



Semi-correlations as a tool to model for skin sensitization

Alla P. Toropova^{*}, Andrey A. Toropov, Emilio Benfenati

Department of Environmental Health Science, Laboratory of Environmental Chemistry and Toxicology, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156, Milano, Italy

ARTICLE INFO

Handling Editor: Dr. Jose Luis Domingo

Keywords:

Local lymph node assay (LLNA)
Direct peptide reactivity assay (DPRA)
Human cell line activation test (h-CLAT)
Skin sensitization
QSAR
CORAL software

ABSTRACT

Semi-correlation specifically assesses the correlation between a binary variable and a continuous variable. Semi-correlations were applied to develop binary models for various endpoints. We applied the semi-correlation to develop models of two kinds of skin sensitization one related to animals (local lymph node assay LLNA) and one to human beings (direct peptide reactivity assay DPRA and/or human cell line activation test h-CLAT). The models refer to binary classification for a two-level strategy: the first level (analysis of all compounds) is used in the format “sensitizer or non-sensitizer”, and the second level (only sensitizers) is a further classification in the format “strong or weak sensitizer”. The ranges of statistical characteristics of the models depend on the endpoint, LLNA or DPRA/h-CLAT: for the first level, sensitivity: 0.69–0.88, specificity: 0.75–0.89, accuracy: 0.77–0.87, Matthew’s correlation coefficient (MCC): 0.54–0.57 and for the second level, sensitivity: 0.70–1.0, specificity: 0.78–0.83, accuracy: 0.77–0.87, MCC: 0.54–0.76. Thus, the described approach can be applied to building up models of the skin sensitization potency.

1. Introduction

Allergic contact dermatitis is a problem-affecting people, exposed to chemicals for occupational reasons or in general as consumers (Owen et al., 2018). The frequency of this skin pathology in the population justifies the need to characterize the skin sensitization potential of chemical substances in order to define ways to protect workers and consumers. This is a non-trivial task since the skin sensitization is caused by a sequence of molecular and cellular events, involving several steps (Roberts and Aptula, 2008). Earlier, the potential for substances to cause skin sensitization was assessed in animal assays, including the guinea pig maximization test and the mouse local lymph node assay (LLNA) (OECD, 2009). However, animal testing is increasingly being limited (REACH, 2016), particularly for certain products such as cosmetics. Indeed, in Europe in vivo tests to obtain data regarding cosmetics are not allowed (EC, 2009).

Quantitative structure-activity relationships (QSARs) are an attractive way to estimate the skin sensitization potential of various compounds. The literature reports several QSAR models for skin sensitization. They vary depending on the endpoint to be modeled, on the algorithm, and the fact that some models are binary classifiers, or attempt to predict the severity of the outcome. In the past, we built up some models (Toropov et al., 2019; Toropova and Toropov, 2017a;

Toropov and Toropova, 2021) with CORAL software (<http://www.insilico.eu/coral>). Using a series of molecular descriptors and adaptive fuzzy partition, Chaudhry et al. developed a model (Chaudhry et al., 2010) that is freely available on the VEGA website (www.vegahub.eu). Toxtree provides structural alerts related to various substances (<https://ec.europa.eu/jrc/en/scientific-tool/toxtree-tool>).

A typical approach for modeling skin sensitization is based on mechanistic modeling (Roberts et al., 2011). The integration of several models has been tried, and in some cases, experimental data have been combined with predicted values (van der Veen et al., 2014; Asturiol et al., 2016; Kumar et al., 2016). More recently, studies, which addressed endpoints able to correlate with skin sensitization (Borba et al., 2021; Zang et al., 2017; Tung et al., 2018, 2019). Semi-correlations are one way to build up a model for skin sensitization, as done for other endpoints (Toropova et al., 2019a,b; Toropova and Toropov, 2017b).

Semi-correlation is a dichotomy that arises from regression and distributes compounds (their descriptors) into classes “active” and “inactive” (Fig. 1). The 2D descriptors for the model are calculated from simplified molecular input line entry systems (SMILES) (Weininger, 1988).

With this approach, we developed a battery of models that can be evaluated together, some related to LLNA, and others to human data.

^{*} Corresponding author.

E-mail address: alla.toropova@marionegri.it (A.P. Toropova).

<https://doi.org/10.1016/j.fct.2021.112580>

Received 23 April 2021; Accepted 20 September 2021

Available online 22 September 2021

0278-6915/© 2021 Elsevier Ltd. All rights reserved.

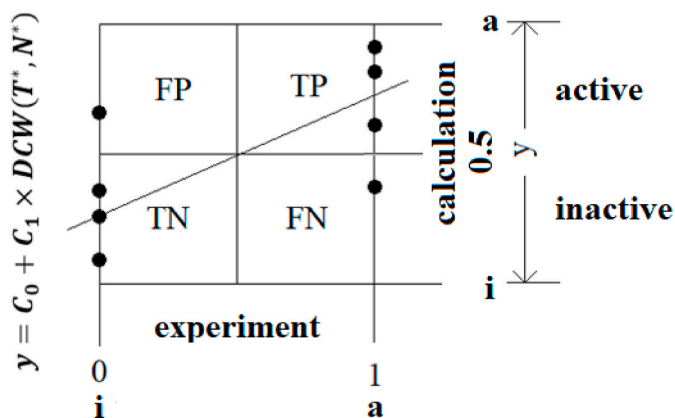


Fig. 1. The general scheme of a model based on semi-correlation. The TN, TP, FN, and FP are true negative, true positive, false negative and false positive results, respectively.

Since it may be important to assess the severity of the effect, the models can identify strong and weak sensitizers.

