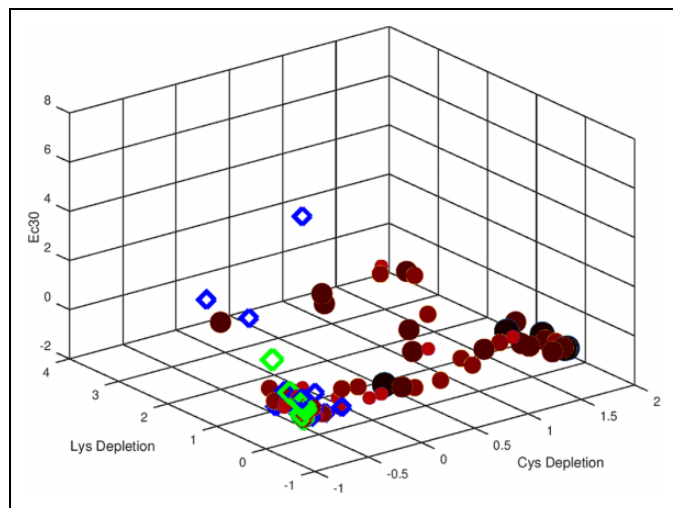


Figure 1. Three-dimensional scores representation of standardized results obtained with *in vitro* and *in chemico* data. Sensitizers are represented in filled reddish circles (darker colors and grater dimensions for more toxic) and non-sensitizers in open diamonds (green and blue colors).

chemicals, whereas group 1, the most populated, contains mostly categories 4 (weak skin sensitizers), 5, and 6 (non-sensitizers) compounds. Category 3 (moderate skin sensitizers) chemicals are evenly included in both groups. Similar classification problems are obtained in the case of 3-level potency classification.

In order to better verify how samples are clustered inside these 2 big groups, 6 smaller groups were considered (Figure 2B). All groups have compounds from 3 or 4 categories demonstrating a large overlap between categories. Furthermore, all groups contain chemicals from adjacent categories, except group 6 that includes compounds from categories 2, 3, and 5.

Furthermore, by performing LDA with this data set, the results were not satisfactory. Best results were obtained via cross validation with LDA following a one-fifth conservative resampling (80% of samples as calibrating set and 20% as testing set, ensuring that all categories are represented in both data sets) which allowed to obtain 61.1% prediction accuracy for hazard identification response (2-level classification: sensitizer or non-sensitizer).

By using MOLS in this system, we only must consider 7 polynomial first-degree models from $P = 4$ (all 3 *in vitro* and *in chemico* descriptors plus a constant, t01) to $P = 2$ (simpler model considering each variable individually, t05-t07). Model performance results are presented in Table 1.

By using only experimental data from *in vitro* and *in chemico* assays, detailed in Table 1, the results indicate that all tested models poorly describe the skin sensitization response either for hazard identification or for potency categorization. Best predictive models for 6-level toxicity response are t05 (“CysDep”) and t02 (“CysDep” and “LysDep”) with very low fitting and predicting accuracy of 34.6% and 30.9%, respectively (corresponding to 53 and 56 fails in 81). Better results are obtained in the case of 3-level potency classification with a fitting ability of 80.2% but a bad result in terms of predicting by achieving only 45.7% accuracy. In the case of 2-level response, best fitting and predicting models are t06 (“EC30”) and t04 (“LysDep” and “EC30”) with better fitting ability (100.0% and 65.4%) but still low predicting ability (69.1%), respectively.

From all these results, we can conclude that it is not possible to fully describe and predict chemical skin sensitization based

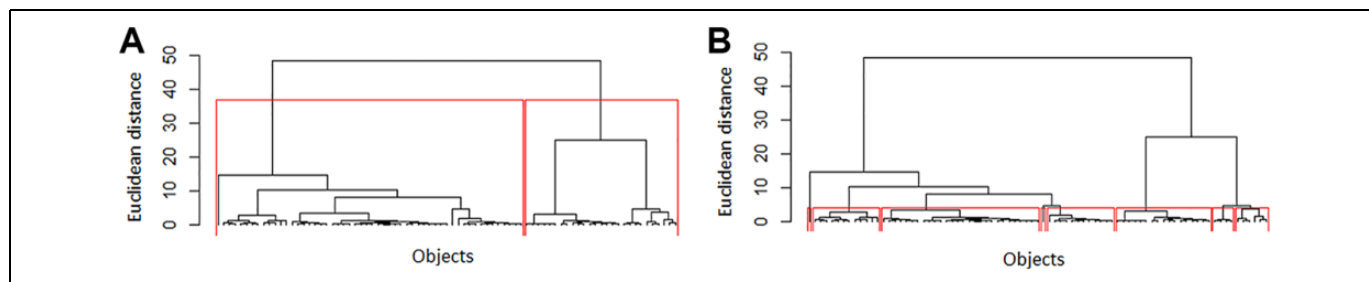


Figure 2. Dendrogram of case1 standardized variables obtained with Ward’s association method for Euclidean distances. Considerations by assuming 2 cluster formation (A) and 6 cluster formation (B).

Table 1. Characterization and Performance of Obtained Models in Case1 (In Vitro and In Chemico).

Code	p	R ²	R ² adj	AIC	BIC	σ_{fit}	FA2	FA3	FA6	CV(LOO)	CV(10-fold)	CV(1/5)	PA2	PA3	PA6
t01	4	0.4680	0.4473	245.2	257.2	1.06	74.1	69.1	32.1	1.19	1.16	1.20	65.4	50.6	30.9
t02	3	0.4672	0.4536	243.3	252.9	1.05	77.8	71.6	34.6	1.17	1.13	1.19	65.4	49.4	30.9
t03	3	0.4649	0.4512	243.7	253.2	1.06	80.2	72.8	35.8	1.15	1.15	1.10	65.4	50.6	27.2
t04	3	0.0690	0.0452	288.5	298.1	1.39	65.4	58.0	28.2	2.13	2.04	1.37	69.1	50.6	23.5
t05	2	0.4641	0.4573	241.8	249.0	1.05	77.8	58.0	34.6	1.13	1.12	1.10	66.7	46.9	30.9
t06	2	0.0402	0.0280	289.0	296.2	1.41	100.0	80.2	23.5	2.03	2.01	1.38	69.1	45.7	23.5
t07	2	0.0340	0.0218	289.5	296.7	1.41	69.1	48.1	21.0	2.15	2.08	1.40	69.1	40.7	21.0

Abbreviations: AIC, Akaike criterion; BIC, Bayesian criterion; Code, model name; CV(LOO), predicted root mean square error obtained with Leave-one-Out cross validation; CV(10-fold), predicted root mean square error obtained by $k = 10$; CV(1/5), predicted root mean square error using 20% of representative sample; FA3 and FA6, fitting accuracy using 3-level and 6-level potency classification; FA2, fitting accuracy obtained with 2-level classification; p, number of parameters; PA2, predicting accuracy obtained with 2-level classification; PA3 and PA6, predicting accuracy using 3-level and 6-level potency classification; σ_{fit} —mean squared model error; R², Pearson coefficient of determination; R²adj, adjusted coefficient of determination.

