

2004

QSAR Study of Skin Sensitization Using Local Lymph Node Assay Data

Adam Fedorowicz
National Institute for Occupational Safety and Health

Lingyi Zheng
West Virginia University

Harshinder Singh
West Virginia University

Eugene Demchuk
West Virginia University

Follow this and additional works at: https://researchrepository.wvu.edu/faculty_publications

Digital Commons Citation

Fedorowicz, Adam; Zheng, Lingyi; Singh, Harshinder; and Demchuk, Eugene, "QSAR Study of Skin Sensitization Using Local Lymph Node Assay Data" (2004). *Faculty & Staff Scholarship*. 2887.
https://researchrepository.wvu.edu/faculty_publications/2887

This Article is brought to you for free and open access by The Research Repository @ WVU. It has been accepted for inclusion in Faculty & Staff Scholarship by an authorized administrator of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.



QSAR Study of Skin Sensitization Using Local Lymph Node Assay Data

Adam Fedorowicz,¹ Lingyi Zheng,² Harshinder Singh^{1,2} and Eugene Demchuk^{1,3}

¹ National Institute for Occupational Safety and Health, Morgantown, WV. E-mail: ajf4@cdc.gov

² Department of Statistics, West Virginia University, Morgantown, WV.

³ School of Pharmacy, West Virginia University, Morgantown, WV.

Received: 28 April 2003 / Accepted: 18 July 2003 / Published: 30 January 2004

Abstract: Allergic Contact Dermatitis (ACD) is a common work-related skin disease that often develops as a result of repetitive skin exposures to a sensitizing chemical agent. A variety of experimental tests have been suggested to assess the skin sensitization potential. We applied a method of Quantitative Structure-Activity Relationship (QSAR) to relate measured and calculated physical-chemical properties of chemical compounds to their sensitization potential. Using statistical methods, each of these properties, called molecular descriptors, was tested for its propensity to predict the sensitization potential. A few of the most informative descriptors were subsequently selected to build a model of skin sensitization. In this work sensitization data for the murine Local Lymph Node Assay (LLNA) were used. In principle, LLNA provides a standardized continuous scale suitable for quantitative assessment of skin sensitization. However, at present many LLNA results are still reported on a dichotomous scale, which is consistent with the scale of guinea pig tests, which were widely used in past years. Therefore, in this study only a dichotomous version of the LLNA data was used. To the statistical end, we relied on the logistic regression approach. This approach provides a statistical tool for investigating and predicting skin sensitization that is expressed only in categorical terms of activity and non-activity. Based on the data of compounds used in this study, our results suggest a QSAR model of ACD that is based on the following descriptors: nDB (number of double bonds), C-003 (number of *CHR3* molecular subfragments), GATS6M (autocorrelation coefficient) and HATS6m (GETAWAY descriptor), although the relevance of the identified descriptors to the continuous ACD QSAR has yet to be shown. The proposed QSAR model gives a percentage of positively predicted responses of 83% on the training set of compounds, and in cross validation it correctly identifies 79% of responses.

Keywords: ACD, LLNA, binary QSAR, logistic regression, skin sensitization.

Introduction

The Bureau of Labor Statistics estimates that occupational skin diseases constitute the second largest group of occupational injuries in the U.S. [1]. Among them, Occupational Contact Dermatitis (OCD) is the most common cause of work-related skin illness comprising up to 95% of registered cases. Allergic Contact Dermatitis (ACD) may lead to severe recurrent forms of OCD because of long-lasting memory of the immune system. ACD, which is an adaptive, T-cell mediated immune response [2], usually develops as a result of repetitive skin exposures to a sensitizing chemical agent. At least a single excessive exposure is essential in the development of the immune response. Information that leads to the development of recommended skin exposure limits that would prevent workers from sensitizing overexposures is an important factor impacting public health. A variety of experimental tests have been suggested to assess the skin sensitization potential of a chemical [3]. Unfortunately, many experimental protocols result in a dichotomous conclusion, more appropriate for denial/acceptance decision-making in design and manufacturing of new chemicals rather than for preventive protection of workers occupationally involved with sensitizing chemical agents. The murine Local Lymph Node Assay (LLNA) has the capacity to provide dose response data that can be used as a standardized continuous scale in the quantitative assessment of skin sensitization.