The present study aims to estimate the statistical quality of the models for skin sensitization applying the semi-correlation approach, generating a sequence of models to assess the severity of the effect.

2. Materials and method

The numerical data on skin sensitization were from the literature (Zang et al., 2017). Two sets are analyzed in (Zang et al., 2017): (i) Dataset I contains LLNA data for 120 substances; and (ii) Dataset II contains human test outcomes (87 substances). The data relative to human skin sensitization obtained to three methods: direct peptide reactivity assay (DPRA), KeratinoSens™, and human cell line activation test (h-CLAT). Thus, this kind of data follows the recent approach to applying a battery of methods, as we discussed in the Introduction. We modeled both endpoints, LLNA, and the other using alternative assays related to the toxicological pathway in humans. Sources of data used in the study are listed in *Supporting information* as Table S5 and Table S6. These are two data sets, one containing LLNA data on 120 substances (Table S5) and the other containing human skin sensitization data on 87 substances (Table S6). All 87 substances in the human sensitization data set are represented in the LLNA data set. The experimental data that applied here, are obtained in accordance with OECD directives (OECD, 2018) as well as obey to recommendations of the Institute for In Vitro Sciences (IIVS, 2021).

We analyzed the corresponding endpoints (from Dataset I and Dataset II) on two levels for each endpoint, resulting in four models. The first level is the detection of sensitizers in a group of substances that contains both sensitizers and non-sensitizers. The second level focuses on the sensitizers only, divided as strong and weak. The definition of the potency follows that in (Zang et al., 2017), which relates to what adopted within the Globally Harmonized System (GHS) of Classification and Labelling of Chemicals: for LLNA, in case of the effect a substance is a strong sensitizer if the estimated test substance concentration that produces a stimulation index of 3 (EC3) is less than (or equal to) 2%, and it is weak if the concentration is >2%; for the human potency, the threshold is at 500 µg/cm² of skin area (Zang et al., 2017).

The corresponding models are calculated according to the scheme represented by Fig. 1 as the following

$$\text{Category1}(\text{SMILES}) = \begin{cases} \text{sensitizer} & \text{if, } y \geq 0.5 \\ \text{non_sensitizer} & \text{if, } y < 0.5 \end{cases} \quad (1)$$

$$\text{Category2}(\text{SMILES}) = \begin{cases} \text{strong sensitizer} & \text{if, } y \geq 0.5 \\ \text{weak sensitizer} & \text{if, } y < 0.5 \end{cases} \quad (2)$$

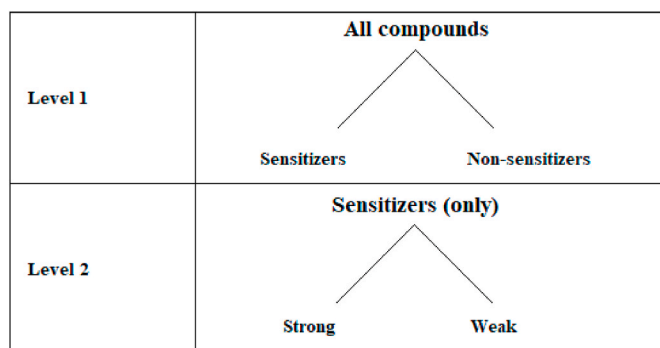


Fig. 2. The scheme of building up binary models on level 1 (Eq. (1)) and level 2 (Eq. (2)).

Table 1

An example of building a self-organized vector of atom-pairs proportions (APP_j) for the following SMILES: CCC(Br) (CC)C(=O)NCC(N)=O.

Atom1	Atom2	n1	n2	Code applied to build up the model
Br	N	1	2	(Br.N)..1.2.
Br	O	1	2	(Br.O)..1.2.
Br	Double bond	1	2	(Br. =).1.2.
N	O	2	2	(N..O)..2.2.
N	Double bond	2	2	(N.. =).2.2.
O	Double bond	2	2	(O.. =).2.2.

Table 2

Examples of the S_k , SS_k and SSS_k (for SMILES = Oc1ccc(O)cc1).

S_k	SS_k	SSS_k
O		
c	c ... O	
1	c ... 1	O ... c ... 1 ...
c	c ... 1	c ... 1 ... c ...
c	c ... c	c ... c ... 1 ...
c	c ... c	c ... c ... c ...
(....	c ... (....	c ... c ... (...
O	O ... (....	c ... (... O ...
(....	O ... (....	(... O ... (...
c	c ... (....	c ... (... O ...
c	c ... c	c ... c ... (...
1	c ... 1	c ... c ... 1 ...

The translations of category 1 to category 2 are the following discriminations: Level 1: 1A = '1'; 1B = '1'; NEG = '0'. Hence sensitizers marked by '1', non-sensitizers marked by '0'. Level 2: 1A = '1'; 1B = '0' (REG removed). Hence strong marked by '1', weak marked by '0'. The scheme of building up and applying semi-correlations is described in the literature (Toropova et al., 2019a,b). Fig. 1 shows the essence of the model based on semi-correlation. Fig. 2 demonstrates the essence of models calculated with Eq. (1) and Eq. (2).

The y is calculated as the following

$$y = C_0 + C_1 \times DCW(T^*, N^*) \quad (3)$$

In other terms, the model is a simple equation, which combines a number of other parameters, as below. The descriptor of the correlation weights (DCW) is calculated as

$$DCW(T, N) = CW(APP) + \sum_{k=1}^{NA} CW(S_k) + \sum_{k=1}^{NA-1} CW(SS_k) + \sum_{k=1}^{NA-2} CW(SSS_k) \quad (4)$$

The APP is the atom pairs' proportions self-organized vector. Table 1 gives examples of the APP . The S_k , SS_k , and SSS_k are the simplified molecular input line entry system (SMILES) fragments (Weininger,

Table 3

The statistical quality of models for Dataset I (Level 1): LLNA data modelling sensitizers versus non-sensitizers.