Table 2. Most Relevant Variables Detected With PLS in Case2 (In Silico Descriptors) and Case3 (In Vitro, In Chemico, and In Silico Descriptors) by Direct and Weighted Criteria.^a

Case	Approach	#	List of most relevant variables
Case2	Direct	12	TSSAA, O-058, AAC, IC0, nR = Cs, C-016, IE, nArCOOH, AMW, BIC0, nArOH, nDB
	Weighted	12	TSSAA, nArCOOH, nR = Cs, C-016, nDB, nArNH2, N-069, IE, nRCHO, C-036, AMW, Me
	Global	17	TSSAA, IE, AMW, Me, nDB, nR = Cs, nArCOOH, nRCHO, nArNH2, nArOH, AAC, IC0, BIC0, C-016, C-036, O-058, N-069
Case3	Direct	6	CysDep, IE, AMW, nDB, MLOGP, O-058
	Weighted	18	CysDep, O-058, nArCOOH, nArNH2, N-069, MAXDN, Me, nR = Cs, C-016, nArCOOR, nArOH, O-057, IC4, SIC4, BIC4, CIC4, ASP, C-040
	Global	22	CysDep, IE, AMW, Me, nDB, MAXDN, ASP, nR = Cs, nArCOOH, nArCOOR, nArNH2, nArOH, MLOGP, IC4, SIC4, CIC4, BIC4, C-016, C-040, O-057, O-058, N-069

Abbreviations: AMW, average molecular weight; IE, Electrophilicity index; Me, Mean atomic Sanderson electronegativity (scale on Carbon atom); nDB, number of double bonds; nR = Cs, number of aliphatic secondary C(sp²); nArCOOH, number of carboxylic acids (aromatic); nArNH2, number of primary amines (aromatic); nArOH, number of aromatic hydroxyls; C-016 - =CHR; N-069, Ar-NH₂/X-NH₂; TSSAA, protein adduct formation.

^aSee table of descriptors in supporting information.

exclusively in these *in vitro* and *in chemico* results and new approaches were tested subsequently (case2 and case3).

Variable Selection

After this preliminary approach, we tried to include more information into our initial system by adding extra *in silico* variables. As more variables were added to the initial system, the more confusing it becomes—PCA was unable to retrieve 80% in a reasonable number of factors ($f = 11$ for case2 and $f = 12$ for case3) and also HCA and LDA yielded worse results than previous ones. The main reason for this failure may be related to the fact that the extra variables can contribute some useful information, while some variables are increasing system information that is not related to the studied response (skin sensitization).

We then tested a new approach starting by selecting the relevant variables that can help to describe skin sensitization. To achieve this goal, we used PLS as soft-modeling approach in order to identify more relevant variables and then, by using model selection criteria with MOLS, we developed several predictive models. In our soft-modeling experience with PLS,^{15,26} we found that this algorithm tends to preserve the same latent

factor variables that may be collinear (have similar information) while they are related to target response—these variables will show similar impact (as loadings). In order to select only most relevant variables, we established 2 criteria, one direct and the other weighted. The direct criterion is exclusively based on variable impact over each latent factor—in this way, we selected the most relevant contributions (with loadings above the mean) along the first 5 latent factors.²² The weighted approach was based on crossing loading information with the described response (amount of justified variability)—this way we are able to emphasize the impact of variables by considering their contribution in the description of the response. We kept in mind that these criteria will not solve the collinearity problem between selected variables, but this gap will be resolved with MOLS, via model refinement (least significant variables removal).

In Table 2, we describe the most relevant variables by detected direct and weighted assessment.

Surprisingly, in both cases (case2 and case3), PLS soft-modeling approach pointed out that only a very small fraction of tested variables (17 in 383 [4.4%] and 22 in 386 [5.7%]) were detected as very relevant. The only evidenced experimental parameter was “CysDep,” either by direct and weighted approaches; “TSSAA” was only detected in *in silico* approach;

Table 3. Characterization and Performance of Obtained Models in Case2 (In Silico).

Code	p	R ²	R ² adj	AIC	BIC	σ_{fit}	FA2	FA3	FA6	CV(LOO)	CV(10-fold)	CV(1/5)	PA2	PA3	PA6
s01	79	0.9997	0.9980	-261.3	-61.4	0.07	100.0	100.0	100.0	0.54	0.75	322.86	91.1	86.7	83.3
s02	79	0.9998	0.9987	-302.7	-102.7	0.05	100.0	100.0	100.0	0.23	0.35	37.30	95.6	90.0	86.7
s03	79	0.9996	0.9967	-218.9	-18.9	0.08	100.0	100.0	100.0	0.04	0.19	20.44	100.0	95.6	91.1
s04	77	0.9998	0.9985	-278.0	-83.0	0.06	100.0	100.0	100.0	0.16	0.34	46.59	98.9	96.7	93.3
s05	77	0.9998	0.9986	-284.6	-89.6	0.06	100.0	100.0	100.0	0.07	0.13	52.68	98.9	96.7	94.4
s06	77	0.9998	0.9988	-300.4	-105.5	0.05	100.0	100.0	100.0	0.23	0.23	84.07	95.6	90.0	86.7
s07	77	0.9996	0.9971	-219.7	-24.7	0.08	100.0	100.0	100.0	0.04	0.21	63.95	98.9	92.2	84.4
s08	77	0.9998	0.9986	-284.6	-89.6	0.06	100.0	100.0	100.0	0.07	0.13	47.34	98.9	95.6	94.4
s09	76	0.9997	0.9984	-265.7	-73.2	0.06	100.0	100.0	100.0	0.54	0.54	137.18	91.1	90.0	85.6
s10	76	0.9996	0.9973	-222.3	-29.8	0.08	100.0	100.0	100.0	0.11	0.14	302.51	97.8	94.4	93.3
s11	75	0.9997	0.9981	-246.1	-56.1	0.06	100.0	100.0	100.0	1.00	1.45	942.73	95.6	93.3	86.7
s12	75	0.9997	0.9981	-246.1	-56.1	0.06	100.0	100.0	100.0	1.00	1.45	942.73	95.6	92.2	86.7
s13	75	0.9994	0.9962	-186.4	3.6	0.09	100.0	100.0	100.0	1.14	2.91	142.01	96.7	93.3	88.9
s14	75	0.9997	0.9981	-246.1	-56.1	0.06	100.0	100.0	100.0	1.00	1.45	942.73	95.6	92.2	86.7
s15	75	0.9997	0.9981	-246.1	-56.1	0.06	100.0	100.0	100.0	1.00	1.46	947.26	95.6	91.1	86.7
s16	75	0.9997	0.9981	-246.1	-56.1	0.06	100.0	100.0	100.0	1.00	1.42	947.06	95.6	91.1	86.7
s17	75	0.9994	0.9967	-198.4	-8.5	0.08	100.0	100.0	100.0	0.06	0.24	20.17	97.8	90.0	82.2
s18	74	0.9998	0.9987	-280.5	-93.1	0.05	100.0	100.0	100.0	0.13	0.26	38.38	98.9	97.8	93.3
s19	73	0.9996	0.9978	-227.9	-42.9	0.07	100.0	100.0	100.0	0.11	0.15	55.84	98.9	96.7	95.6
s20	72	0.9996	0.9979	-229.1	-46.6	0.07	100.0	100.0	100.0	0.35	1.53	111.10	95.6	93.3	88.9
s21	72	0.9996	0.9979	-229.1	-46.6	0.07	100.0	100.0	100.0	0.15	1.05	91.37	95.6	92.2	88.9
s22	72	0.9996	0.9979	-229.1	-46.6	0.07	100.0	100.0	100.0	2.35	0.15	111.10	95.6	93.3	88.9
s23	72	0.9996	0.9979	-229.1	-46.6	0.07	100.0	100.0	100.0	2.35	0.15	111.10	95.6	91.1	88.9
s24	72	0.9996	0.9979	-229.1	-46.6	0.07	100.0	100.0	100.0	2.35	0.15	111.10	95.6	93.3	88.9
s25	71	0.9986	0.9933	-121.3	58.7	0.12	100.0	100.0	100.0	0.15	0.26	192.97	96.7	92.2	86.7
s26	71	0.9997	0.9988	-278.1	-98.2	0.05	100.0	100.0	100.0	0.06	0.10	5.09	98.9	97.8	95.6
s27	71	0.9997	0.9988	-278.1	-98.2	0.05	100.0	100.0	100.0	0.06	0.12	7.29	98.9	97.8	95.6
s28	70	0.9993	0.9971	-194.4	-16.9	0.08	100.0	100.0	100.0	0.11	0.14	5.49	98.9	96.7	92.2
s29	69	0.9985	0.9938	-124.1	50.9	0.12	100.0	100.0	100.0	0.12	0.16	27.25	96.7	94.4	88.9
s30	68	0.9993	0.9972	-193.1	-20.6	0.08	100.0	100.0	100.0	0.71	1.38	36.51	95.6	94.4	88.9
s31	67	0.9990	0.9960	-158.3	11.7	0.09	100.0	100.0	100.0	3.63	2.16	45.42	95.6	93.3	85.6
s32	67	0.9990	0.9960	-158.3	11.7	0.09	100.0	100.0	100.0	3.63	2.16	45.42	95.6	93.3	85.6
s33	60	0.9958	0.9875	-46.4	106.1	0.16	100.0	100.0	100.0	0.12	0.28	8.30	94.4	90.0	80.0
s34	51	0.9633	0.9162	130.7	260.7	0.43	100.0	100.0	92.2	0.16	0.17	1.90	85.6	71.1	46.7

