

Quantitative structure–property relationships modeling of skin irritation

Sharath Golla, Sundar Madihally, Robert L. Robinson Jr., Khaled A.M. Gasem *

School of Chemical Engineering, Oklahoma State University, 423 Engineering North, Stillwater, OK 74078, USA

ARTICLE INFO

Article history:

Received 29 August 2008

Accepted 28 October 2008

Available online 5 November 2008

Keywords:

Skin irritation

Draize test

Primary irritation index

Non-linear QSPR models

Neural networks

ABSTRACT

Interest in developing procedures for estimating skin irritation potential of chemicals has been increasing as a result of concerns regarding animal welfare and costs involved in experimental irritation studies. In response to these concerns, a number of expert systems and quantitative structure–activity relationship (QSAR) models have been proposed for predicting the skin irritation potential of compounds. However, these models require as input independent estimates of several physiochemical properties. Hence, to predict skin irritation potential using these models often requires additional models capable of estimating the physiochemical properties of diverse structures; a requirement that most literature QSARs fail to meet.

In the work reported here, we developed a skin irritation QSPR model based on rabbit Draize test data for 186 compounds, which included chemicals from diverse molecular classes. The effectiveness of using a combination of traditional, functional group and structural descriptors has been studied. Our non-linear QSPR model is capable of predicting the skin irritation potential of chemical compounds with an R^2 of 0.78. Further, the final set of descriptors used to model skin irritation was analyzed to elucidate the effects of molecular size, reactivity and skin penetration on skin irritation.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Skin diseases and injuries are the most common job-related problems in industries such as manufacturing, food production, construction, machine tool operation, printing, metal plating, leather processing, engine service, landscaping, farming, or forestry. Since the skin is often exposed to chemicals, the potential for a chemical from a particular product to cause skin irritation needs to be evaluated as part of the overall safety assessment process. Common everyday routines like cleaning the house, working with paints or dyes, and gardening can lead to chemical exposure leading to skin irritation. As a result, all industrially important chemicals are assessed for their skin irritation or corrosive ability, and the results are listed in material safety data sheets (MSDS) to ensure the safety of the workers. According to the Organization for Economic Co-operation and Development (OECD) test guideline 404 (OECD, 1992), skin irritation is defined as “the production of reversible damage of the skin following the application of a test substance for up to 4 h”. For many years, the Draize rabbit skin test has been widely accepted for assessing the skin irritation potential of chemicals. In this test, 0.5 ml of the chemical of interest is

applied to albino rabbit skin for four hours and kept under clinical observation for 14 days for signs of erythema and edema (Draize et al., 1944). Although the Draize test provides reliable estimates of skin irritation, it is often criticized as being cruel to the test animals; hence, the incentive for developing predictive procedures for estimating skin irritation potential of chemicals has been increasing (Botham et al., 1998).

The relevance of the rabbit test to estimate human skin irritation hazards has been seriously questioned in the literature as the penetration and reactivity characteristics of rabbit and human skin are different. A detailed review of the limitations of the Draize test is provided by Robinson et al. (2001). In studying the skin penetration characteristics of human and rabbit skin, Bartek et al. (1972) found that human skin is more resistant to penetration than rabbit skin. Phillips et al. (1972) compared the rabbit and human skin response to certain irritants; the Draize rabbit test predicted accurately the severe human skin irritants and non-irritants, but it failed to predict accurately the mild and moderate skin irritants. Several chemicals considered unsafe by the rabbit test proved non-irritating to human skin, which can be attributed to higher penetration resistance of human skin. In the past, various *in vitro* and *in silico* skin irritation test methods have been proposed as replacements for *in vivo* tests; however, none of the alternate test procedures has been accepted as a valid replacement to the Draize test by OECD (Gerner and Schlede, 2002).

In the hazard identification of chemicals, the purpose of testing is to assess the irritation/corrosion potential according to

Abbreviations: ANN, Artificial neural networks; OECD, Organization for economic co-operation and development; QSPR, Quantitative structure–property relationship.

* Corresponding author. Tel.: +1 405 744 5280; fax: +1 405 744 6338.

E-mail address: gasem@okstate.edu (K.A.M. Gasem).

classification schemes defined by regulatory authorities. Current regulatory proposals recommend a tiered (stepwise) approach to hazard identification in which chemicals can be classified as irritating/corrosive on the basis of results from non-animal methods. Since testing on animals is only required as a last step to confirm negative results generated by non-animal tests, these testing strategies contribute to the reduction and refinement of *in vivo* tests. The non-animal methods act as a partial replacement of animal testing and are used to place chemicals into two or more categories of irritation/corrosion potential, without generating too many false-positive results. Promising *in vitro* methods that have been evaluated include EpiDerm, EPISKIN, PREDISKIN, the non-perfused pig ear model, and the mouse skin integrity function test (SIFT) (Botham et al., 1998). *In silico* methods for predicting skin irritation include expert systems and structure–activity models that express the skin irritation potential of a chemical as a function of a set of physiochemical properties and structural descriptors. The application of these *in silico* models is limited by the availability of the necessary input physiochemical properties, such as octanol/water partition coefficient, melting point, lipid solubility, aqueous solubility, surface tension and vapor pressure. Often the properties of a novel chemical are not readily available, rendering these *in silico* models inapplicable. The limited applicability of these procedures necessitates the development of novel prediction methods that use only structural descriptors to estimate the skin irritation potential.

Quantitative structure–property relationship (QSPR) modeling offers an attractive alternative because successful models have the potential to provide reliable property estimates based solely on chemical structure information. That is, if structural information can be successfully decoded, the properties can be determined simply from the chemical structure.

QSPR studies have gained impetus with the advent of high-speed computers. This has led to both the development of new and more complex molecular descriptors and to the application of QSPR models to properties that were previously infeasible due to computational intensity. QSPR models are now well established and are used to correlate varied, and often complex, physiochemical properties of molecules. A detailed review of its applications can be found elsewhere (Godavorthy, 2004; Godavorthy et al., 2008).