IIC	APP	Set	TP	TN	FP	FN	Sen*	Spe	Acc	MCC
unused	unused	Active training	12	10	0	0	1	1	1	1
		Passive training	16	4	0	1	0.94	1	0.95	0.87
		Calibration	9	4	3	6	0.60	0.57	0.59	0.16
		Validation	7	6	3	6	0.54	0.67	0.59	0.20
Used	unused	Active training	9	7	3	3	0.75	0.70	0.73	0.45
		Passive training	14	3	1	3	0.82	0.75	0.81	0.49
		Calibration	14	6	1	1	0.93	0.86	0.91	0.79
		Validation	11	5	4	2	0.84	0.56	0.73	0.42
unused	used	Active training	12	10	0	0	1	1	1	1
		Passive training	1	4	0	16	0.06	1	0.24	0.11
		Calibration	10	3	4	5	0.67	0.43	0.59	0.09
		Validation	7	7	2	6	0.54	0.78	0.64	0.32
Used	used	Active training	10	9	1	2	0.83	0.90	0.86	0.73
		Passive training	10	4	0	7	0.59	1	0.67	0.46
		Calibration	14	7	0	1	0.93	1	0.95	0.90
		Validation	9	8	1	4	0.69	0.89	0.77	0.57

*) Sen-sensitivity; Spe-specificity; Acc-accuracy; MCC- Matthew's correlation coefficient; the "unused" means that corresponding element of building up a model (i.e. IIC or APP) is not used; the "used" means that corresponding element of building up a model (i.e. IIC or APP) is used.

1988) that represent the corresponding compound. Thus, all these variables are simple chemical data easily extracted by our software from the chemical structure represented through the SMILES format. The $CW(S_k)$, $CW(SS_k)$, $CW(SSS_k)$, $CW(APP)$ are correlation weights for the corresponding SMILES attributes.

Table 2 provides an example of these fragments for a chemical. The $CW(S_k)$, $CW(SS_k)$, and $CW(SSS_k)$ are correlation weights of these SMILES fragments calculated with the Monte Carlo technique (Kumar and Kumar, 2020a,b; Kumar et al., 2019; Toropov and Toropova, 2019; Toropov et al., 2020; Toropova et al., 2019a,b, 2021; Toropova and Toropov, 2017c).

The T is the threshold in order to define rare and not rare SMILES fragments: if the frequency of a SMILES fragment in the active training set is less than T , its correlation weight is zero (i.e. the fragment is ignored as uninformative).

The N is the number of epochs of the Monte Carlo optimization (one epoch is the sequence of modifications of all correlation weights for fragments which are not rare ones according to the threshold T).

The original set of substances has been randomly split into four subsets: active training, passive training, calibration, and validation. Each of the above sets has a specific task. The active training set is a builder of the model, i.e. compounds of this set are used for building up the model. The passive training set is an inspector of the model, i.e. compounds of this set are used to check up whether the model is satisfactory for substances which are absent in the active training set. The task of the calibration set is to detect the start of the overtraining. Compounds of the validation set are used for the final validation of the predictive potential of the model.

In other words, the model initially is developed with the active training set, the passive training and the calibration sets are used to select the best parameters, and the validation set is used only when the model is constructed in order to evaluate the predictions for substances, which are invisible during developing of the model.

The distribution of available data into the above sets was random, but equivalent. Each of the above set contains about 25% of compounds from Dataset I or Dataset II.

Models obtained using the Monte Carlo optimization are based on two target functions TF_1 and TF_2 :

$$TF_1 = r_{aTRN} + r_{pTRN} - |r_{aTRN} - r_{pTRN}| * 0.1 \quad (5)$$

$$TF_2 = TF_1 + IIC_{CLB} * 0.5 \quad (6)$$

The r_{aTRN} and r_{pTRN} are the correlation coefficients between the observed and predicted endpoint values for the active training and passive training sets, respectively.

The index of ideality of correlation IIC_{CLB} is calculated with data on the calibration set as follows:

$$IIC_C = r_C \frac{\min(-MAE_C, +MAE_C)}{\max(-MAE_C, +MAE_C)} \quad (7)$$

$$\min(x, y) = \begin{cases} x, & \text{if } x < y \\ y, & \text{otherwise} \end{cases} \quad (8)$$

$$\max(x, y) = \begin{cases} x, & \text{if } x > y \\ y, & \text{otherwise} \end{cases} \quad (9)$$

$$-MAE_C = \frac{1}{N} \sum |\Delta_k|, \quad -N \text{ is the number of } \Delta_k < 0 \quad (10)$$

$$+MAE_C = \frac{1}{N} \sum |\Delta_k|, \quad +N \text{ is the number of } \Delta_k \geq 0 \quad (11)$$

$$\Delta_k = \text{observed}_k - \text{calculated}_k \quad (12)$$

The observed_k and calculated_k are corresponding values of the endpoint.

The statistical quality of models estimated with the following criteria:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (13)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (14)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (15)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (16)$$

The TP, TN, FP, and FN are the numbers of true positive, true negative, false positive, and false negative predictions, respectively.

3. Results

3.1. Level 1 for Dataset I

Table 3 sets out the statistical characteristics of the model for Dataset I, LLNA, on level 1, sensitizer versus non-sensitizer substances. The results are those obtained on the different subsets, but the most important values are those related to the validation set, which contains substances

Table 4

The statistical quality of models for Dataset I (Level 2): LLNA data modelling strong versus weak sensitizers.