Abbreviations: AIC, Akaike criteria; BIC, Bayesian criteria; CV(LOO), predicted root mean square error obtained with Leave-one-Out cross validation; Code, model name; CV(10-fold), predicted error obtained by $k = 10$; CV(1/5), predicted error using 20% of representative sample; FA3 and FA6, fitting accuracy using 3-level and 6-level toxicity classification; FA2, fitting accuracy obtained with 2-level classification; p, number of parameters; PA-3 and PA6, predicting accuracy using 3-level and 6-level potency categories; PA2, predicting accuracy obtained with 2-level classification; R², Pearson coefficient of determination; R²adj, adjusted coefficient of determination; σ_{fit} , mean squared model error.

and several predictors were detected in both cases (IE, AMW, Me, nDB, the number of terminal primary carbons sp² [nR = Cs], nArCOOH, nArNH₂, nArOH, C-016, O-058, N-069).

Model Selection

PLS is extremely robust to deal with multivariate systems ($n \ll m$) but presents a huge difficulty in what concerns interpretation based on original variables. Since we are interested in understanding sensitizing potency, we preferred to use hard-modeling with MOLS to keep original response dependencies. With these PLS-selected variables, explained in Table 2, we developed 18 (17 dependencies and a constant) and 23 (22 dependencies and a constant) starting models with $p = 1$ initial parameter in MOLS and performed AIC and BIC optimization in ascending mode ("forward"). After convergence, the selected models were simplified by "backward" and "backward-forward"

methodologies. With this process, we ended with 34 and 48 models for case2 and case3, respectively. Model performance evaluation is summarized in Tables 3 and 4.

From Tables 3 and 4, we see that all MOLS models are extremely accurate in describing the response (for most models R² and R²adj > 0.999 and FA6 = FA3 = FA2 = 100%). However, these models present several dependencies that may be responsible for their difficulty in predicting new results. By comparing predicting cross-validation errors, from CV(LOO) to CV(1/5), predictions are extremely dependent upon training set dimensions, suggesting that these models are able to increase their prediction ability as the training set increases.

Analyzing the results for case2 (*in silico* data; Table 3) and case3 (*in vitro*, *in chemico* and *in silico* data; Table 4), we selected model "s26" ($p = 71$) and model "m20" ($p = 66$) as the best representative models, respectively. Model "s26" presents a predicting accuracy of 95.6% (6 fails in 90), 97.8% (2

Table 4. Characterization and Performance of Obtained Models in Case3 (In Vitro, In Chemico, and In Silico).

Code	p	R ²	R ² adj	AIC	BIC	σ_{fit}	FA2	FA3	FA6	CV(LOO)	CV(10-fold)	CV(1/5)	PA2	PA3	PA6
m01	72	0.9998	0.9984	-266.4	-91.6	0.06	100.0	100.0	100.0	0.15	0.27	16.63	95.1	88.9	81.5
m02	71	0.9997	0.9979	-238.7	-66.3	0.06	100.0	100.0	100.0	2.44	7.13	52.26	95.1	86.4	75.3
m03	71	0.9998	0.9985	-262.3	-89.9	0.06	100.0	100.0	100.0	3.79	8.39	48.74	97.5	92.6	86.4
m04	70	0.9996	0.9974	-214.8	-44.8	0.07	100.0	100.0	100.0	0.96	1.43	618.07	95.1	90.1	80.2
m05	70	0.9998	0.9984	-251.8	-81.8	0.06	100.0	100.0	100.0	0.57	0.71	132.88	95.1	91.4	86.4
m06	69	0.9998	0.9988	-272.4	-104.8	0.05	100.0	100.0	100.0	0.07	0.10	100.72	98.8	98.8	95.1
m07	69	0.9998	0.9985	-251.9	-84.3	0.06	100.0	100.0	100.0	0.10	0.23	128.59	95.1	91.4	88.9
m08	69	0.9997	0.9978	-223.7	-56.1	0.07	100.0	100.0	100.0	0.64	2.52	733.07	100.0	95.1	92.6
m09	69	0.9998	0.9985	-255.2	-87.6	0.05	100.0	100.0	100.0	1.89	10.81	175.46	95.1	92.6	92.6
m10	69	0.9998	0.9985	-255.1	-87.5	0.05	100.0	100.0	100.0	2.93	10.05	181.04	95.1	93.8	92.6
m11	69	0.9998	0.9988	-272.4	-104.8	0.05	100.0	100.0	100.0	0.07	0.10	100.72	98.8	98.8	95.1
m12	69	0.9996	0.9971	-199.7	-32.1	0.08	100.0	100.0	100.0	2.56	5.24	588.81	93.8	92.6	87.7
m13	68	0.9996	0.9978	-217.2	-52.0	0.07	100.0	100.0	100.0	0.35	1.25	84.12	95.1	92.6	85.2
m14	68	0.9997	0.9984	-241.9	-76.7	0.06	100.0	100.0	100.0	0.90	1.68	291.92	93.8	92.6	92.6
m15	68	0.9994	0.9962	-174.2	-9.0	0.09	100.0	100.0	100.0	2.20	3.25	148.48	95.1	88.9	84.0
m16	67	0.9996	0.9974	-202.1	-39.3	0.07	100.0	100.0	100.0	0.90	1.68	437.82	93.8	96.3	92.6
m17	67	0.9996	0.9974	-202.1	-39.3	0.07	100.0	100.0	100.0	0.86	1.05	88.58	93.8	90.1	87.7
m18	67	0.9994	0.9965	-177.6	-14.7	0.08	100.0	100.0	100.0	0.12	0.18	5.89	96.3	92.6	92.6
m19	67	0.9994	0.9965	-176.1	-13.3	0.08	100.0	100.0	100.0	2.43	6.71	16.87	93.8	87.7	86.4
m20	66	0.9997	0.9986	-247.3	-86.9	0.05	100.0	100.0	100.0	0.05	0.05	2.81	98.8	97.5	96.3
m21	66	0.9997	0.9986	-247.4	-87.0	0.05	100.0	100.0	100.0	0.10	0.20	3.25	96.3	92.6	92.6
m22	66	0.9995	0.9976	-202.5	-42.1	0.07	100.0	100.0	100.0	0.21	0.25	3.42	97.5	90.1	88.9
m23	66	0.9996	0.9980	-217.4	-57.0	0.06	100.0	100.0	100.0	1.00	1.23	19.72	96.3	87.7	84.0
m24	66	0.9990	0.9946	-137.5	23.0	0.11	100.0	100.0	100.0	4.67	4.18	55.70	96.3	92.6	87.7
m25	65	0.9996	0.9980	-216.7	-58.6	0.06	100.0	100.0	100.0	0.27	1.11	10.90	95.1	92.6	88.9
m26	65	0.9997	0.9986	-242.5	-84.5	0.05	100.0	100.0	100.0	0.71	1.08	43.68	96.3	95.1	95.1
m27	65	0.9997	0.9985	-239.5	-81.5	0.05	100.0	100.0	100.0	3.97	7.11	43.56	93.8	90.1	86.4
m28	65	0.9997	0.9985	-239.5	-81.4	0.05	100.0	100.0	100.0	5.52	6.65	42.88	93.8	88.9	86.4
m29	65	0.9997	0.9986	-242.5	-84.5	0.05	100.0	100.0	100.0	0.71	1.08	43.49	96.3	96.3	95.1
m30	64	0.9994	0.9973	-188.9	-33.2	0.07	100.0	100.0	100.0	0.95	1.50	31.96	96.3	87.7	81.5
m31	64	0.9996	0.9982	-219.0	-63.3	0.06	100.0	100.0	100.0	0.13	0.14	1.47	96.3	91.4	87.7
m32	64	0.9994	0.9973	-189.7	-34.0	0.07	100.0	100.0	100.0	1.33	0.87	38.45	98.8	95.1	93.8
m33	63	0.9995	0.9978	-202.5	-49.3	0.07	100.0	100.0	100.0	0.85	0.85	9.60	96.3	90.1	91.4
m34	63	0.9996	0.9983	-224.6	-71.3	0.06	100.0	100.0	100.0	0.75	1.51	12.04	95.1	87.7	86.4
m35	63	0.9993	0.9970	-177.7	-24.4	0.08	100.0	100.0	100.0	0.12	0.23	1.63	97.5	92.6	92.6
m36	63	0.9993	0.9970	-177.0	-23.7	0.08	100.0	100.0	100.0	7.47	9.93	18.26	97.5	91.4	92.6
m37	63	0.9993	0.9970	-176.9	-23.7	0.08	100.0	100.0	100.0	0.11	2.74	76.50	97.5	92.6	91.4
m38	63	0.9994	0.9975	-192.7	-39.5	0.07	100.0	100.0	100.0	2.44	2.66	24.13	93.8	88.9	87.7
m39	62	0.9995	0.9981	-210.2	-59.3	0.06	100.0	100.0	100.0	0.27	0.06	1.49	100.0	96.3	96.3
m40	62	0.9995	0.9978	-198.9	-48.1	0.07	100.0	100.0	100.0	0.20	0.25	2.59	97.5	91.4	88.9
m41	62	0.9992	0.9967	-167.6	-16.8	0.08	100.0	100.0	100.0	0.27	3.00	85.22	98.8	87.7	88.9
m42	62	0.9996	0.9981	-213.4	-62.6	0.06	100.0	100.0	100.0	0.08	0.22	2.36	100.0	96.3	90.1
m43	61	0.9991	0.9964	-157.1	-8.6	0.09	100.0	100.0	100.0	5.85	13.89	50.60	97.5	91.4	88.9
m44	61	0.9994	0.9974	-184.6	-36.2	0.07	100.0	100.0	100.0	0.97	1.33	31.36	98.8	88.9	91.4
m45	61	0.9992	0.9967	-164.2	-15.8	0.08	100.0	100.0	100.0	0.88	5.07	21.15	93.8	86.4	87.7
m46	60	0.9993	0.9974	-183.5	-37.5	0.07	100.0	100.0	100.0	0.20	0.24	2.53	96.3	87.7	87.7
m47	59	0.9992	0.9970	-167.9	-24.3	0.08	100.0	100.0	100.0	1.08	0.93	34.81	98.8	95.1	93.8
m48	58	0.9988	0.9958	-139.7	1.6	0.09	100.0	100.0	100.0	3.65	3.96	38.39	97.5	92.6	91.4