A combination of methods in statistics and computational chemistry, commonly referred to as Quantitative Structure-Activity Relationship (QSAR) modeling, complements the experimental approach. A method of QSAR is based on the examination of measured and calculated molecular descriptors, with known biological activity, in this work the sensitization potential, and then relating a few of the most informative descriptors to the target bioactivity. The structure-activity relationships constructed this way provide a means of investigating and predicting the sensitization potential of the chemicals.

We rely on LLNA data to quantify the skin sensitization potential [4]. At present, the LLNA data are (1) outnumbered by the long history of guinea pig assays, and (2) often reported as dichotomous and congruous to the guinea pig data. Therefore, the work has been started using LLNA data in a dichotomous format to identify molecular descriptors that may be effective in the continuous-scale LLNA QSAR. The work began from building a database of chemical names, structures, properties and bioactivities, along with the design of appropriate software. Our immediate goal is to identify a pool of potentially informative molecular descriptor classes that are most appropriate for QSAR modeling to predict skin sensitization potential. In the present work, a QSAR based on a logistic regression is proposed. The logistic regression permits construction of standard QSAR equations, in which the activity data are represented only in terms of activity (1) or non-activity (0) values. In order to evaluate molecular properties, which can be associated with LLNA data on skin sensitization, 1204 molecular descriptors were calculated and tested for their significance in predicting the skin sensitization potential. Only a limited number of molecular descriptors were found to be statistically associated with skin sensitization.

Materials and Methods

In the present study, a pool of 54 LLNA-tested compounds was used, of which 25 were sensitizers and 29 were negative controls [5, 6]. The molecular structures of these compounds were first encoded using the SMILES notation and subsequently transformed into three-dimensional co-ordinates using *Cerius²* from Accelrys, Inc (Accelrys, San Diego, USA, <http://www.accelrys.com/cerius2>). The *Dragon 2.1* software developed by Milano Chemometrics and QSAR Research Group was used to calculate a total of 1204 molecular descriptors (<http://www.disat.unimib.it/chm/Dragon.htm>), for each of the studied compounds. The statistical analysis was carried out using the *SAS 8.2* statistical package [7].

The linear probability model is inadequate for modeling the probability of positive LLNA sensitization response, since it is heteroscedastic and often leads to uninterpretable results. The logistic regression is a more appropriate statistical tool than linear probability models, when the response variable is *binary* (dichotomous). The properties of the logistic function ensure that whatever estimate of the response one obtains, it is always a number between 0 and 1 that can be easily translated into a binary response using an appropriate threshold value (usually 0.5). The S-shape of the logistic function is another important feature, which is particularly appealing in epidemiology studies when a single variable X is viewed as representing an index that combines contributions of several risk factors and $\pi(\mathbf{X})$ represents the risk for a given value of X in single variable logistic regression models. Depending on the choice of cumulative distribution function F , the probability of positive response of the LLNA sensitization test $P\{S=1|X_1, X_2, \dots, X_N\} = F(\mathbf{X}\beta)$ – can be represented either by the probit or the logistic regression model [8]. In the present study, we used the logistic regression model, where $\pi(\mathbf{X}) = P\{S=1|X_1, X_2, \dots, X_N\}$ that depends on molecular descriptors X_1, X_2, \dots, X_N , is modeled in the form

$$\pi(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_N X_N}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_N X_N}}$$

or

$$\log \left(\frac{\pi(\mathbf{X})}{1 - \pi(\mathbf{X})} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_N X_N,$$

where $\beta_0, \beta_1, \dots, \beta_N$ are regression coefficients.

The validity of logistic regression models was tested using cross validation, which, in general, treats $n-1$ out of n training observations as a training set [9]. It re-estimates the parameters of the model, and then classifies the remaining n -th observation based on the new parameter estimates. This is repeated for each of the n training observations. The misclassification rate for each group is the proportion of sample observations in the group that are misclassified. This method achieves an almost unbiased estimate but with a relatively large variance.

The most predictive molecular descriptors were identified in several stages. At first, the statistical

quality of a single-descriptor logistic model, the P-value, was assessed for each of the descriptors. Descriptors with the P-value above 0.05 were then omitted from the further analysis. The remaining potentially predictive descriptors were subsequently used in an exhaustive search through all possible combinations of 1, 2, 3 and 4-descriptor models, along with a stepwise regression algorithm, which does not restrict the number of descriptors in the model. However, the total number of descriptors was limited to four, following a commonly used QSAR ‘rule of thumb’, which sets the lower limit of about 15 molecules per one fitted parameter in the model. QSAR models which identified positive sensitizers with probability above 75% were analyzed in detail. The validity of these results was additionally verified using cross validation.