As outlined in the next section, various QSPR models have been proposed for predicting skin irritation potential of molecules. However, these models are restricted to a limited number of classes of molecules, thus lacking general applicability. Hence, we have developed a skin irritation QSPR model using rabbit Draize test data for 189 compounds consisting of chemicals from various classes. We have evaluated the predictive ability of several QSPR models based on literature descriptors, group contribution descriptors and structural descriptors, used alone or in combination. A QSPR model utilizing only structural descriptors was developed first, then its predictive ability was improved by using a combination of literature, functional group and structural descriptors. Robust artificial neural network models with superior capabilities were used for non-linear model development. The physical significance of the final set of descriptors was studied and external validation of the QSPR model was performed using data from human-patch tests.

2. Materials and methods

2.1. Skin irritation models

Although *in vivo* tests for skin irritation cannot be replaced entirely, they can be reduced significantly by initial screening of

chemical compounds using *in vitro* and *in silico* methods. While a number of *in vitro* methods have been proposed in the literature, to date there are no validated *in vitro* tests that can replace the Draize test. A detailed review of the currently available *in vitro* techniques is given elsewhere (Chew and Maibach, 2006; Welss et al., 2004). The *in silico* techniques offer an attractive alternative due to their ease of use and low cost relative to the *in vitro* techniques. In this section, a review of the *in silico* techniques available in the literature is presented. A comprehensive review of the available *in silico* techniques for predicting skin irritation potential of chemicals has been reported by Saliner et al. (2008).

Physiochemical properties found to be significant for modeling skin irritation include molecular volume, dipole moment, molecular shape, the overall number of atoms, steric effects, molar refractivity, acidic dissociation constant (pK_a), absolute hardness, and the octanol/water partition coefficient ($\log K_{ow}$) (Saliner et al., 2008). Among these properties, $\log K_{ow}$ has been found to be most significant since it provides a quantitative measure of partitioning between aqueous and lipid phases (Combes and Judson, 1995). One of the earliest attempts to model skin irritation using structure–activity relationships (SARs) was done by Enslein et al. (1987). Separate models were developed for aromatic and aliphatic compounds using molecular connectivity indices, sub-structural keys and molecular length parameters. Since the database used did not provide numerical scores, the models use skin irritation severity ratings. Smith et al. (2000b) developed a SAR model to discriminate skin irritant esters from non-irritant esters using nineteen physiochemical parameters that represent transport, electronic and steric properties. Human skin irritation data of 42 esters were used to generate ten sub-models using multiple random sampling of the database. The sensitivity and specificity values ranged from 0.846 to 0.923 and 0.615 to 0.923, respectively. The results indicate that physiochemical parameters of esters relate to their skin irritation effects in humans, and chemical partitioning and intermolecular reactions are important components of the response. Smith et al. (2000a) developed an iterative SAR model for human skin irritation. The model predictions were validated experimentally, and these test results were incorporated into the database to refine the model. A total of 34 irritants from the rabbit test were selected, of which 16 were predicted by the model to be positive and 18 to be negative. These chemicals were further tested experimentally using human-patch tests, and the results were incorporated into the database to refine the model. However, since the SAR models were based on limited data, the accuracy of the models was not satisfactory, thus emphasizing the need for experimental validation of models and their further refinement as new data become available.

Hayashi et al. (1999) studied QSAR for skin irritation potential for 24 phenols. Absolute hardness, LUMO energy and $\log K_{ow}$ were used to fit a regression function to the skin irritation scores obtained from a rabbit Draize test study. An R value of 0.85 was obtained. The model predictions were further validated using a set of six additional phenols, and good correlations with the expected skin irritation scores were observed. Berner et al. (1994) studied the influence of pK_a on skin irritation and found the two to be highly correlated for the chemical compounds examined. Kodithala et al. (2002) used a membrane-interaction QSAR (MI-QSAR) technique to predict the skin irritation potential of 20 hydroxy organic compounds. The MI-QSAR skin irritation predictions were compared to the traditional 2-D-QSAR predictions to prove the superiority of MI-QSAR approach.

In addition to the QSAR models, a number of expert systems that use a knowledge base to determine the skin irritation potential of a chemical have been developed. These expert systems find the skin irritation potential of a chemical either by relating the existing physiochemical properties of chemicals or through pattern

recognition algorithms. A number of papers describing the currently available expert systems have been published (Botham et al., 1998; Combes and Judson, 1995; Dearden et al., 1997; Saliner et al., 2008). Some of the widely used expert systems are German Federal Institute for Risk Assessment Decision Support System (BfR-DSS), DEREK for Windows (DEREKfW), HazardExpert, MULTICASE, Computer-Aided Chemistry (CACHe) program, Substructure-Based Computerized Chemical Selection Expert System (SuCCSES), OASIS, and TOXicity Prediction Komputer-Assisted Technology (TOPKAT). Saliner et al. (2008) have provided a detailed review of each of these expert systems. The expert systems are applicable to a wide range of chemical structures in contrast to the currently available QSAR models. However, the application of these models to predict skin irritation requires as input several physiochemical properties such as octanol/water partition coefficient, melting point, lipid solubility, aqueous solubility, surface tension and vapor pressure. Thus, prediction of skin irritation potential using these models often requires other models capable of estimating the required input physiochemical properties.

2.2. QSPR methodology

The development of a QSPR model for skin irritation involves several distinct steps: (a) compilation of an experimental dataset, (b) generation and optimization of 3-D molecular structures, (c) calculation of molecular descriptors, (d) reduction of the number of descriptors, and (e) development of a regression model. To begin, a dataset of molecules chosen from reliable sources must be assembled. Once this dataset is analyzed using adequate ranges of both target property values and chemical diversity, the 3-D structures of the molecule are generated using commercial molecular visualization software. Structure optimization and descriptor generation are performed using commercial QSPR software. Model development usually begins with linear analysis using algorithms such as multiple-linear regression analysis, principal-component analysis, heuristic analysis or partial least squares. In the optimal scenario, the best set of descriptors is then used to develop a non-linear QSPR model using artificial neural network algorithms.

2.2.1. Dataset compilation

High-quality data generated through good laboratory practices and complying with the OECD test guidelines have been used for modeling. The majority of the data are acquired from Bagley et al. (1996), who cite Draize skin irritation test data for 176 chemicals. However, since 29 structures were tested more than once, a total of 215 data points are obtained. Other sources of Draize test data include publications by Hayashi et al. (1999) and Kodithala et al. (2002), who report skin irritation data for 30 phenols and 22 hydroxy organic compounds, respectively. The primary irritation index (PII) scores reported in these sources were used for modeling.