Abbreviations: AIC, Akaike criteria; BIC, Bayesian criteria; CV(LOO), predicted root mean square error obtained with Leave-one-Out cross validation; Code, model name; CV(10-fold), predicted error obtained by $k = 10$; CV(1/5), predicted error using 20% of representative sample; FA3 and FA6, fitting accuracy using 3-level and 6-level toxicity classification; FA2, fitting accuracy obtained with 2-level classification; p, number of parameters; PA-3 and PA6, predicting accuracy using 3-level and 6-level potency categories; PA2, predicting accuracy obtained with 2-level classification; R^2 , Pearson coefficient of determination; R^2_{adj} , adjusted coefficient of determination; σ_{fit} , mean squared model error.

fails in 90), and 98.9% (1 fail in 90) for 6-level, 3-level, and 2-level toxicity response, respectively. In turn, model “m20” was able to predict 96.3% (3 fails in 81), 97.5% (2 fails in 81), and 98.8% (1 fail in 81) of same 6-level, 3-level, and 2-level

responses, respectively. Considering these results, we can conclude that the experimental information is crucial to increase prediction with simpler models (7.0% reduction in the number of parameters).

Table 5. Performance Evolution of Model “s26” (Case2) by Successive Removal of Least Significant Parameters.

p	R ²	R ² adj	AIC	BIC	σ_{fit}	FA2	FA3	FA6	CV(LOO)	CV (10-fold)	CV (1/5)	PA2	PA3	PA6
71	0.9997	0.9988	-278.1	-98.2	0.05	100.0	100.0	100.0	0.06	0.10	5.09	98.9	97.8	95.6
70	0.9993	0.9971	-194.4	-16.9	0.08	100.0	100.0	100.0	1.15	1.77	5.49	98.9	96.7	92.2
68	0.9989	0.9955	-150.5	21.9	0.10	100.0	100.0	100.0	0.97	1.13	5.69	98.9	96.7	87.8
66	0.9984	0.9942	-123.0	44.5	0.11	100.0	100.0	100.0	1.05	1.08	1.77	97.8	93.3	83.3
64	0.9979	0.9927	-99.1	63.4	0.13	100.0	100.0	100.0	0.30	0.29	0.89	97.8	94.4	85.6
62	0.9965	0.9889	-58.8	98.7	0.16	100.0	100.0	100.0	0.34	0.66	1.05	97.8	93.3	87.8
60	0.9950	0.9852	-31.3	121.2	0.18	100.0	100.0	100.0	0.25	0.46	1.00	97.8	90.0	82.2
58	0.9934	0.9816	-9.6	137.9	0.20	100.0	100.0	100.0	0.23	0.52	1.17	96.7	88.9	82.2
56	0.9923	0.9798	0.3	142.8	0.21	100.0	100.0	100.0	0.48	1.22	1.10	96.7	90.0	80.0
54	0.9905	0.9765	15.0	152.5	0.23	100.0	100.0	98.9	0.53	0.88	1.04	96.7	91.1	84.4
52	0.9886	0.9733	27.5	160.0	0.24	100.0	100.0	100.0	0.59	1.13	1.10	96.7	88.9	77.8
50	0.9874	0.9721	32.1	159.5	0.25	100.0	100.0	97.8	0.43	0.48	1.06	95.6	86.7	78.9
48	0.9843	0.9667	48.3	170.8	0.27	100.0	100.0	97.8	0.28	0.27	1.34	95.6	87.8	75.6
46	0.9785	0.9566	72.4	189.9	0.31	100.0	98.9	96.7	0.73	0.61	1.59	91.1	83.3	66.7
44	0.9706	0.9430	96.8	209.3	0.35	100.0	98.9	94.4	1.26	1.57	2.15	87.8	78.9	67.8
42	0.9622	0.9299	115.3	222.8	0.39	96.7	94.4	93.3	1.76	2.19	1.91	86.7	80.0	65.6
40	0.9555	0.9208	125.9	228.4	0.41	95.6	93.3	88.9	0.88	1.20	2.08	86.7	75.6	62.2

Abbreviations: AIC, Akaike criteria; BIC, Bayesian criteria; CV(LOO), predicted root mean square error obtained with Leave-one-Out cross validation; Code, model name; CV(10-fold), predicted error obtained by k = 10; CV(1/5), predicted error using 20% of representative sample; FA3 and FA6, fitting accuracy using 3-level and 6-level toxicity classification; FA2, fitting accuracy obtained with 2-level classification; p, number of parameters; PA-3 and PA6, predicting accuracy using 3-level and 6-level potency categories; PA2, predicting accuracy obtained with 2-level classification; R², Pearson coefficient of determination; R²adj, adjusted coefficient of determination; σ_{fit} , mean squared model error.