Results and Discussion

Overall 420 descriptors (out of 1204) were found to be statically significant at the P-level of 0.05. Table 1 shows the top part of the list of descriptors with P-values below the 0.01 threshold.

The selection of the classes of molecular descriptors with P-value below 0.01 is hypothesized to have an association with immunological activity measured by Local Lymph Node Assay, where the three dimensional structure recognition of a given antigen is responsible for the immunological response. Most of these descriptors are either: 1) radial distribution functions (RDF); 2) topological properties; 3) GETAWAY descriptors or 4) BCUT descriptors. Descriptors that belong to the class of radial distribution function descriptors [10] are based on the distance distribution in the geometrical representation of the molecule. In addition to interatomic distances in the entire molecule, the RDF also provides valuable information about bond distances, ring types, planar and non-planar systems, atom types and other important structural motifs. By using different weighting schemes, which include atom types, electronegativity, atom mass or van der Waals radii, RDF can be adjusted to select among those atoms of molecule, which give rise to an important descriptor in deriving an appropriate QSAR. The Topological descriptors are based on molecular graphs as a source of probability distributions to which the information theory definitions apply [11]. They can be considered as a quantitative measure expressing the lack of structural homogeneity or the diversity of the graph, and in this way they are related to the symmetry associated with structure. The GETAWAY class of descriptors represents recently proposed [GEometry, Topology and Atom-Weights Assembly] group of descriptors, which are based on a leverage matrix similar to that defined in statistics and usually used for regression diagnostics. These molecular descriptors match the three dimensional molecular geometry provided by the molecular influence matrix and atom relatedness by molecular topology, with chemical information by using various atomic weight schemes [12,13]. Therefore, this class of descriptors is highly sensitive to the 3-dimensional molecular structure. Combined with appropriate weighting schemes the GETAWAY descriptors are used to compare molecules or even conformers taking into account their molecular shape, size symmetry and atom distribution, which are ‘scaled’ using specific atomic property.

Table 1. Descriptors.

No.	Symbol	Definition	Class of Descriptors	P-Value
1	C-003	CHR3	Atom-centered fragments	0.0005
2	RDF040p	Radial Distribution Function -4.0 / weighted by atomic polarizabilities	RDF	0.0024
3	nDB	Number of double bonds	Constitutional	0.0029
4	RDF040v	Radial Distribution Function -4.0 / weighted by atomic van der Waals volumes	RDF	0.0039
5	TI2	Second Mohar index TI2	Topological	0.0040
6	GATS6m	Geary autocorrelation – lag 6 / weighted by atomic masses	2D autocorrelations	0.0042
7	Rtu+	R maximal index / unweighted	GETAWAY	0.0045
8	RTe+	R maximal index / weighted by Sanderson electronegativities	GETAWAY	0.0049
9	BEHp2	Highest eigenvalue n. 2 of Burden matrix / weighted by atomic polarizabilities	BCUT	0.0051
10	RDF050e	Radial Distribution Function -5.0 / weighted by atomic Sanderson electronegativities	RDF	0.0061
11	X3v	Valence connectivity index chi-3	Topological	0.0061
12	S2K	2-path Kier alpha-modified shape index	Topological	0.0070
13	RDF065p	Radial Distribution Function -6.5 / weighted by atomic polarizabilities	RDF	0.0072
14	X1v	Valence connectivity index chi-1	Topological	0.0074
15	E2m	2nd component accessibility directional WHIM index / weighted by atomic masses	WHIM	0.0078
16	Htp	H total index / weighted by atomic polarizabilities	GETAWAY	0.0082
17	RDF075p	Radial Distribution Function -7.5 / weighted by atomic polarizabilities	RDF	0.0082
18	X0v	Valence connectivity index chi-0	Topological	0.0085
19	RDF075v	Radial Distribution Function -7.5 / weighted by atomic van der Waals volumes	RDF	0.0089
20	RDF065u	Radial Distribution Function -6.5 / unweighted	RDF	0.0092
21	RDF050u	Radial Distribution Function -5.0 / unweighted	RDF	0.0095
22	BEHe2	Highest eigenvalue n. 2 of Burden matrix / weighted by Sanderson electronegativities	BCUT	0.0097