2.2.1.1. Oklahoma State University (OSU) Draize test database. A database containing experimental Draize skin irritation test data of 205 molecules has been compiled from the above sources. The database contains experimental data including the chemical classes represented by acids, acrylates, alcohols, aldehydes, amides, amines, brominated derivatives, chlorinated solvents, esters, ethers, fatty acids, halogenated aromatics, hydrocarbons, inorganics, ketones, nitrils, phenols, sulphur containing compounds, and triglycerides. The skin irritation effects of the chemicals have been graded according to the scale proposed by OECD test guideline 404 (Draize et al., 1944; OECD, 1992), as shown in Table 1. The chemicals are graded for erythema and edema using experimental data and a PII is calculated from these grades using the following equation:

Table 1

Grading scale for skin reactions OECD (1992).

Erythema and eschar formation grade	Scale
No erythema	0
Very slight erythema (barely perceptible)	1
Well defined erythema	2
Moderate to severe erythema	3
Severe erythema (beet redness) to slight eschar formation (injuries in depth)	4
Edema formation grade	
No edema	0
Very slight edema (barely perceptible)	1
Slight edema (edges of area well defined by definite raising)	2
Moderate edema (raised approximately 1 mm)	3
Severe edema (raised more than 1 mm and extending beyond the area of exposure)	4

Primary Irritation Index (PII)

$$= \frac{\sum \text{erythema grades at 24/48/72 h} + \sum \text{edema grades at 24/48/72 h}}{3 \times \text{number of animals}} \quad (1)$$

The numerical value of PII ranges from 0 to 8, where chemicals with a PII value of zero have no erythema and edema formation, and those with a value of eight have severe effects. Chemicals for which there are multiple test data, the PII values were averaged to obtain a unique skin irritation score. Since the Draize test data available in the open literature were limited to a few hundred compounds, efforts were made to utilize all reliable experimental data for modeling. Molecules for which the PII could not be determined due to the severity of effects were given a score of eight, the maximum possible PII value. Experimental data for chemical mixtures and fragrance oils were not included in the database.

Of the 205 molecules, only 189 structures were considered for modeling. The other 16 molecules were rejected due to one the following reasons: (a) a structure could not be found, (b) the structure could not be optimized using Chem3-D (ChemBioOffice 11.0, 2008), or (c) molecular descriptors could not be generated for the structure.

For a model to be able to predict skin irritation efficiently, the range of potencies used to train the model should be adequate. Fig. 1 shows the distribution of PII values for the data considered in the present modeling.

2.2.2. Structure generation and optimization

After database compilation, the next step in QSPR modeling is the generation of a chemical structure for each molecule included

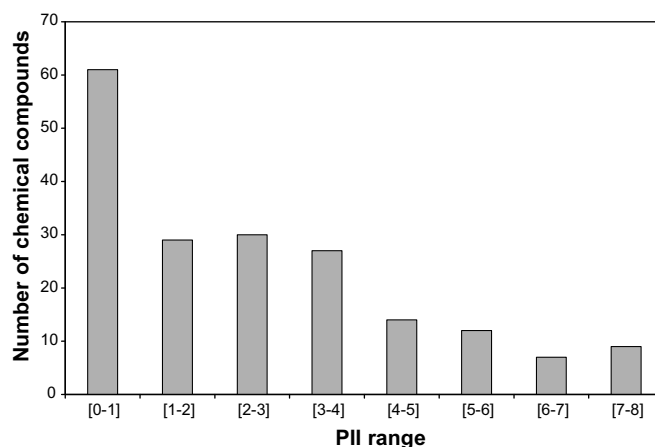


Fig. 1. Distribution of PII values for the data used for QSPR modeling.

in the modeling effort. Various schemes for chemical representation have been proposed in the literature. For example, Open Babel software (Morley, 2006) includes about 80 different representations for a given molecular structure. The most common way of representing a chemical is a two-dimensional (2-D) sketch. However, using a 2-D representation does not provide complete description of the molecule and lacks shape and surface distribution information. To develop an effective QSPR model, the representations of the molecules should provide all the necessary structural information. This requires information about the atoms present, along with three dimensional (3-D) coordinates that provide a full spatial depiction of the molecule. A commercial package, ChemDraw, was used to generate the 2-D structures of the molecules. The 3-D structures were generated for these molecules using Chem3-D Ultra (Chembiooffice 11.0, 2008, Cambridge Software). Since more than one set of 3-D coordinates that satisfy the structural constraints (bond length and bond angle) can be generated for a given molecule, the conformation with the lowest energy must be located. The structures were initially optimized using the Chem3-D module available in Chem3-D Ultra. To locate the lowest energy configuration, multiple initializations were used during the structure optimization. AMPAC 6.0 (AMPAC 6.0., 1997, Semichem Inc.) was then used to further refine the 3-D geometry of the structures. The output files from AMPAC were used to calculate various descriptors.

2.2.3. Descriptor generation

The final optimized structure from AMPAC was used as input to commercial QSPR software that can generate over 1200 molecular descriptors for a given molecule. A variety of constitutional, topological, geometrical, thermodynamic, quantum-chemical and electrostatic descriptors are generated by CODESSA (CODESSA 2.7.8., 2007, Semichem Inc.) and 154 functional group descriptors are generated by Dragon (Dragon 5.0, 2007). The number of descriptors calculated for a given molecule depends on the structural complexity of the molecule. Descriptors that were not pertinent for a given molecule were set to zero in the subsequent QSPR model development.

2.2.4. Descriptor reduction

Not all the descriptors generated for a molecule are significant in modeling a particular property. Retaining a large number of the available descriptors in model development efforts causes dimensionality problems and diminishes the performance of a QSPR model, especially when non-linear algorithms are used in model development (Godavarthy, 2004). Descriptor reduction is the process of automating the discovery of potentially useful correlations from large sets of descriptor data. This process involves the identification of the most relevant set of descriptors for model development and is an important step in QSPR modeling (Jegadeesan, 2006). Different methods for reduction are available in the literature. The most widely used techniques are the principal-component analysis (PCA), partial least squares (PLS), genetic algorithms (GA), and neural networks (NNs) (Godavarthy, 2004). Most existing QSPR models are multiple-linear regression correlations, which require *a priori* assumption of a specific form for the mathematical relationship between the property and its descriptors. However, linear regression analysis ignores any non-linear relationships that might exist between the descriptors and properties. The use of linear approaches often leads to loss of critical information and results in models with poor predictive abilities (Godavarthy et al., 2008).