Table 6. Characterization and Performance of Model “m20” (Case3) During Its Simplification (Removal of Least Significant Parameter).

P	R ²	R ² adj	AIC	BIC	σ_{fit}	FA2	FA3	FA6	CV(LOO)	CV(10-fold)	CV(1/5)	PA2	PA3	PA6
66	0.9997	0.9986	-247.3	-86.9	0.05	100.0	100.0	100.0	0.21	0.22	64.81	98.8	97.5	96.3
64	0.9996	0.9981	-216.6	-61.0	0.06	100.0	100.0	100.0	0.03	0.14	1.13	100.0	98.8	96.3
62	0.9994	0.9973	-183.3	-32.5	0.07	100.0	100.0	100.0	0.05	0.37	0.87	100.0	98.8	97.5
60	0.9988	0.9953	-134.5	11.6	0.10	100.0	100.0	100.0	0.14	0.31	2.20	97.5	93.8	88.9
58	0.9982	0.9937	-107.6	33.7	0.11	100.0	100.0	100.0	1.19	2.78	2.30	97.5	92.6	88.9
56	0.9967	0.9893	-61.7	74.8	0.15	100.0	100.0	100.0	0.53	0.19	1.34	98.8	93.8	90.1
54	0.9951	0.9855	-34.5	97.2	0.17	100.0	100.0	100.0	0.20	0.45	0.70	95.1	91.4	87.7
52	0.9928	0.9801	-7.3	119.6	0.20	100.0	100.0	100.0	0.53	1.44	0.58	91.4	86.4	76.5
50	0.9906	0.9756	10.7	132.8	0.22	100.0	100.0	100.0	0.42	1.15	0.95	91.4	85.2	76.5
48	0.9889	0.9730	19.9	137.2	0.23	100.0	100.0	100.0	0.56	0.53	0.81	91.4	85.2	75.3
46	0.9840	0.9634	45.4	158.0	0.27	100.0	98.8	97.5	0.77	0.27	0.95	92.6	84.0	67.9
44	0.9817	0.9604	52.3	160.1	0.28	100.0	98.8	96.3	0.85	0.22	1.04	92.6	82.7	66.7
42	0.9761	0.9511	69.7	172.7	0.32	98.8	97.5	93.8	1.68	5.14	1.98	93.8	85.2	70.4
40	0.9702	0.9419	83.7	181.8	0.34	97.5	95.1	93.8	1.78	3.26	2.42	91.4	82.7	71.6
38	0.9664	0.9375	89.4	182.8	0.36	96.3	93.8	93.8	1.42	1.87	1.64	91.4	81.5	65.4
36	0.9572	0.9238	105.2	193.7	0.39	96.3	93.8	91.4	2.68	4.59	1.89	91.4	80.2	65.4
34	0.9479	0.9113	117.1	200.9	0.42	97.5	95.1	92.6	2.41	3.44	1.72	90.1	79.0	58.0
32	0.9370	0.8971	128.4	207.4	0.46	97.5	93.8	86.4	2.94	2.96	1.93	88.9	76.5	54.3
30	0.9228	0.8789	140.9	215.1	0.50	95.1	90.1	80.2	2.91	4.15	2.68	84.0	75.3	58.0

Abbreviations: AIC, Akaike criteria; BIC, Bayesian criteria; CV(LOO), predicted root mean square error obtained with Leave-one-Out cross validation; Code, model name; CV(10-fold), predicted error obtained by k = 10; CV(1/5), predicted error using 20% of representative sample; FA3 and FA6, fitting accuracy using 3-level and 6-level toxicity classification; FA2, fitting accuracy obtained with 2-level classification; p, number of parameters; PA-3 and PA6, predicting accuracy using 3-level and 6-level potency categories; PA2, predicting accuracy obtained with 2-level classification; R², Pearson coefficient of determination; R²adj, adjusted coefficient of determination; σ_{fit} , mean squared model error.

In order to check whether further model simplification was still possible, we processed them with “least significant parameter removal” criterion.²² In Tables 5 and 6, we present this simplification process.

From Table 5, we can conclude that is not possible to simplify the case2 model “s26” (p = 71) without seriously compromising

its performance—the removal of one single parameter implies an irreversible loss in the predictive accuracy of 6-level response—from 4 fails to 7 fails out of 90. A smaller impact is observed in the 3-level case by increasing from 2 to 3 fails out of 90. However, in terms of 2-level response, the predictive ability is maintained with p = 68 parameters (1 fail in 90) and still

significant till $p = 48$ (4 fails in 90) where PA3 and PA6 are very low (11 and 22 fails in 90, respectively).

In case3, selected model “m20” ($p = 66$) may be refined to $p = 62$ parameters with an additional improvement in predictive ability—from 3 fails to 2 and from 2 to 1 fails in 81 in PA6 and PA3 and from 1 fail to no fails in PA2. This is a clear evidence of an “overfitting” error in multivariate models. Continuing model simplification, predictive ability decreases but it is still possible to accurately predict 2-level toxicity with $p = 56$ (1 fail in 81) while PA3 and PA6 remains above 90% (with 5 and 8 fails in 81, respectively). With $p = 48$ (72.7% of initial parameters), it is still possible to correctly describe the response (100% in FA6, FA3, and FA2), but predictive ability considerably decreases (7 fails in 81 for 2-level response). It is still possible to go further in the model simplification to $p = 42$ (63.6% of initial parameters) where fitting ability (93.8%, 97.5%, and 98.8% for FA6, FA3, and FA2) and predictive ability for 2-level response (93.8% in PA2) remain significant.

By comparing both cases, we can see that data from *in vitro* and *in chemico* assays seem to be relevant to describe and predict toxicity and the best results are obtained when combined with some *in silico* parameters.

According to Table 6, the initial model (“m20” with $p = 66$) preserves information related to all experimental descriptors (CysDep, LysDep, and EC30) and 63 (62 plus a constant) *in silico* predictors (electrophilicity index [EI], number of nitrogen atoms [nN], number of halogen atoms [nX], nR09, TI1, TI2, ww, Whete, Jhetp, MAXDP, PW4, BAC, Lop, D/Dr03, T(N.Cl), GGI1, GGI5, GGI6, JGI2, JGI3, JGI7, JGI9, J3D, ADDD, G1, PJI3, L/Bw, QYYp, G(N.S), G(O.S), nCp, nCs, the number of substituted benzenes sp2 [nC_b], nCconj, nR = Cs, the number of non-terminal carbons sp [nR#C-], number of aliphatic esters [nRCOOR], nRCHO, nRNH2, number of hydroxyl groups [nROH], nArOH, nOht, nHBonds, hydrophilic factor [Hy], AMR, MLOGP, SRW05, piPC05, piPC06, piPC08, piPC09, SIC0, CIC0, BIC1, C-001, C-005, C-008, C-024, H-051, H-052, O-056, S-107).

By simplification, with the removal of 4 variables (nR09, Lop, nRCHO, H-052), the new model gains maximal prediction accuracy (97.5%, 98.8%, and 100.0% accuracy for 6-level, 3-level, and 2-level responses, respectively).

By further simplification, with the removal of extra 6 *in silico* predictors (Whete, JGI2, J3D, nCconj, MLOGP, O-056), the model remains accurate, describing all responses (100.0% for FA6, FA3, and FA2) but losing some predictive ability—there is a generic increase in fails from 0 to 1, from 1 to 5, and from 2 to 8 in PA2, PA3, and PA6, respectively.

By removing 8 extra variables, we encounter a very interesting situation where all responses are very well described (100% for FA2, FA3, and FA6), and we still get a good prediction for 2-level response (91.4%), but results for 3-level and 6-level responses were not as good (only about 85.2% and 75.3%).