The latter include electronegativity, atom mass, and van der Waals radii. BCUT is a class of molecular descriptors defined as eigenvalues of the modified connectivity matrix, which is also called the Burden matrix B [11]. These descriptors have been demonstrated to reflect relevant aspects of molecular structure, and are therefore useful in similarity searching and comparison [11, 14]. The next group of descriptors is based on 2-dimensional autocorrelation functions applied to a molecular graph, which is a 2-dimensional structural representation of a molecule. This class of descriptors expresses a correlation between numerical values of the graph entries, which can be statistically weighted using various atomic properties, at intervals equal to the given lag value [11]. WHIM descriptors are the molecular descriptors based on statistical indices calculated on the projections of the atoms along principal axes [11, 15]. They are built in such a way as to capture relevant molecular 3-dimensional information regarding molecular size, shape, symmetry, and atom distribution with respect to invariant reference frames.

The fact that these classes of descriptors are derived either from three- (radial distribution function, GETAWAY and WHIM) or two- (2-D autocorrelation function, topological and BCUT) dimensional representation of a molecule, seems to indicate a connection between the molecular structure of sensitizing chemical and its skin sensitization potential. This would be consistent with the highly stereoselective and specific requirements for immunological responses to larger proteins. These data suggest that for low molecular weight chemicals, the expression of explicit molecular patterns and motifs may be necessary to invoke a reaction from the immune system. These molecular patterns can be expressed in terms of 2- and 3-dimensional molecular descriptors that after appropriate validation can be used to construct QSAR models of skin sensitization.

Even descriptors that at the first look seem not to be related to the 3D molecular structure, like the number of double bonds or the number of CHR_3 groups, in fact, do define molecular sub-fragments that can be considered as 'structure making' factors. For example, the number of double bonds between two carbon atoms is associated with the cis-trans isomerism or may indicate the presence of an aromatic ring. The number of double bonds might also be associated with the hydrophobicity and reactivity of the studied compounds. Another important structural element, which contains a double bond, is the carbonyl $\text{C}=\text{O}$ group. The C-003 descriptor, which is a counter of the CHR_3 groups or strictly speaking tertiary carbon atoms, also points at structural motifs that seem to be important in determination of the molecular shape, which is particularly important in the study of skin sensitization.

Sophisticated representation of all but two identified descriptor classes impedes a simple interpretation of the mechanism of immunological response in skin sensitization. Therefore, in this study we rely on QSAR modeling only as an instrument of predicting the immunological activity. Several tested QSAR models showed interesting results. We found that the best classification results were achieved with the 3-, 4-parameter models, although we have identified several above-average models that include only 2 or even 1 descriptor (Table 2). The differences in classification between the best models were minimal, which seems to suggest that future QSAR studies of skin sensitization, based

Table 2. Comparison of the best performing logistic models containing 1, 2, 3 and 4 descriptors. Most of presented descriptors are described in Table 1 or in the text, apart from: BELv2, which is a BCUT descriptor weighted by atomic van der Waals volumes; Mor13m, which is a 3D-Morse descriptor weighted by atomic masses; TIE is E-state topological parameter, and C-002 is a counter of CH_2R_2 molecular sub-fragments.

Model	Percentage of correctly predicted responses	
	Cross validation	Model
GATS6m	59%	65%
BELv2, Mor13m	69%	76%
nDB, C-003, GATS6m	76%	78%
nDB, TIE, C-003	70%	74%
E2m, TI2, C-003	74%	78%
E2m, RTe+, C-003	70%	72%
nDB, C-003, E2m, C-002	79%	80%
nDB, GATS6m, HATS6e, C-003	79%	83%
nDB, RTe+, E2m, C-003	78%	78%
nDB, C-003, GATS6m, TIE	78%	80%

on Local Lymph Node data, may yield different models with exceptional validity. These models may provide additional information about molecular factors, which are important for skin sensitization. The best model that we have identified so far consists of 4 descriptors:

$$\log\left(\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})}\right) = -1.63 + 1.46 \cdot [nDB] + 183.9 \cdot [GATS6m] - 5.97 \cdot [HATS6e] + 3.02 \cdot [C-003]$$

where:

- i. nDB is the number of double bonds.
- ii. GATS6m is the mass-weighted Geary graph spatial autocorrelation coefficient of the sixth lag. The Geary coefficient is a distance-type function varying from zero to infinity. Strong autocorrelation produces low values of this index; moreover, positive autocorrelation translates into values between 0 and 1 whereas negative autocorrelation produces values larger than 1.
- iii. HATS6e is the GETAWAY descriptor weighted by the atomic Sanderson electronegativities. This descriptor encodes information about molecular shape, size, and atom distribution. Application of the Sanderson electronegativities as weighting coefficients, takes into account, to some degree, charge distribution inside a molecule.
- iv. C-003 is the atom-centered fragments descriptor, indicating the presence of the CHR_3 molecular sub-fragment.