To ensure that the non-linear relationships are accounted for in the QSPR models, non-linear transformations of all the descriptors were calculated (including the log and the inverse of the descriptors), thus an expanded set of descriptors was generated. This

expanded set was used to find the best set of descriptors through sequential selection. The best descriptor from the expanded set was selected and combined sequentially with the remaining descriptors. Then, the best two-descriptor combination was retained and combined with each of the remaining descriptors. The sequence continued until a set of 40 descriptors was identified. Heuristic regression available in CODESSA was used to further reduce the number of descriptors. The optimal set of descriptors was retained for artificial neural network (ANN) analysis.

2.3. Literature descriptors

If a mechanistic interpretation of the relationship between a property and its descriptor set can be formulated, it provides increased confidence in the accuracy and validity of the model (Barratt, 2000). However, such interpretation is often not possible for all the descriptors, due to the complexities involved. A detailed review of the literature points to chemical reactivity, molecular size, and skin permeability as important determinants of skin irritation. The chemical reactivity of a molecule can be determined using the HOMO and LUMO energies. Molecules with low HOMO–LUMO energy gaps have low kinetic stability and are favorably inclined to add or remove electrons, thus forming activated complexes (Aihara, 2000). Although skin permeability is significant in modeling skin irritation, the absence of readily available skin permeation prediction models often leads to the use of log K_{ow} values to represent the skin permeation ability of a chemical.

2.4. QSPR model development

The objectives of the present work were (a) demonstrating the ability of non-linear QSPR modeling to predict the skin irritation potential of a chemical compounds, (b) examining the efficacy of using a combination of literature, functional group and structural descriptors for model development, (c) using descriptors from multiple QSPR software to identify the most significant descriptors, and (d) validating the final QSPR model using an external dataset. The following specific case studies were conducted to achieve these objectives:

- Case 1: Develop a QSPR model using literature descriptors.
- Case 2: Identify any significant functional group contributions using functional group descriptors from Dragon.
- Case 3: Develop a QSPR model using only structural descriptors from CODESSA.
- Case 4: Develop a final robust QSPR model using a combination of literature, functional group and structural descriptors.
- Case 5: Validate the final QSPR model using skin irritation data from other experimental studies.

These case studies, conducted in the sequence above, provided valuable guidance to the development of the final QSPR model.

2.4.1. Linear model development

Selecting the set of descriptors that best encodes information on the property of interest is a difficult optimization problem. Fast and reliable regression techniques are required to obtain a reduced set of descriptors after discarding both the correlated and the insignificant descriptors through orthogonalization and non-linear sequential reduction (Godavarthy, 2004). Multi-linear techniques available in commercial software packages were used to obtain linear regression models. The following linear regression analysis techniques are provided in CODESSA: (a) multi-linear regression, (b) principal-component regression, (c) partial least squares

regression, and (d) heuristic regression. In the search for the best multi-linear correlation for a large set of descriptors, the major problem is mutual collinearity of descriptors, which leads to instability of regression coefficients, overestimated standard errors, and critical loss of predictive information (Godavorthy, 2004). One way to avoid this effect is to transform the non-orthogonal set of the natural descriptors into a set of orthogonal factors. The latter, being linear combinations of the natural descriptors, preserves the information content while removing the negative effects of the multi-collinearity from the subsequent regressions. In this work, the heuristic regression analysis technique available in CODESSA was used to obtain the best linear regression model after orthogonalization and non-linear sequential reduction of descriptors.

2.4.2. Non-linear model development

Linear regression analysis ignores any non-linear relationships that may exist between property and descriptors. To ensure that non-linear relationships are accounted for in the QSPR models, the optimal descriptor set from the heuristic analysis was retained for ANN analysis.

ANN Analysis: The efficacy and efficiency of supervised learning in multilayer neural networks depends strongly on the network topology, the transfer functions, the learning rule, and the initial values of the weights. Optimal instances for these items are usually unknown *a priori* because they depend mainly on the particular training set to be considered and on the nature of the solution. A feed forward back-propagation neural network model with an improved network architecture, consistency, data randomization, allocation of training/validation data, and automated network initialization was used for the non-linear model development (Demuth et al., 2008). The model searched for all possible combinations of hidden layer units that result in a degree of freedom ratio (ratio of the number of network connections and the number of data points) value greater than two. The input dataset was divided into training, cross-validation and testing sets with 70% of data assigned to training set, 15% to cross-validation set and 15% to testing set. Overtraining of the network, which results in poor predictive capability, was avoided by use of the cross-validation set with an early-stopping method. The training set was used for computing the gradient and updating the network weights and biases. The cross-validation set chosen was representative of all molecular types in the training set for optimal performance. The error on the cross-validation set was monitored during the training process. The cross-validation error normally decreases during the initial phase of training, as does the training set error. However, when the network begins to over-fit the data, the error in the cross-validation set typically begins to rise. When the cross-validation error increased for a specified number of iterations, the training was stopped, and the weights and biases at the minimum cross-validation error were retained. Once the weights and biases are determined, the testing set provides a means of evaluating the network predictive capability.

By training the network starting from several different initial conditions, the robustness of the network performance can be verified. Multiple randomizations of the data and initializations of the weights were used to obtain the best network, as suggested by Iyer

and Rhinehart (Iyer and Rhinehart, 1999). The inputs and targets were normalized to have zero mean and unity standard deviation, which ensures that exceptionally large-valued descriptors do not bias the network. The Nguyen-Widrow algorithm was used to initialize weights and biases, which were updated using a Levenberg-Marquardt optimization technique. The transfer functions and the performance function of the network were tailored to find the best possible network.

The final network was evaluated further using correlation coefficient (R^2), sum-of-squared errors, average-absolute deviations, weighted-average-absolute deviations, root-mean-squared error (RMSE), and number of wrong classifications, when applicable.