<i>IIC</i>	<i>APP</i>	Set	TP	TN	FP	FN	Sen*	Spe	Acc	MCC
unused	unused	Active training	7	8	0	0	1	1	1	1
		Passive training	6	6	1	0	1	0.86	0.92	0.86
		Calibration	2	4	3	5	0.29	0.57	0.43	−0.15
		Validation	5	7	2	1	0.83	0.78	0.80	0.60
used	unused	Active training	6	7	1	1	0.85	0.88	0.87	0.73
		Passive training	4	3	4	2	0.67	0.43	0.54	0.10
		Calibration	7	6	1	0	1	0.86	0.93	0.87
		Validation	3	4	5	3	0.50	0.44	0.47	−0.05
unused	used	Active training	6	8	0	1	0.86	1	0.93	0.87
		Passive training	6	4	3	0	1	0.57	0.77	0.62
		Calibration	1	6	1	6	0.14	0.86	0.50	0
		Validation	3	5	4	3	0.50	0.56	0.53	0.05
used	used	Active training	6	8	0	1	0.86	1	0.93	0.87
		Passive training	6	5	2	0	1	0.71	0.85	0.73
		Calibration	4	6	1	3	0.57	0.86	0.71	0.45
		Validation	6	7	2	0	1	0.78	0.87	0.76

*) Sen-sensitivity; Spe-specificity; Acc-accuracy; MCC- Matthew's correlation coefficient; the "unused" means that corresponding element of building up a model (i.e. *IIC* or *APP*) is not used; the "used" means that corresponding element of building up a model (i.e. *IIC* or *APP*) is used.

Table 5

The statistical quality of models for Dataset II (Level 1): human data modelling sensitizers versus non-sensitizers.

<i>IIC</i>	<i>APP</i>	Set	TP	TN	FP	FN	Sen*	Spe	Acc	MCC
unused	unused	Active training	13	9	0	0	1	1	1	1
		Passive training	12	4	0	6	0.67	1	0.73	0.52
		Calibration	7	8	0	7	0.50	1	0.68	0.52
		Validation	8	8	1	4	0.67	0.89	0.76	0.55
used	unused	Active training	9	7	2	4	0.69	0.78	0.73	0.46
		Passive training	12	4	0	6	0.67	1	0.73	0.52
		Calibration	14	7	1	0	1	0.87	0.95	0.90
		Validation	8	7	2	4	0.67	0.78	0.71	0.44
unused	used	Active training	13	9	0	0	1	1	1	1
		Passive training	2	4	0	16	0.11	1	0.27	0.15
		Calibration	3	7	1	11	0.21	0.87	0.45	0.11
		Validation	8	7	2	4	0.67	0.78	0.71	0.44
used	used	Active training	11	6	3	2	0.85	0.67	0.77	0.52
		Passive training	14	4	0	4	0.78	1	0.82	0.62
		Calibration	12	8	0	2	0.86	1	0.91	0.83
		Validation	11	9	0	1	0.92	1	0.95	0.91

*) Sen-sensitivity; Spe-specificity; Acc-accuracy; MCC- Matthew's correlation coefficient; the "unused" means that corresponding element of building up a model (i.e. *IIC* or *APP*) is not used; the "used" means that corresponding element of building up a model (i.e. *IIC* or *APP*) is used.

not used to build up the model. The results indicate the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), which are used to calculate the sensitivity, specificity, accuracy, and Matthews correlation coefficient (MCC). The most interesting parameter

is probably the MCC because it takes into account the fact that the positive and negative substances are not balanced. MCC has values ranging from −1 to 1, and values above 0.5 indicate a good model. On level 1, the model with the best predictive potential is obtained by

Table 6

The statistical quality of models for Dataset II (Level 2): human data modelling strong versus weak sensitizers.

<i>IIC</i>	<i>APP</i>	Set	TP	TN	FP	FN	Sen*	Spe	Acc	MCC
unused	unused	Active training	9	6	0	0	1	1	1	1
		Passive training	8	6	0	0	1	1	1	1
		Calibration	3	6	2	3	0.50	0.75	0.64	0.26
		Validation	2	5	6	1	0.67	0.45	0.50	0.10
used	unused	Active training	9	6	0	0	1	1	1	1
		Passive training	8	6	0	0	1	1	1	1
		Calibration	2	5	3	4	0.33	0.62	0.50	−0.04
		Validation	2	4	7	1	0.67	0.36	0.43	0.03
unused	used	Active training	9	6	0	0	1	1	1	1
		Passive training	8	6	0	0	1	1	1	1
		Calibration	2	6	2	4	0.33	0.75	0.57	0.09
		Validation	1	8	3	2	0.33	0.73	0.64	0.05
used	used	Active training	9	6	0	0	1	1	1	1
		Passive training	8	6	0	0	1	1	1	1
		Calibration	5	7	1	1	0.83	0.87	0.86	0.71
		Validation	3	9	2	0	1	0.82	0.86	0.70

*) Sen-sensitivity; Spe-specificity; Acc-accuracy; MCC- Matthew's correlation coefficient; the "unused" means that corresponding element of building up a model (i.e. *IIC* or *APP*) is not used; the "used" means that corresponding element of building up a model (i.e. *IIC* or *APP*) is used.

applying *IIC* and *APP*. The model is the following:

$$y = 0.2272 (\pm 0.0183) + 0.0613 (\pm 0.0020) \times \text{DCW}(1,10) \quad (17)$$

The accuracy of this model is 0.77; MCC is 0.57.