After stripping another extra 6 predictors, we still get a quite satisfactory model in describing all responses (98.8%, 97.5%, and 93.8% for FA2, FA3, and FA6) while higher performance

differences are evidenced in terms of prediction—a quite satisfactory prediction for 2-level response case (PA2 = 93.8), a reasonable prediction in 3-level response (PA3 = 85.2%) and a bad result for 6-level response (PA6 = 70.4%). This last relatively reliable model maintains dependencies with 41 predictors—2 experimental (CysDep and Ec30) and 39 *in silico* (EI, nN, TI1, ww, Jhetp, PW4, BAC, D/Dr03, T(N.Cl), GGI1, GGI5, GGI6, JGI3, JGI7, JGI9, G1, PJI3, QYYp, G(N.S), nCp, nCs, nCb-, nR = Cs, nRCOOR, nRNH2, nROH, nArOH, nOht, Hy, AMR, SRW05, piPC05, piPC06, piPC09, SIC0, CIC0, C-001, C-008, H-051).

Discussion

Results demonstrated the advantage of using PLS and MOLS combined with a first-degree polynomial approach in order to retrieve information for describing and predicting skin sensitization.

Recalling Table 2, PLS was able to select 22 predictors, where only one of the experimental descriptors was evidenced (CysDep). Comparing PLS and MOLS results on variable relevance, they only coincide in 3 predictors (CysDep, IE, nR = Cs). Previous works involving soft-modeling (PLS) and hard-modeling (MOLS) in response description and prediction^{19,20,22} also showed similar results, while PLS is only focused in relating predictor subspace information with response subspace information and, independently of variable correlation, MOLS, with all statistical support, is able to detect internal subspace correlations and thus allows to select most relevant and independent ones. However, PLS was important in our approach given it allowed the generation of initial seeding predictors for model generating and building processes.

By analyzing the molecular descriptors that demonstrated to be relevant for the description of the response, and by considering the key events beyond the AOP for skin sensitization, it is possible to find common points. These molecules are frequently strong electrophiles with polarized bonds and have functional groups like amides, aldehydes and ketones. Therefore, it is easy to understand the presence of the EI as a molecular descriptor related with the KE1 since it is a descriptor of reactivity that measures the global electrophilic nature of a molecule within a relative scale. The relevance of descriptors like the nROH or the number of aromatic hydroxyl groups is also clear. The substantially greater electronegativity of oxygen in comparison with carbon and hydrogen is responsible for the polarization of the covalent bonds in this functional group. In the same way, the nN is also related to KE1 because this atom is a part of the constitution of an amide, described as one of the reactive functional groups in skin sensitizers.

As a consequence of the skin sensitizers structural characteristics, they react mostly with cysteine and lysine residues via bimolecular and aromatic nucleophilic substitution (S_N2 and S_NAr), acylation reactions, Michael addition, and Schiff base formation, as shown in Figure 3.^{32,38,39}

Thus, descriptors like nRCOOR, nX, nCb-, nR = Cs, and nR#C-, are correlated with the reactivity between proteins and

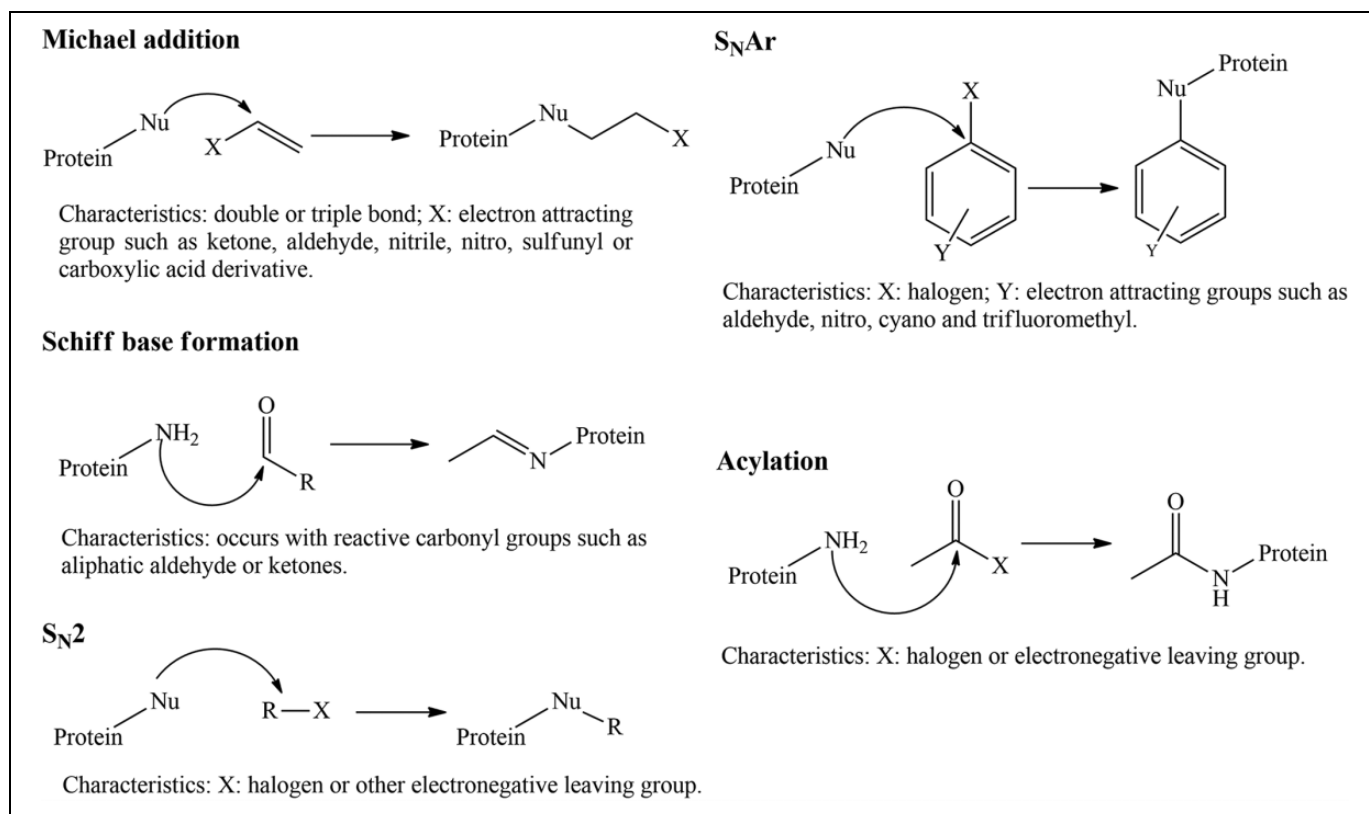


Figure 3. Mechanisms of reactions between proteins and allergens, promoting skin sensitization. Nu indicates nucleophile; S_N2, bimolecular nucleophilic substitution; S_NAr, aromatic nucleophilic substitution. Adapted from reference.³²

allergens. Aliphatic esters are carboxylic acid derivatives and for this reason can react with proteins via Michael addition. This reaction also involve double or triple bonds and consequently sp² and sp carbons (nR = Cs and nR#C-). For the occurrence of KE1 via S_N2, S_NAr, or acylation, the nX might be determinant because in general they are good leaving groups. In addition, the reactivity via S_NAr is also benefited by the presence of substituted benzenes (nC_b-).