3. Results and discussion

Schultz et al. (2003) suggested that an ideal QSPR model should: (1) consider an adequate number of molecules for sufficient statistical representation, (2) have a wide range of quantified toxic potency, and (3) yield to mechanistic interpretation. In this work, skin irritation data for 189 molecules of diverse classes with a wide range of potency distribution have been considered. Further, to interpret mechanistically the final set of descriptors, the correlation of the descriptors with the molecular size, reactivity and penetration ability (factors that are believed to highly influence the skin irritation potential) was studied. Four case studies were conducted to build an effective skin irritation QSPR model. Each case study, analyzed in sequence, provided valuable guidance to the development of the final QSPR model.

Table 2 presents a brief summary of the results obtained for the Cases 1–4. Initially, models were developed using only the literature descriptors (Case 1). Although the performance of these models was comparable to the models cited in the literature, errors exceeded the desired level. Therefore, approximately 1200 structural descriptors and 154 functional descriptors were analyzed to obtain a QSPR model with the best predictive ability. Non-linear transformations of the descriptors were evaluated to identify any non-linear property-descriptor relationships during descriptor reduction. Although the inclusion of these descriptors improved the model predictions, additional improvement was desired, and hence, non-linear neural network based models were developed.

Our first non-linear QSPR model used only descriptors identified in the literature. Fig. 2 illustrates the property predictions obtained in Case 1. The symbols in the plots denote the property estimate obtained from the model while the solid line represents perfect predictions (45° line). R^2 and RMSE values of 0.41 and 1.7 were obtained, respectively. The results show that using only the literature descriptors for modeling skin irritation is inadequate, and additional structural descriptors are required to account for skin irritation effects of diverse chemical classes.

The functional group contributions to the property predictions were studied using 154 functional group descriptors evaluated using Dragon (Case 2). The significant functional groups identified in this case study are listed in Table 3, and Fig. 3 presents the property predictions. Details on the functional groups used for model development are illustrated in the Dragon manual. To further

Table 2
Summary of results obtained for modeling of skin irritation.

Case	Descriptors used	N ^a	Nd ^b	Neural network architecture	RMSE	R ²
1	Literature	189	4	4-7-1-1	1.7	0.41
2	Functional group	189	16	16-5-1-1	1.3	0.62
3	Structural	189	19	19-4-1-1	1.2	0.69
4	Literature, functional, and structural	186	13	13-3-8-1	1.1	0.78

^a N = Number of molecules.

^b Nd = Number of descriptors.

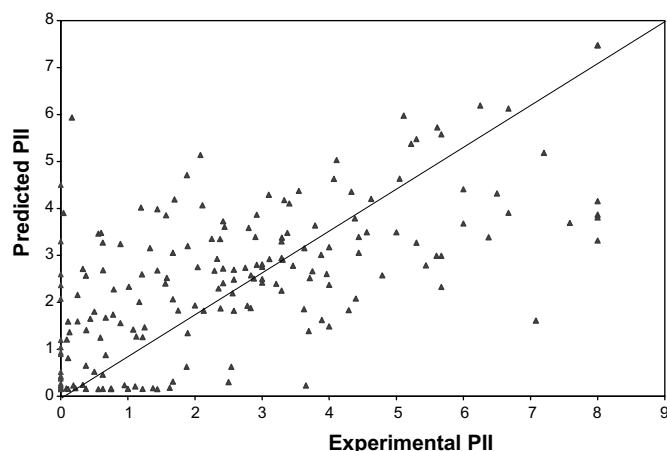


Fig. 2. QSPR model predictions using literature descriptors (Case 1).

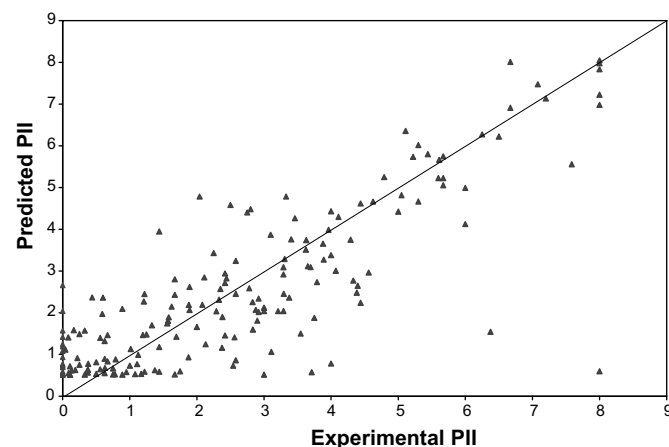


Fig. 4. QSPR model predictions using structural descriptors from CODESSA (Case 3).

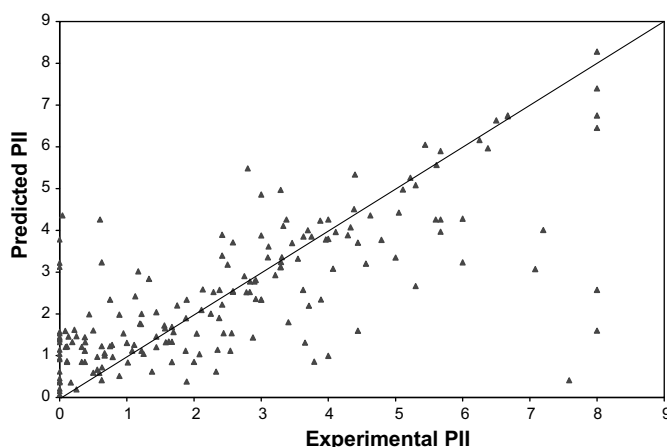


Fig. 3. QSPR model predictions using functional group descriptors (Case 2).