3.2. Level 2 for Dataset I

Table 4 sets out the statistical characteristics of the model for Dataset I on level 2, which refers to the different potency of sensitizer substances: weak or strong for substances assessed using the LLNA. On level 2 the model with the best predictive potential is based on applying *IIC* and *APP*. The model is the following:

$$y = 0.3963 (\pm 0.0158) + 0.0652 (\pm 0.0015) \times \text{DCW}(1,15) \quad (18)$$

The accuracy of this model is 0.87; MCC is 0.76.

The accuracy for this model is 0.81.

3.3. Level 1 for Dataset II

Table 5 sets out the statistical characteristics of the model for Dataset II, a human test on level 1, to distinguish sensitizers from non-sensitizers. On level 1 the model with the best predictive potential is obtained applying *IIC* and *APP*. The model is the following:

$$y = 0.4265 (\pm 0.0204) + 0.07937 (\pm 0.00259) * \text{DCW}(1,10) \quad (19)$$

The accuracy of this model is 0.95; MCC is 0.91.

The accuracy of this model is 0.87.

3.4. Level 2 for Dataset II

Table 6 sets out the statistical characteristics of the model for Dataset II on level 2, which refers to human data assessed in their potency. On level 2 the model with the best predictive potential is obtained applying *IIC* and *APP*. The model is the following:

$$y = 0.5219 (\pm 0.021) + 0.09457 (\pm 0.00365) * \text{DCW}(1,10) \quad (20)$$

The accuracy of this model is 0.86; MCC is 0.70.

Supplementary materials section contains data on Dataset I and Dataset II for level 1 and level 2, i.e. the distribution into the active training (+), passive training (−), calibration (#), and validation (*) sets (Tables S1, S2, S3, and S4) respectively. Table S5 contains chemical names for the Dataset I; Table S6 contains chemical names for the Dataset II.

4. Discussion

The development of computational approaches to study skin sensitization is one of the fundamental directions of toxicology (Jaworska et al., 2015; Kleinstreuer et al., 2018; Hoffmann et al., 2018; Wilm et al., 2018, 2019). In order to build up reliable models for skin sensitization new assays, new algorithms, new theoretical knowledge, and integration of these heterogeneous components are necessary. New assays generate new data, which form the basis of the models. While initially the data mainly related to LLNA, now there are data from *in vitro* and *in chemico* studies, which can improve the overall target to assess the effects on human beings. Furthermore, the tendency to move away from *in vivo* data, such as LLNA, is related to the fact that in Europe *in vivo* studies are banned for cosmetics, and in other countries and for other industrial sectors *in vivo* experiments are often considered as a last resource.

The development of conceptually new models needs improvement of the technique of establishing QSAR models. One of the key way to the improvement for QSAR is the development of new statistical criteria for the estimation of the QSAR quality, especially the predictive potential for QSAR. The conceptual novelty of the index of ideality of correlation (Toropov et al., 2019; Kumar and Kumar, 2020a,b; Toropov and

Toropova, 2019; Toropova et al., 2019b) is the ability to take into account the determination coefficient together with the dispersion of predictions for an endpoint expressed in terms of mean absolute error. The semi-correlation (Toropova et al., 2019a,b) examined here is a special new tool of the structure-activity analysis. The index of ideality of correlation has been designed as a tool to improve the predictive potential for QSAR i.e. for regression models with a continuous range of endpoint values from some minimum to maximum. This index has demonstrated the attractive potential to improve the predictive potential of the semi-correlations.

Atom pair's proportions is another conceptually new way to improve the predictive potential of models. Factually, molecular structure can be interpreted as a random event. The definition of the event can be a list of atoms of a molecule. More specific can be the answer to the question "whether the molecule contains atoms A1 and A2"? A further level of the individuality of a molecule as a random event can be the answer to a question about the proportion of A1 and A2 in the molecule, e.g. 1:1, 1:2, 2:1, 1:3, 3:1, 2:3, 3:2, and so on. Using the Monte Carlo optimization, correlation weights for these events (situations in molecules) can be defined. The data can improve the predictive potential of a model or conversely, the data can decrease the predictive potential of a model. The estimation of the data can be obtained after carrying out the corresponding computational experiments.

Multiple models also require integrating the results from the different models, and for this reason, there are more and more studies proposing not one single model, but several ones, to be combined. For instance, the CAESAR model for skin sensitization, freely available on the VEGA website, is one of the first examples of hybrid models, combining the results of several models (Chaudhry et al., 2010). Bayesian methods have been applied to integrate heterogeneous data, some from experiments (van der Veen et al., 2014). Based on Bayesian networks, *in silico* models have been implemented and are ready to use: one is SaferSkin. Machine learning has also been employed to integrate data from different models, as in the case of CAESAR, and then also combining data from types of experiments. This heterogeneous data, relating not only to LLNA experiments, prompted the use of different algorithms, also offering freely available software programs (Borba et al., 2021; Zang et al., 2017; Tung et al., 2018, 2019). This shows in practice what we mentioned about the evolution of *in silico* models for skin sensitization, generating novel algorithms, novel data, and integration of heterogeneous data.

Anyhow, the use of new models still requires careful evaluation. The use of algorithms facilitates the integration of the data and covers data gaps, but the conceptual scheme for integration of the results, to address the effects on human beings, calls for further study and more data, to verify the relative contributions of heterogeneous data to model the human effect. We, therefore, preferred to leave the results from the LLNA and the human studies separate; an expert may evaluate them separately and integrate them by case.