As mentioned above, the choice of these descriptor classes, and particularly these four molecular descriptors, indicates a plausible connection between the proposed QSAR model of skin sensitization and molecular stereospecificity of the immunological response, where the 3-dimensional information about a sensitizing agent is the most critical component of the receptor-ligand interaction [16]. These four descriptors show that the presence of specific molecular motifs, like double bonds or tertiary carbon atoms or molecular patterns modeled by descriptors GATS6m and HATS6e, is an important factor in predicting the skin sensitization potential of a chemical.

The proposed QSAR model gives a percentage of positively predicted responses of 83% on the training set of compounds, and in cross validation it correctly identifies 79% of responses. The results of proposed QSAR model are summarized in Table 3.

Table 4 presents the list of compounds tested in this study, together with corresponding Local Lymph Node Activity data and the activity estimated by the application of the proposed QSAR model.

Table 3. Model Summary.

	Percentage of correctly predicted responses	Percentage of correctly identified active compounds	Percentage of correctly identified inactive compounds
Model	83%	72%	93%
Cross validation	79%	68%	90%

Table 4. LLNA-tested compounds.

No.	Compound	CAS	LLNA	Predicted skin sensitization
1	chlorobenzene	108-90-7	0	0
2	geraniol	106-24-1	0	1
3	phenol	108-95-2	0	0
4	2-chloroethanol	107-07-3	0	0
5	benzaldehyde	100-52-7	0	1
6	1-bromobutane	109-65-9	0	0
7	1-butanol	71-36-3	0	0
8	2-4-dichloronitrobenzene	611-06-3	0	0
9	isopropanol	67-63-0	0	0
10	glycerol	56-81-5	0	0
11	hexane	110-54-3	0	0
12	streptozotocin	18883-66-4	0	0
13	4-aminobenzoic acid	150-13-0	0	0
14	2-acetamidofluorene	53-96-3	0	0
15	benzalkonium chloride	8001-54-5	0	0

Table 4 (continued).

No.	Compound	CAS	LLNA	Predicted skin sensitization
16	dimethyl-isophthalate	1459-93-4	0	0
17	ethyl-methanesulfonate	62-50-0	0	0
18	4-hydroxybenzoic acid	99-96-7	0	0
19	lactic acid	598-82-3	0	0
20	4-methoxyacetophenone	100-06-1	0	0
21	6-Methylcoumarin	92-48-8	0	0
22	methyl-4-hydroxybenzoate	99-76-3	0	0
23	methyl salicylate	119-36-8	0	0
24	2-nitrofluorene	607-57-8	0	0
25	propylene glycol	57-55-6	0	0
26	propyl paraben	94-13-3	0	0
27	resorcinol	108-46-3	0	0
28	salicylic acid	69-72-7	0	0
29	di-2-furanylethanedione	492-94-4	0	0
30	12-bromo-1-dodecanol	3344-77-2	1	1
31	3-amino-5-mercapto-1,2,4-triazole	16691-43-3	1	0
32	chloramine-T	127-65-1	1	1
33	benzocaine	94-09-7	1	0
34	urushiol V	53237-59-5	1	1
35	2-aminophenol	95-55-6	1	0
36	phthalic anhydride	85-44-9	1	1
37	cinnamic aldehyde	104-55-2	1	1
38	camphorquinone	10373-78-1	1	1
39	2-hydroxyethyl-acrylate	818-61-1	1	1
40	N-nitroso-N-methylurea	684-93-5	1	1
41	diethyl-sulfate	64-67-5	1	1
42	1-2-Benzisothiazol-3[2H]-one	2634-33-5	1	1
43	butyl-glycidil ether	2426-08-6	1	0
44	methyl-2-nonynoate	111-80-8	1	1
45	2-vinylpyridine	100-69-6	1	1
46	propyl gallate	121-79-9	1	0
47	ethylene-glycol-dimethacrylate	97-90-5	1	0
48	imidazolidinyl urea	39236-46-9	1	1
49	tetrachlorosalicylanilide	1154-59-2	1	0
50	oxazolone	1564-29-0	1	1
51	acetyl-isovaleryl	13706-86-0	1	1
52	hydroxycitronellal	107-75-5	1	1
53	methylene diphenyl diisocyanate	101-68-8	1	1
54	dodecyl methanesulphonate	51323-71-8	1	1