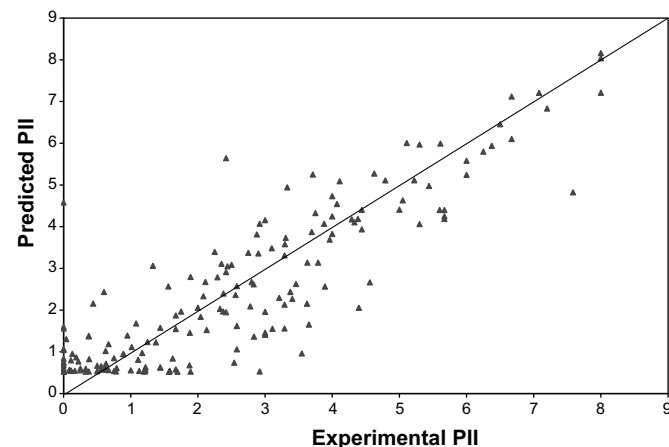


Fig. 5. QSPR model predictions using literature, functional group and structural descriptors (Case 4).

improve the property predictions, 1200 structural descriptors were generated using CODESSA, and a non-linear prediction model was developed (Case 3). The descriptors used and the model predictions are illustrated in Table 3 and Fig. 4, respectively. R^2 and RMSE values of 0.69 and 1.2 were obtained, respectively. This model used

only structural descriptors; hence, application of this model to any new structure requires no *a priori* estimation of physiochemical properties of chemicals.

Table 3
Descriptors used for model development.

Case 1	Case 2	Case 3	Case 4
HOMO energy	nHAcc	FNSA-2 Fractional PNSA (PNSA2/TMSA) (Quantum-chemical PC)	nHAcc
LUMO energy	nRNH2	Avg 1-electron react. index for a O atom	Min nucleoph. react. index for a N atom
Log K_{ow}	nRNR2	Tot molecular 2-center resonance energy	Min partial charge for a C atom [Zefirov's PC]
Molecular weight	nCH2RX	Avg nucleoph. react. index for a N atom	Max partial charge for a H atom [Zefirov's PC]
	nArX	Moment of inertia C	Tot molecular 2-center resonance energy
	nRCN	RNCS Relative negative charged SA SAMNEG* RNCG (Zefirov's PC)	Avg 1-electron react. index for a O atom
	nCrt	Min electroph. react. index for a Cl atom	Average Structural Information content (order 1)
	nR=CX2	DPSA-1 Difference in CPSAs (PPSA1-PNSA1) (Quantum-Chemical PC)	nArOR
	nCconj	Kier shape index (order 3)	Relative number of double bonds
	nCrq	Tot dipole of the molecule	nCRX3
	nCRX3	Kier shape index (order 2)	nCrt
	nROH	PNSA-2 Total charge weighted PNSA (Quantum-Chemical PC)	nOHs
	nCp	Min 1-electron react. index for a S atom	nCp
	nCs	Min total interaction for a C–C bond	
	nArOH	ALFA polarizability (DIP)	
		Max 1-electron react. index for a N atom	
		Min 1-electron react. index for a O atom	
		Max net atomic charge for a F atom	
		DPSA-1 Difference in CPSAs (PPSA1-PNSA1) (Zefirov's PC)	

A final model, combining the literature, functional and structural descriptors was developed to obtain the best predictive ability (Case 4). Three molecules that were assigned a PII score of eight were found to be outliers and subsequently removed. A neural network model with a 13–3–8–1 architecture utilizing 13 descriptors was found to give the best predictions for the remaining 186 chemical compounds. Table 3 presents the final set of descriptors used for model development. The final model for skin irritation has an R^2 of 0.78. Further, the RMSE value obtained was 1.1, showing the efficacy of using a combined set of descriptors. Fig. 5 shows the property estimates obtained from the final QSPR model.

An ideal QSPR model should yield to mechanistic interpretation of the input parameters used for model predictions. For this, the descriptors obtained were further investigated to identify any correlation with molecular size, reactivity and skin permeation ability

using CODESSA. The values of $\log K_{ow}$ and $\log K_p$ for the chemical compounds were obtained using QSPR prediction models developed previously by our group. Fig. 6 illustrates that the descriptors in the final model are indeed highly correlated with these properties.

To further validate the predictive capability of the QSPR model, an external dataset from Tornier et al. (2006), not included in the model development, was used. The external dataset contained 22 chemical compounds. A list of the molecules used for the external validation is provided in Table 4. This dataset differed from the one that was used for model development in that it contained primarily skin irritation classification values from European Union (EU), human patch, SkinEthic direct application and SkinEthic patch test (Tornier et al., 2006). Fig. 7 shows a comparison of the classification system used in the EU and human-patch test. More details on the

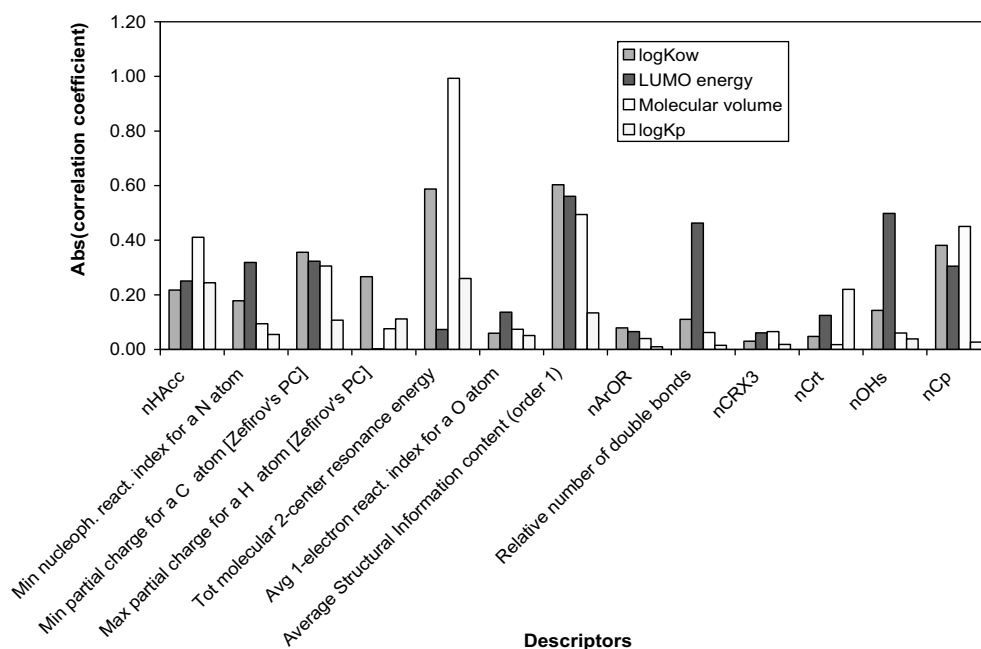


Fig. 6. Correlation of the final set of descriptors with the molecular size, reactivity and skin penetration ability.