We developed a number of models, applying different strategies. The best results (for instance considering the accuracy and the MCC values on the external validation set) were obtained when we used the *IIC* and *APP* conditions (Tables 3–6). Another indication of interest is the fact that the sensitivity and specificity values are closer when both these conditions are applied. Indeed, if the model is skewed towards a high sensitivity and low specificity, or vice versa, this is a weakness (which is anyhow captured by the MCC value). The role of *IIC* and *APP* is independent and both are quite important. Indeed, depending on the cases, models with only *IIC* are better than those with the only *APP*, or vice versa, but anyhow when both are used this is preferable. *APP* relates to the use of larger chemical information, related to the co-presence of certain atoms. Instead, *IIC* is related to the algorithm and the improvement is related to the reduced errors in both directions (positive and negative), based on the optimization done on the calibration set.

All models applying *APP* and *IIC* produced good results. The results can be compared appropriately by those from the study in (Zang et al.,

Table 7

Comparison of statistical quality of different models for skin sensitization: in the format sensitizer/non-sensitizer (i.e. the level 2).

Activity	n	Sensitivity	Specificity	Accuracy	MCC	Reference
LLNA	120	–	–	0.81	–	Zang et al. (2017)
Human outcomes	87	–	–	0.75	–	Zang et al. (2017)
LLNA	295	0.98	0.95	–	–	Bergal et al. (2020)
LLNA	–	0.70	0.79	0.45	–	Asturiol et al. (2016)
LLNA	–	0.91	0.64	0.82	–	Urbisch et al. (2016)
LLNA	–	0.98	0.85	0.93	–	Kumar et al. (2016)
LLNA	96	0.50	0.90	0.63	–	Emter et al. (2010)
LLNA	67	0.88	0.79	0.85	–	Natsch and Emter (2008)
LLNA	–	0.93	0.64	0.78	–	van der Veen et al. (2014)
LLNA	120	0.85	0.82	0.84	0.63	This work, Eq. (17)
Human outcomes	87	0.84	0.90	0.86	0.72	This work, Eq. (19)

2017) because the data used to build up the models are the same. In the case of the model for LLNA, the accuracy of the external validation set for our model is equal to 0.77 for level 1, and 0.87 for level 2. The value for level 2 in the literature (Zang et al., 2017) is 0.89, very close to our figure, and the overall accuracy (levels 1 and 2) is 0.77 (lower than the value we obtained).

In the case of the models for Dataset II, human data, the accuracy of our models on the external validation sets are 0.95 for level 1, and 0.86 for level 2. In the literature (Zang et al., 2017) the overall accuracy values (level 1 and 2) are 0.71 and 0.81, respectively. Thus, both values from the literature are lower.

Other models have given slightly higher accuracy, about 0.89 (Borba et al., 2021; Tung et al., 2018). However, it is hard to make comparisons, because the endpoints are different, and they cover several toxicological targets and datasets simultaneously.

However, our models do offer advantages. The integrated model, as in (Zang et al., 2017), uses a more complex approach, which implies machine learning tools and a series of values as input, including physico-chemical properties, and data from in chemico (direct peptide assay) and two in vitro tests. Thus, the strategy used in (Zang et al., 2017) calls for more effort and resources, with a large set of data, while our model only needs the molecular structure. Therefore, it can be used even without any preparation of the substance to be analyzed; this is not possible when experimental data are required. The complexity and availability of experimental values limit the other methods, relying on heterogeneous data needing experimental values. Furthermore, in our model the information relative to the chemical structure is very concise and simple, basically extracted from the SMILES, while in other cases a set of physicochemical parameters or molecular descriptors has to be obtained. Table 7 contains the comparison of the CORAL model with a similar model (sensitizer/non-sensitizer) suggested in the literature (Bergal et al., 2020; Asturiol et al., 2016; Urbisch et al., 2016; Kumar et al., 2016; Emter et al., 2010; Natsch and Emter, 2008; van der Veen et al., 2014).

The advantage of the CORAL model is applying exclusively data on molecular structure represented by SMILES together with data on an endpoint without additional data on 3D molecular structure, descriptors of quantum mechanics, or additional physicochemical and biochemical parameters.

On the other hand, the disadvantage of the CORAL model as well as the Monte Carlo approach, in general, is a dispersion of the statistical characteristics observed in several runs of the Monte Carlo optimization. Under such circumstances, the study of average values becomes the necessary practice. The statistical characteristics represented here are average on groups of corresponding runs of described versions of the Monte Carlo optimization (Tables 3–6). Supplementary materials section and freely available CORAL software (<http://www.insilico.eu/coral>) give the possibility to apply suggested models as well as to study runs of the described optimization. Finally, one can apply the CORAL software for other SAR/QSAR tasks.

5. Conclusions

The semi-correlation is a specific case among traditional correlations. Semi-correlations can serve as a tool for SAR/QSAR analysis. The suggested approach gives reasonably good models for skin sensitization, distinguishing sensitizers and non-sensitizers, and dividing sensitizers as weak or strong. Application of the self-organized vector of atom-pairs proportions improves the predictive potential of models for skin sensitization. Similarly, attempts to build up these models without the index of ideality of correlation indicated the advisability of including this criterion. The models we developed related to different kinds of data, such as LLNA, a widely used test, and data related to humans as well. The battery of results for skin sensitization should assist evaluators in their decisions. A major advantage of the present models is their intrinsic simplicity, since they only need the SMILES structure of the substances, while other models need multiple data (predicted and experimental) and complex information regarding the compounds.

Author contributions

Conceptualization, A.P.T., A.A.T., E.B.; data curation, A.P.T.; A.A.T.; E.B.; writing—original draft preparation, A.P.T.; A.A.T.; E.B.; writing—review and editing, A.P.T.; A.A.T.; E.B.; supervision, E.B.; project administration, E.B. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful for the contribution of the project LIFE-CONCERT (LIFE17 GIE/IT/000461) for financial support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.fct.2021.112580>.