Skin penetration properties have a relevant role to trigger the AOP for skin sensitization. Sensitizers are compounds capable of penetrating the skin, and once there, they react with skin proteins as described previously. In this regard, the inclusion of the octanol–water partition coefficient (LogP; descriptor MLOGP) and the Hy are also relevant because the penetration depends on the chemicals lipophilicity or hydrophobicity. It is also known that the surface charge of the molecules influences skin penetration and their reactivity, which was confirmed by the selection of 7 topological charge descriptors (GGI1, GGI5, GGI6, JGI2, JGI3, JGI7, and JGI9). On the other hand, the walk and path variables describe molecular complexity that also influences the above events. Eight of these descriptors turned out to be relevant (SRW05, piPC05, piPC06, piPC08, piPC09, SIC0, CIC0, and BIC1). It should also be noted that the geometry of the molecules influences reactivity, which supports the selection of 8 of these descriptors (J3D, ADDD, G1, PJI3, L/Bw, QYYp, G[N.S], and G[O.S]).

Supporting the confidence on the variables selected by our approach, previous studies revealed that descriptors correlated with the reactivity, the energy of highest occupied and lowest unoccupied molecular orbitals, and with the skin penetration are preferred in the description of the skin sensitization response.⁴⁰

In fact, there are several other methods described in the literature to evaluate risk and potency of skin sensitizers. However, very few use human information to train models, being hazard retrieved from animal methods such as the LLNA or GPMT, the most commonly used. This constitutes a serious limitation because the information obtained by these methods does not adequately correlate with human clinical responses. The potency categorization of chemicals is also predicted in some approaches but considering just 3 or 4 categories. In this study, 6-levels of classification have been considered allowing a more accurate classification of chemical potency.

For hazard identification (sensitizers vs non-sensitizers), our method revealed a better performance in relation to the methods already available (generally inferior to 90%). The model with the best performance until now was described by Strickland and colleagues.⁴¹ This defined approach includes information from 3 *in chemico* and *in vitro* assays (DPRA, h-CLAT, and KeratinoSens), 6 physicochemical properties, and 12 *in silico* descriptors from QSAR Toolbox version 3.2. This model showed 92% of accuracy and included the average of

cysteine and lysine depletion, information obtained with h-CLAT assay and with the Toolbox and LogP as predictors. The study also reveals the importance of cysteine depletion and LogP in the skin sensitization response, 2 descriptors also identified in our approach (CysDep and MLOGP). However, the aforementioned method has a performance equal to, if not worse than our models and the experimental assays considered are more complex, making the prediction less practical and more expensive.

Models to predict the potency of chemicals generally show lower accuracies than for hazard identification and our models are no exception. However, our accuracy reveals an advance relative to other methods described in the literature. Zang and colleagues described a model to predict skin sensitization potency considering the GHS classification for LLNA and humans, using 6 physicochemical properties and experimental data from DPRA, h-CLAT, and KeratinoSens assays.⁴² Authors described the best model as having an accuracy of 89% and 81% relatively to LLNA and human data, respectively. When compared to our approach, we achieved an accuracy for 6-levels potency in humans equal, if not better, considering or not experimental data. Once again, the experimental assays used by Zang et al adds limitations to the model, such as the costs and the time expended.

Conclusions

In this work, we evaluated the feasibility of experimental and computational descriptors to properly predict the skin sensitization potential of chemicals and their potency. To achieve this goal, we used a data set of known human skin allergens and non-allergens with different potencies to build classifiers for high-throughput virtual screening. Two skin sensitization responses were considered, 1 to categorize the potency of chemicals (according to a 6-level and 3-level scale) and the other for hazard classification, aiming to distinguish skin sensitizers from non-sensitizers (2-level scale).

In an initial modeling stage, PLS was used to extract the most relevant predictors from a huge pool of variables (over 380), using a direct and a weighted criteria. With this selected pool, several models were developed and refined with MOLS optimization based on first degree polynomial dependencies, according to the AIC and BIC criteria. Via hard-modeling, we were able to retrieve direct dependencies between predictors and responses, evidencing their relevance in describing and predicting new values. This modeling methodology is relatively simple to implement and present the advantages of evidencing direct relationship between predictors and the adverse outcome pathway for skin sensitization.

From our results, we conclude that by using exclusively *in silico* calculated descriptors, it is possible to avoid time-consuming experimental tests, but the models are less accurate than the ones additionally including *in vitro* and *in chemico* information. However, in practical terms, it can be advantageous to consider both, using in a first approach, the models based on *in silico* data and then the models based

on *in vitro*, *in chemico*, and *in silico* data to confirm the results. For instance, it is possible to estimate the potential of a chemical to induce skin sensitization, as well as its respective potency, in a fast and cheap way and, importantly, without the need of having the chemical synthesized or isolated. Using this strategy, it is possible for chemical, cosmetic, and pharmaceutical industries to make better and more informed decisions in initial chemical screening steps, before investing resources in experimental assays.

Author Contributions

Silva, F. contributed to conception and design, contributed to acquisition, analysis, and interpretation, drafted manuscript, and critically revised manuscript; Brites, G. contributed to conception, contributed to acquisition, and critically revised manuscript; Ferreira, I. contributed to conception, contributed to acquisition and critically revised manuscript; Silva, A. contributed to conception, contributed to acquisition, and critically revised manuscript; Neves, B. contributed to conception and design, contributed to analysis and interpretation, drafted manuscript, and critically revised manuscript; Pereira, J. contributed to conception and design, contributed to acquisition, analysis, and interpretation, drafted manuscript, and critically revised manuscript; Cruz, M. contributed to conception and design, contributed to interpretation, drafted manuscript, and critically revised manuscript. All authors gave final approval and agree to be accountable for all aspects of work ensuring integrity and accuracy.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The Coimbra Chemistry Centre (CQC) is supported by Fundação para a Ciência e a Tecnologia (FCT), through the project UI0313/QUI/2013, also co-funded by FEDER/COMPETE 2020-UE. The Center for Neuroscience and Cell Biology (CNC) is supported by European Regional Development Fund (ERDF) through the Competitiveness Factors Operational Program – COMPETE 2020 and by National Funds through the FCT within the scope of the Strategic Project with the reference: POCI-01-0145-FEDER-007440. This work was also financed by the ERDF, through the Centro 2020 Regional Operational Programme: project CENTRO-01-0145-FEDER-000012-HealthyAging2020, the COMPETE 2020 – Operational Programme for Competitiveness and Internationalisation, the Portuguese national funds via FCT, ref. CENTRO-01-0145-FEDER-029369. Isabel Ferreira and Gonçalo Brites are supported by the FCT through an individual PhD fellowship (SFRH/BD/110717/2015 and PD/BDE/142926/2018, respectively).

ORCID iD

Filipa A. L. S. Silva  <https://orcid.org/0000-0002-7400-267X>

Supplemental Material

Supplemental material for this article is available online.