Table 4

External validation of the model and corresponding experimental data.

Chemical compound	EU class	Human-patch class	SkinEthic direct application class	SkinEthic patch class	Predicted PII	Predicted class
2-Propanol	NC	NC	I	NI	0.6	NI
Isopropyl palmitate	NC	NC	NI	NI	3.9	I
Dimethyl sulphoxide	NC	R38	I	NI/I	5.1	I
Lactic acid	NC	R38	I	I	0.5	NI
Triethanolamine	NC	NC	I	NI	1.1	NI
Dodecanol	NC	NC	NI	NI	3.3	I
Tween 80	NC	NC	NI	NI	5.9	I
Propylene glycol	NC	NC	NI	NI	0.5	NI
Ethanol	NC	NC	I	NI	2.2	I
Octanoic acid	R34	R38	I	I	3.6	I
Heptanoic acid	R34	R38	I	I	4.8	I
1-Decanol	R38	NC	I	I	3.9	I
Decanoic acid	R38	R38	I	I	2.6	I
Dodecanoic acid	R38	NC	I	I	2.4	I
N,N-Dimethyl-N-dodecylaminobetaine	R38	R38	I	NI/I	5.8	I
Acetic acid	R38	NC	I	I	2.8	I
Hydrochloric acid	R38	NC	I	NI	3.2	I
Benzalkonium chloride	R38	R38	I	I	2.9	I
Octanol	R38	NC	I	I	4.9	I
Geraniol	R38	NC	I	I	2.7	I
Linalyl acetate	R38	NC	I	I	4.7	I
Hexanol	R38	NC	I	I	6.8	I

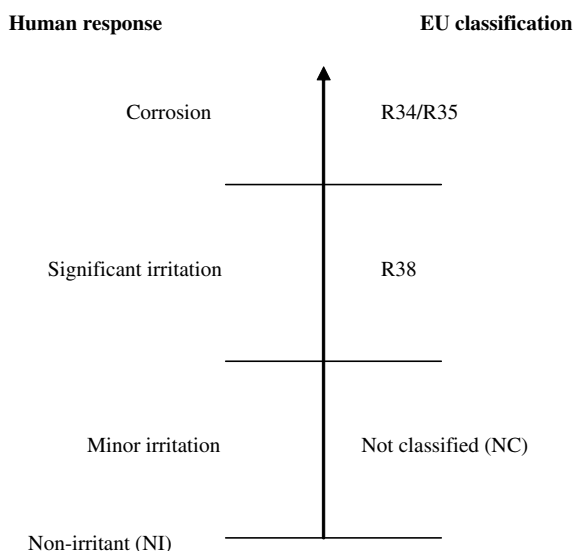


Fig. 7. Classification of skin irritation hazard (Basketter et al., 1997).

classification used are provided elsewhere (Basketter et al., 1997). The predictions obtained from the model were compared to the experimental results reported in the external dataset as illustrated in the Table 4. Classifying the molecules with PII in the range 0–2 as non-irritants (NI) and 2–8 as irritants (I), the model predictions were found to be in good agreement with the experimental data for a large part of the dataset. Of the 22 molecules validated, lactic acid is the only molecule that was not consistent with any of the test procedures. Although the secondary validation process provides valuable insights regarding the predictive ability of the developed QSPR model, only a guarded judgment can be made based on this study since the test data reported in the external dataset mostly come from different experimental techniques.

Table 5 illustrates a comparison of the current QSPR model results with literature skin irritation prediction models. The current model for skin irritation improves on other similar literature models in several aspects, including (a) the use of a large dataset consisting of a wide array of functional groups of significance to most chemical processes; (b) use of descriptors from multiple QSPR software to assure model superiority and stability; (c) use of non-linear transformations during descriptor reduction to obtain the most suitable set of descriptors; (d) mechanistic interpretation of the final non-linear QSPR models; and (e) extensive validation of models to assure robustness and predictive ability.

Table 5
Comparison of current QSPR model with similar literature models.

Author (reference)	No. of compounds	Chemical class	No. of inputs to the model	Model type	R^2
Hayashi et al. (1999)	13	Phenols	2	Linear	0.67
Kodithala et al. (2002)	11	Phenols	2	Linear	0.52
	20	Hydroxy alcohols	1	Linear	0.54
	20	Hydroxy alcohols	2	Linear	0.62
	20	Hydroxy alcohols	3	Linear	0.76
	20	Hydroxy alcohols	4	Linear	0.86
	13	Aliphatic alcohols	3	Linear	0.95
	9	Phenols	4	Linear	0.94
This work	186	Diverse classes	13	Non-linear	0.78

As evidenced by the results, the QSPR model is capable of predicting the skin irritation potential of a diverse set of molecules with varying structural complexities.

4. Conclusions

Following are the conclusions drawn based on this study:

1. The QSPR model was able to predict the skin irritation potential of diverse chemical compounds successfully, with an R^2 of 0.78.
2. A combination of literature, functional group, and structural descriptors are effective in QSPR modeling of skin irritation.
3. The final set of descriptors obtained showed good correlation with the molecular size, reactivity and skin penetration characteristics of chemical compounds, thus providing a mechanistic interpretation.
4. The final QSPR model was effective in estimating the skin irritation potential of diverse chemical compounds for an external dataset containing 22 compounds.
5. Although the database employed contained experimental data for diverse compounds, expansion of the database as additional experimental data become available would facilitate further development of the model.
6. The accuracy and reliability of the QSPR model depends on the availability and accuracy of the experimental data. As the skin penetration and reactivity characteristics of human skin and rabbit skin are different, development of reliable test procedures that mimic the human skin more closely would further benefit our modeling capability.

Conflict of interest statement

The authors state that there are no conflicts of interest with the publication of this manuscript.

Acknowledgement

Financial support for this research was provided by the National Institutes of Health; the National Institute of Biomedical Imaging and Bioengineering (1R21EB005749).