References

- Asturiol, D., Casati, S., Worth, A., 2016. Consensus of classification trees for skin sensitisation hazard prediction. *Toxicol. Vitro* 36, 197–209. <https://doi.org/10.1016/j.tiv.2016.07.014>.
- Bergal, M., Puginier, M., Gerbeix, C., Groux, H., Roso, A., Cottrez, F., Milius, A., 2020. In vitro testing strategy for assessing the skin sensitizing potential of “difficult to test” cosmetic ingredients. *Toxicol. Vitro* 65, 104781. <https://doi.org/10.1016/j.tiv.2020.104781>.
- Borba, J.V.B., Braga, R.C., Alves, V.M., Muratov, E.N., Kleinstreuer, N., Tropsha, A., Horta Andrade, C., 2020. Pred-skin: a web portal for accurate prediction of human

- skin sensitizers. *Chem. Res. Toxicol.* 34 (2), 258–267. <https://doi.org/10.1021/acs.chemrestox.0c00186>.
- Chaudhry, Q., Piclin, N., Cotterill, J., Pintore, M., Price, N.R., Chrétien, J.R., Roncaglioni, A., 2010. Global QSAR models of skin sensitizers for regulatory purposes. *Chem. Cent. J.* 4 (Suppl. 1), S5. <https://doi.org/10.1186/1752-153X-4-S1-S5>.
- EC, 2009. European parliament, regulation (EC) No 1223/2009 of the European parliament and of the council of 30 november 2009 on cosmetic products. *Orkesterjournalen L* 342, 59, 22.12.2009.
- Emter, R., Ellis, G., Natsch, A., 2010. Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. *Toxicol. Appl. Pharmacol.* 245 (3), 281–290. <https://doi.org/10.1016/j.taap.2010.03.009>.
- Hoffmann, S., Kleinstreuer, N., Alépée, N., Allen, D., Api, A.M., Ashikaga, T., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Goebel, C., Kern, P.S., Klaric, M., Kühnl, J., Lalko, J.F., Martinozzi-Teissier, S., Mewes, K., Miyazawa, M., Parakhia, R., van Vliet, E., Zang, Q., Petersohn, D., 2018. Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database. *Crit. Rev. Toxicol.* 48 (5), 344–358. <https://doi.org/10.1080/10408444.2018.1429385>.
- IIVS, 2021. Institute for in vitro Sciences (Accessed March, 2021). <https://iivs.org/testing-services/assays/sensitization/>.
- Jaworska, J.S., Natsch, A., Ryan, C., Strickland, J., Ashikaga, T., Miyazawa, M., 2015. Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. *Arch. Toxicol.* 89 (12), 2355–2383. <https://doi.org/10.1007/s00204-015-1634-2>.
- Kleinstreuer, N.C., Hoffmann, S., Alépée, N., Allen, D., Ashikaga, T., Casey, W., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Göbel, C., Kern, P.S., Klaric, M., Kühnl, J., Martinozzi-Teissier, S., Mewes, K., Miyazawa, M., Strickland, J., van Vliet, E., Zang, Q., Petersohn, D., 2018. Non-animal methods to predict skin sensitization (II): an assessment of defined approaches. *Crit. Rev. Toxicol.* 48 (5), 359–374. <https://doi.org/10.1080/10408444.2018.1429386>.
- Kumar, P., Kumar, A., 2020a. Nucleobase sequence based building up of reliable QSAR models with the index of ideality correlation using Monte Carlo method. *J. Biomol. Struct. Dyn.* 38, 3296–3306. <https://doi.org/10.1080/07391102.2019.1656109>.
- Kumar, P., Kumar, A., 2020b. CORAL: QSAR models of CB1 cannabinoid receptor inhibitors based on local and global SMILES attributes with the index of ideality of correlation and the correlation contradiction index. *Chemometr. Intell. Lab. Syst.* 200, 103982. <https://doi.org/10.1016/j.chemolab.2020.103982>.
- Kumar, K.L.S., Tangadpalliwa, S.R., Desai, A., Singh, V.K., Jere, A., 2016. Integrated computational solution for predicting skin sensitization potential of molecules. *PloS One* 11 (6), e0155419. <https://doi.org/10.1371/journal.pone.0155419>.
- Kumar, P., Kumar, A., Sindhu, J., 2019. In silico design of diacylglycerol acyltransferase-1 (DGAT1) inhibitors based on SMILES descriptors using Monte-Carlo method. *SAR QSAR Environ. Res.* 30, 525–541. <https://doi.org/10.1080/1062936X.2019.1629998>.
- Natsch, A., Emter, R., 2008. Skin sensitizers induce antioxidant response element dependent genes: application to the in vitro testing of the sensitization potential of chemicals. *Toxicol. Sci.* 102 (1), 110–119. <https://doi.org/10.1093/toxsci/kfm259>.
- OECD, 2009. Test No. 429: Skin Sensitisation. OECD Guidelines for the Testing of Chemicals, Section, vol. 4. OECD Publishing, Paris, France. <https://doi.org/10.1787/9789264071100-en>.
- OECD, 2018. Organisation for economic Co-operation and development. Test No 442E. <http://www.oecd.org/env/test-no-442e-in-vitro-skin-sensitisation-9789264264359-en.htm> (Accessed March, 2021).
- Owen, J.L., Vakharia, P.P., Silverberg, J.I., 2018. The role and diagnosis of allergic contact dermatitis in patients with atopic dermatitis. *Am. J. Clin. Dermatol.* 19, 293–302. <https://doi.org/10.1007/s40257-017-0340-7>.