References

1. Nguyen HL, Yiannias JA. Contact dermatitis to medications and skin products. *Clin Rev Allerg Immu.* 2019;56(1):41-59. doi:10.1007/s12016-018-8705-0
2. Luís A, Demétrio J, Silva A, Ferreira I, Cruz MT, Neves BM. Oxidative stress-dependent activation of the eIF2 α -ATFr unfolded protein response branch by skin sensitizer 1-fluoro-2, 4-dinitrobenzene modulates dendritic-like cell maturation and inflammatory status in a biphasic manner. *Free Radic Biol Med.* 2014;77:217-229. doi:10.1016/j.freeradbiomed.2014.09.008
3. Anderson SE, Siegel PD, Meade BJ. The LLNA: a brief review of recent advances and limitations. *J Allergy.* 2011;2011:424203. doi:10.1155/2011/424203
4. Johnston GA, Exton LS, Mustapa MFM, et al. British Association of Dermatologists' guidelines for the management of contact dermatitis 2017. *Br J Dermatol.* 2017;176(2):317-329. doi:10.1111/bjd.15239
5. 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. *Off J Eur Union.* 2003;66(L):26-35.
6. Organization for Economic Co-operation and Development. *The Adverse Outcome Pathway for Skin Sensitization Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence; Part 2: Use of the AOP to Develop Chemical Categories and Integrated Assessment and Testing Approaches.* OECD Publishing; 2012.
7. Organization for Economic Co-operation and Development. *Test No. 442C: In: Chemico Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA).* OECD Publishing; 2015.
8. Organization for Economic Co-operation and Development. *Test No. 442D. In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method. In: OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects.* OECD Publishing; 2015.
9. Organization for Economic Co-operation and Development. *Test No. 442E: In Vitro Skin Sensitisation Assays Addressing the Key Event on Activation of Dendritic Cells on the Adverse Outcome Pathway for Skin Sensitisation.* OECD Publishing; 2017.
10. Hoffmann S, Kleinstreuer N, Alépée N, et al. Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database. *Crit Rev Toxicol.* 2018;48(5):344-358. doi:10408444.2018.1429385
11. Kleinstreuer N, Hoffmann S, Alépée N, et al. Non-animal methods to predict skin sensitization (II): an assessment of defined approaches. *Crit Rev Toxicol.* 2018;48(5):359-374. doi:10.1080/10408444.2018.1429386
12. Kirwan GM, Johansson E, Kleemann R, et al. Building multivariate systems biology models. *Anal Sci.* 2012;84(16):7064-7071. doi:10.1021/ac301269r
13. Cramer RD. Partial least squares (PLS): its strengths and limitations. *Perspect Drug Dis Des.* 1993;1(2):269-278. doi:10.1007/BF02174528
14. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst.* 2001;58(2):109-130. doi:10.1016/S0169-7439(01)00155-1
15. Barroso M, Pereira JC, Pais AA, Arnaut LG, Formosinho SJ. Molecular factor analysis in atom-transfer reactions. *Mol Phys.* 2006;104(5-7):731-743. doi:10.1080/00268970500417085
16. Pereira JC, Serpa C, Arnaut LG, Formosinho SJ. Molecular factor analysis in self-exchange electron transfer reactions in solution. *J Mol Liq.* 2010;156(1):3-9. doi:10.1016/j.molliq.2010.07.007
17. Cova TF, Pereira JL, Pais AA. Is standard multivariate analysis sufficient in clinical and epidemiological studies? *J Biomed Inform.* 2013;46(1):75-86. doi:10.1016/j.jbi.2012.09.005
18. Lopes AC, Cova TF, Pais AA, Pereira JL, Colaco B, Cabrita AM. Improving discrimination in the grading of rat mammary tumors using two-dimensional mapping of histopathological observations. *Exp Toxicol Pathol.* 2014;66(1):73-80. doi:10.1016/j.etp.2013.09.001
19. Amaral R, Pereira JC, Pais AA, Santos LM. Is axenicity crucial to cryopreserve microalgae? *Cryobiology.* 2013;67(3):312-320. doi:10.1016/j.cryobiol.2013.09.006
20. Gabriel D, Matias T, Pereira JC, Araújo R. (2013). Predicting gas emissions in a cement kiln plant using hard and soft modeling strategies. In: *2013 IEEE 18th Conference on Emerging Technologies & Factory Automation (ETFA).* IEEE;1-8. doi:10.1109/ETFA.2013.6648036
21. De Almeida Brehm F, De Azevedo J C R, Da Costa Pereira J, Burrows HD. Direct estimation of dissolved organic carbon using synchronous fluorescence and independent component analysis (ICA): advantages of a multivariate calibration. *Environ. Monit Assess.* 2015;187(11):703. doi:10.1007/s10661-015-4857-z
22. Pereira JC, Marques JMC, Włodarczyk E, Fenert B, Zarzycki PK. Toward the understanding of micro-TLC behavior of various dyes on silica and cellulose stationary phases using a data mining approach. *J AOAC Int.* 2018;101(5):1437-1447. doi:10.5740/jaoacint.18-0061
23. Basketter DA, Alépée N, Ashikaga T, et al. Categorization of chemicals according to their relative human skin sensitizing potency. *Dermatitis.* 2014;25(1):11-21. doi:10.1097/DER.0000000000000003
24. United Nations. *Globally Harmonized System of Classification and Labelling of Chemicals.* United Nations, 2011.
25. Basketter D. Skin sensitisation, adverse outcome pathways and alternatives. *Altern Lab Anim.* 2016;44(5):431-436. doi:10.1177/026119291604400501
26. Bauch C, Kolle S, Ramirez T, et al. Putting the parts together: combining in vitro methods to test for skin sensitizing potentials. *Regul Toxicol Pharmacol.* 2012;63(3):489-504. doi:10.1016/j.yrtph.2012.05.013
27. Urbisch D, Mehling A, Guth K, et al. Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul Toxicol Pharmacol.* 2015;71(2):337-351. doi:10.1016/j.yrtph.2014.12.008
28. Asturiol D, Casati S, Worth A. Consensus of classification trees for skin sensitisation hazard prediction. *Toxicol In Vitro.* 2016;36:197-209. doi:10.1016/j.tiv.2016.07.014
29. O'Brien J, Wilson I, Orton T, Pognan F. Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of

- mammalian cell cytotoxicity. *Eur J Biochem.* 2000;267(17): 5421-5426. doi:10.1046/j.1432-1327.2000.01606.x
30. Tetko IV, Gasteiger J, Todeschini R, et al. Virtual computational chemistry laboratory – design and description. *J Comput Aided Mol Des.* 2005;19(6):453-463. doi:10.1007/s10822-005-8694-y
31. VCCLAB, Virtual Computational Chemistry Laboratory. 2005. Accessed March 23, 2019. <http://www.vcclab.org>
32. Enoch SJ, Cronin MTD, Schultz TW, Madden JC. Quantitative and mechanistic read across for predicting the skin sensitization potential of alkenes acting via Michael addition. *Chem Res Toxicol.* 2008;21(2):513-520. doi:10.1021/tx700322 g
33. Parr RG, Szentpaly LV, Liu S. Electrophilicity index. *J Am Chem Soc.* 1999;121(9):1922-1924. doi:10.1021/ja983494x
34. Eaton JW, Bateman D, Hauberg S, et al. *GNU Octave Version 3.8.1 Manual: a High-Level Interactive Language for Numerical Computations, CreateSpace Independent Publishing Platform.* 2014; ISBN 1441413006. <http://www.gnu.org/software/octave/doc/interpreter/>
35. *R Core Team and contributors worldwide, The R Stats Package, Ver.: 3.7.0.* Accessed March 23, 2019. <https://www.rdocumentation.org/packages/stats>
36. *RCore Team and contributors worldwide, Choose a model by AIC in a Stepwise Algorithm.* Accessed March 23, 2019. <https://www.rdocumentation.org/packages/stats/versions/3.5.3/topics/step>
37. *R Core Team and contributors worldwide, Akaike's An Information Criterion.* Accessed March 23, 2019. <https://www.rdocumentation.org/packages/stats/versions/3.5.3/topics/AIC>
38. Aptula AO, Patlewicz G, Roberts DW. Skin sensitization: reaction mechanistic applicability domains for structure–activity relationships. *Chem Res Toxicol.* 2005;18(9):1420-1426. doi:10.1021/tx050075 m
39. Rustemeyer T, Van Hoogstraten IMW, Von Blomberg BME. Mechanisms of allergic contact dermatitis. In: John SM, Johansen JD, Rustemeyer T, Elsner P, Maibach HI, eds. *Kanerva's Occupational Dermatology.* Cham: Springer. 2019;1-41. doi:10.1007/978-3-642-02035-3_14
40. Wilm A, Jochen K, Kirchmair J. Computational approaches for skin sensitization prediction. *Crit Rev Toxicol.* 2018;48(9): 738-760. doi:10.1080/10408444.2018.1528207
41. Strickland J, Zang Q, Paris M, et al. Multivariate models for prediction of human skin sensitization hazard. *J Appl Toxicol.* 2017;37(3):347-360. doi:10.1002/jat.3366
42. Zang Q, Paris M, Lehmann DM, et al. Prediction of skin sensitization potency using machine learning approaches. *J Appl Toxicol.* 2017;37(7):792-805. doi:10.1002/jat.3424