References

- Aihara, J., 2000. Correlation found between the HOMO–LUMO energy separation and the chemical reactivity at the most reactive site for isolated-pentagon isomers of fullerenes. *Physical Chemistry Chemical Physics* 2, 3121–3125.
- Bagley, D.M., Gardner, J.R., Holland, G., Lewis, R.W., Regnier, J.F., Stringer, D.A., Walker, A.P., 1996. Skin irritation: reference chemicals data bank. *Toxicology in Vitro* 10, 1–6.
- Barratt, M.D., 2000. Prediction of toxicity from chemical structure. *Cell Biology and Toxicology* 16, 1–13.
- Bartek, M.J., Labudde, J.A., Maibach, H.I., 1972. Skin permeability *in vivo*: comparison in rat, rabbit, pig and man. *Journal of Investigative Dermatology* 58, 114–123.
- Basketter, D.A., Chamberlain, M., Griffiths, H.A., Rowson, M., Whittle, E., York, M., 1997. The classification of skin irritants by human-patch test. *Food and Chemical Toxicology* 35, 845–852.
- Berner, B., Nangia, A., Maibach, H.I., 1994. Influence of pKa on skin irritancy potential of chemicals. *In Vitro Skin Toxicology, Irritation, Phototoxicity, Sensitization*, 29–36.
- Botham, P.A., Earl, L.K., Fentem, J.H., Roguet, R., van de Sandt, J.J.M., 1998. Alternative methods for skin irritation testing: the current status. *Alternatives to Laboratory Animals* 22, 195–211.
- ChemBioOffice 11.0, 2008. Cambridge Software, Cambridge, MA.
- Chew, A.L. and Maibach, H.I., 2006. *In vitro* methods to predict skin irritation. *Irritant Dermatitis*.
- Combes, R.D., Judson, P., 1995. The use of artificial intelligence systems for predicting toxicity. *Pesticide Science* 45, 179–194.
- Dearden, J.C., Barratt, M.D., Benigni, R., Bristol, D.W., Combes, R.D., Cronin, M.T.D., Judson, P.N., Payne, M.P., Richard, A.M., Tichy, M., 1997. The development and

- validation of expert systems for predicting toxicity. *Alternatives to Laboratory Animals* 25, 223–252.
- Demuth, H., Beale, M., Hagan, M., 2008. *Neural Network Toolbox™* 6. MathWorks Inc.
- Dragon 5.0, 2007. Milano Chemometrics, Milano, Italy.
- Draize, J.H., Woodard, G., Calvery, H.O., 1944. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *Journal of Pharmacology and Experimental Therapeutics* 82, 377–390.
- Enslein, K., Borgstedt, H.H., Blake, B.W., Hart, J.B., 1987. Prediction of rabbit skin irritation severity by structure–activity relationships. *In Vitro Toxicology* 1, 129–147.
- Gerner, I., Schleder, E., 2002. Introduction of *in vitro* data into local irritation/corrosion testing strategies by means of SAR considerations: assessment of chemicals. *Toxicology Letters* 127, 169–175.
- Godavarthy, S.S., 2004. Design of improved solvents for extractive distillation. Ph. D. Dissertation, Oklahoma State University, Stillwater, OK.
- Godavarthy, S.S., Robinson, R.L., Gasem, K.A.M., 2008. Improved structure–property relationship models for prediction of critical properties. *Fluid Phase Equilibria* 264, 122–136.
- Hayashi, M., Nakamura, Y., Higashi, K., Kato, H., Kishida, F., Kaneko, H., 1999. A quantitative structure–activity relationship study of the skin irritation potential of phenols. *Toxicology in Vitro* 13, 915–922.
- Iyer, M.S., Rhinehart, R.R., 1999. A method to determine the required number of neural network training repetitions. *IEEE Transactions on Neural Networks* 10, 427–432.
- Jegadeesan, A., 2006. Structure-based generalized models for selected pure-fluid saturation properties. Master Thesis, Oklahoma State University, Stillwater, OK.
- Kodithala, K., Hopfinger, A.J., Thompson, E.D., Robinson, M.K., 2002. Prediction of skin irritation from organic chemicals using membrane-interaction QSAR analysis. *Toxicological Sciences* 66, 336–346.
- Morley, C., 2006. OpenBabelGUI 2.0.2.
- OECD, 1992. Guideline for Testing of Chemicals No. 404: Acute dermal irritation/corrosion, Paris, France. Available from: <<http://www.oecd.org>>.
- Phillips, L., Steinberg, M., Maibach, H.I., Akers, W.A., 1972. A comparison of rabbit and human skin response to certain irritants. *Toxicology and Applied Pharmacology* 21, 369–382.
- Robinson, M.K., McFadden, J.P., Basketter, D.A., 2001. Validity and ethics of the human 4-h patch test as an alternative method to assess acute skin irritation potential. *Review Article* 45, 1–12.
- Saliner, A.G., Patlewicz, G., Worth, A.P., 2008. A review of QSAR models for skin and eye irritation and corrosion. *QSAR & Combinatorial Science* 27, 49–59.
- Schultz, T.W., Cronin, M.T.D., Walker, J.D., Aptula, A.O., 2003. Quantitative structure–activity relationships (QSARs) in toxicology: a historical perspective. *Journal of Molecular Structure: theochem* 622, 1–22.
- Smith, J.S., Macina, O.T., Sussman, N.B., Karol, M.H., Maibach, H.I., 2000a. Experimental validation of a structure–activity relationship model of skin irritation by esters. *Quantitative Structure–Activity Relationship* 19, 467–474.
- Smith, J.S., Macina, O.T., Sussman, N.B., Luster, M.I., Karol, M.H., 2000b. A robust structure–activity relationship (SAR) model for esters that cause skin irritation in humans. *Toxicological Sciences* 55, 215–222.
- Tornier, C., Rosdy, M., Maibach, H.I., 2006. *In vitro* skin irritation testing on reconstituted human epidermis: reproducibility for 50 chemicals tested with two protocols. *Toxicology in Vitro* 20, 401–416.
- Welss, T., Basketter, D.A., Schröder, K.R., 2004. *In vitro* skin irritation: facts and future. State of the art review of mechanisms and models. *Toxicology in Vitro* 18, 231–243.