Conclusions

The main goal of the present study was to evaluate classes of molecular descriptors that later can be used in a comprehensive QSAR model of contact sensitization based on a larger set of compounds tested in LLNA. Our current results demonstrate that the most promising molecular descriptors are derived either from three or two dimensional molecular structure indices, which are based on radial distribution functions, or topological indices, or autocorrelation functions. These classes of descriptors seem to be naturally related to the sensitizing activity as they associate the immunological response with a three dimensional structure and shape of the sensitizing agents. These results suggest that it is possible, by using only a few appropriate parameters, to build comprehensive QSAR models of ACD. However, the relevance of the identified descriptors to the continuous-scale ACD QSAR has yet to be shown. Further work will be focused on populating the QSAR database with continuous-scale ACD data and an expansion of the database. New predictive QSARs are expected to be useful in screening larger sets of compounds for their potential impact on the skin, and thus may suggest a useful order of priorities in experimental testing.

Acknowledgment

This research was supported by the National Occupational Research Agenda (NORA) Dermal Exposure Research Program.

References

1. Worker Health Chartbook, 2000. Nonfatal Illness. DHHS (NIOSH) Publication No. 2002-120, April 2002.
2. Engelhard, V. H. How cells process antigens. *Sci. Am.* **1994**, *8*, 44-51.
3. Hewitt, P.; Maibach, H. I. *Dermatotoxicology*. In: *Handbook of Occupational Dermatology*; Kanerva, L.; Eisner, P.; Wahiberg, J. E.; Maibach, H.I.; Eds., Springer: Berlin, 2000.
4. *The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds*, NIH Publication No. 99-4494, February 1999.
5. Ashby, J.; Basketter, D. A.; Paton, D.; Kimber, I. Structure-activity relationships in skin sensitization using murine local lymph node assay. *Toxicology* **1995**, *102*, 177-194.
6. Haneke, K. E.; Tice, R. R.; Carson, B. L.; Margolin, B. H.; Stokes, W. S. ICCVAM evaluation of the murine local lymph node assay. III. Data analyses completed by the national toxicology program interagency center for the evaluation of alternative toxicological methods. *Regul. Toxicol. Pharm.* **2001**, *34*, 274-286
7. SAS Institute., *SAS/STAT User's Guide*, Version 8, Cary, NC, SAS Institute Inc., 1999, p. 1901.
8. Agresti, A. *Categorical Data Analysis*, John Wiley & Sons, New York, 1990, pp 79-129.
9. Hawkins, D. M., Basak, S. C., Mills D. Assessing model fit by cross-validation. *J. Chem. Inf. Comp. Sci.* **2003**, *43*, 579-586.

10. Hemmer, M. C.; Steinhauer, V.; Gasteiger J. J. *Vib. Spectrosc.* **1999**, *19*, 151-164.
11. Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*. Wiley-VCH: Weinheim, Germany, 2000.
12. Consonni, V.; Todeschini R.; Pavan M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. *J. Chem. Inf. Comp. Sci.* **2002**, *42*, 682-692.
13. Consonni, V.; Todeschini, R.; Pavan M.; Gramatica, P. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *J. Chem. Inf. Comp. Sci.* **2002**, *42*, 693-705.
14. Pearlman, R.S.; Smith, K.M. Metric validation and the receptor-relevant subspace concept. *J. Chem. Inf. Comp. Sci.* **1999**, *39*, 28-35.
15. Todeschini, R.; Gramatica, P. 3D-modelling and prediction by WHIM descriptors. Part 5. Theory development and chemical meaning of WHIM descriptors. *Quant. Struct.-Act. Rel.* **1997**, *16*, 113-119.
16. Hansson, C., Thörneby-Anderson, K. Stereochemical considerations on concomitant allergic contact dermatitis to ester of the *cis-trans* isomeric compounds maleic acid and fumaric acid. *Skin Pharmacol. Appl.* **2003**, *16*, 117-122.