- REACH, 2016. https://echa.europa.eu/view-article/-/journal_content/title/register-ants-to-use-alternative-test-methods-for-skin-sensitisation.
- Roberts, D.W., Aptula, A.O., 2008. Determinants of skin sensitisation potential. *J. Appl. Toxicol.* 28, 377–387. <https://doi.org/10.1002/jat.1289>.
- Roberts, D.W., Aptula, A.O., Patlewicz, G.Y., 2011. Chemistry-based risk assessment for skin sensitization: quantitative mechanistic modelling for the SnAr domain. *Chem. Res. Toxicol.* 24, 1003–1011. <https://doi.org/10.1021/tx100420w>.
- Toropov, A.A., Toropova, A.P., 2019. QSAR as a random event: criteria of predictive potential for a chance model. *Struct. Chem.* 30, 1677–1683. <https://doi.org/10.1007/s11224-019-01361-6>.
- Toropov, A.A., Toropova, A.P., Selvestrel, G., Benfenati, E., 2019. Idealization of correlations between optimal simplified molecular input-line entry system-based descriptors and skin sensitization. *SAR QSAR Environ. Res.* 30, 447–455. <https://doi.org/10.1080/1062936X.2019.1615547>.
- Toropov, A.A., Toropova, A.P., Veselinović, A.M., Leszczynska, D., Leszczynski, J., 2020. SARS-CoV Mpro inhibitory activity of aromatic disulfide compounds: QSAR model. *J. Biomol. Struct. Dyn.* <https://doi.org/10.1080/07391102.2020.1818627>. Published online: 09 Sep. 2020.
- Toropova, A.P., Toropov, A.A., 2017a. Hybrid optimal descriptors as a tool to predict skin sensitization in accordance to OECD principles. *Toxicol. Lett.* 275, 57–66. <https://doi.org/10.1016/j.toxlet.2021.01.015>.
- Toropova, A.P., Toropov, A.A., 2017b. Hybrid optimal descriptors as a tool to predict skin sensitization in accordance to OECD principles. *Toxicol. Lett.* 275, 57–66. <https://doi.org/10.1016/j.toxlet.2017.03.023>.
- Toropova, A.P., Toropov, A.A., 2017c. CORAL: binary classifications (active/inactive) for drug-induced liver injury. *Toxicol. Lett.* 268, 51–57. <https://doi.org/10.1016/j.toxlet.2017.01.011>.
- Toropova, A.P., Toropov, A.A., Benfenati, E., 2019a. Semi-correlations as a tool to build up categorical (active/inactive) model of GABA_A receptor modulator activity. *Struct. Chem.* 30, 853–861. <https://doi.org/10.1007/s11224-018-1226-x>.
- Toropova, A.P., Toropov, A.A., Veselinović, A.M., Veselinović, J.B., Leszczynska, D., Leszczynski, J., 2019b. Semi-correlations combined with the index of ideality of correlation: a tool to build up model of mutagenic potential. *Mol. Cell. Biochem.* 452, 133–140. <https://doi.org/10.1007/s11010-018-3419-4>.
- Toropova, A.P., Toropov, A.A., Roncaglioni, A., Benfenati, E., 2021. The index of ideality of correlation improves the predictive potential of models of the antioxidant activity of tripeptides from frog skin (*Litoria rubella*). *Comput. Biol. Med.* 133, 104370. <https://doi.org/10.1016/j.combiomed.2021.104370>.
- Tung, C.-W., Wang, C.-C., Wang, S.-S., 2018. Mechanism-informed read-across assessment of skin sensitizers based on SkinSensDB. *Regul. Toxicol. Pharmacol.* 94, 276–282. <https://doi.org/10.1016/j.yrtph.2018.02.014>.
- Tung, C.-W., Lin, Y.-H., Wang, S.-S., 2019. Transfer learning for predicting human skin sensitizers. *Arch. Toxicol.* 93, 931–940. <https://doi.org/10.1007/s00204-019-02420-x>.
- Urbisch, D., Honarvar, N., Kolle, S.N., Mehling, A., Ramirez, T., Teubner, W., Landsiedel, R., 2016. Peptide reactivity associated with skin sensitization: the QSAR Toolbox and TIMES compared to the DPRA. *Toxicol. in Vitro* 34, 194–203. <https://doi.org/10.1016/j.tiv.2016.04.005>.
- van der Veen, J.W., Rorije, E., Emter, R., Natsch, A., van Loveren, H., Ezendam, J., 2014. Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals. *Regul. Toxicol. Pharmacol.* 69, 371–379. <https://doi.org/10.1016/j.yrtph.2014.04.018>.
- Weininger, D., 1988. SMILES, a chemical language and information system: 1: introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* 28, 31–36. <https://doi.org/10.1021/ci00057a005>.
- Wilm, A., Kühnl, J., Kirchmair, J., 2018. Computational approaches for skin sensitization prediction. *Crit. Rev. Toxicol.* 48 (9), 738–760.
- Wilm, A., Stork, C., Bauer, C., Schepky, A., Kühnl, J., Kirchmair, J., 2019. Skin doctor: machine learning models for skin sensitization prediction that provide estimates and indicators of prediction reliability. *Int. J. Mol. Sci.* 20 (19), 4833. <https://doi.org/10.3390/ijms20194833>.
- Zang, Q., Paris, M., Lehmann, D.M., Bell, S., Kleinstreuer, N., Allen, D., Matheson, J., Jacobs, A., Casey, W., Strickland, J., 2017. Prediction of skin sensitization potency using machine learning approaches. *J. Appl. Toxicol.* 37, 792–805. <https://doi.org/10.1002/jat.3424>.