Quantum Mechanical Structure−Activity Relationship Analyses for Skin Sensitization

Matthew D. Miller,[†] David M. Yourtee,[‡] Alan G. Glaros,[§] Cecil C. Chappelow,[||] J. David Eick,[§] and Andrew J. Holder*,[†]

Department of Chemistry, University of Missouri−Kansas City, Kansas City, Missouri 64110, Schools of Pharmacy and Medicine, University of Missouri−Kansas City, Kansas City, Missouri 64110, School of Dentistry, University of Missouri−Kansas City, Kansas City, Missouri 64110, and Midwest Research Institute, Kansas City, Missouri 64110

Received January 18, 2005

Allergic contact dermatitis (ACD) results in inflammation of the skin due to sensitization of the immunologic system to a particular substance. The sensitization process is limited by the compound's ability to both permeate and react with proteins in the integumentary system. Currently, only in vivo animal tests such as the local lymph node assay (LLNA) are recognized by regulatory authorities for risk assessment of ACD. A quantitative structure−activity relationship has been developed to predict relative potency, which allows for the prediction of relative sensitization potentials. The experimental values used in this study include EC3 values (the concentration at which the stimulation index equals 3) from LLNA tests. The predictions in this model enable categorization of the compounds into three groups on the basis of risk of sensitization and enable screening of candidate molecules using rapid SAM1 semiempirical calculations prior to animal testing. The model may also be used to reduce the number of animals subjected to testing by providing estimated concentrations required for useful data of risk assessment. The effect of averaging available literature values on predictive ability is also investigated. The model includes halogenated compounds, aromatic compounds, alcohols, aldehydes, and ketones. The computational investigation resulted in a two-descriptor model that is consistent with the assumed mechanism for sensitization.

INTRODUCTION

Allergic contact dermatitis (ACD) is the most common of all human allergic reactions.[1] This immunological response results in substantial deleterious effects such as erythema and edema. Currently, more than 3500 chemical substances are known to elicit ACD.[2] The development of contact dermatitis is directly related to the duration and extent of exposure as well as the potency of the allergen,[3] and many investigators have reviewed its process.[1,3−6]

ACD occurs in two phases, the first of which is the induction phase. Upon exposure to the sensitizing agent, or hapten, the compound must cross the *stratum corneum*. Several properties influence the ability of the hapten to penetrate this barrier, including hydrophobicity (estimated by the octanol/water partition coefficient, log *P*), the size and shape of the compound, and the presence of highly charged sites in the molecule. Once the chemical has penetrated the skin barrier, it must be present in sufficient quantity and capable of partitioning between relevant cellular regions for adequate bioavailability[6] to produce a response. The hapten must then covalently coordinate with a protein present in the cell. Further background information pertaining to local lymph node assay (LLNA) research,[1,7,8] test procedures,[9] and comparisons with other sensitization test methods[3,10−13] is available in the literature.

The LLNA measures an experimental isotope reading at a specified (w/v) percent concentration. A control reading with a concentration of 0 is tested concurrently, and a stimulation index (SI) is derived relative to this control. A SI value of 3 is a generally accepted value to classify a compound as a sensitizer.[10,14] Thus, the concentration of the compound at which the test animal became sensitized is an obvious metric to determine the relative potency of the substance. We observe that this value can be linearly determined using a simple mathematical expression with the following definition of variables: x_1 and x_2 are measured concentrations above and below SI = 3; y_1 and y_2 are calculated SI values for respective concentrations; y_3 is the calculated concentration for SI = 3. The fundamental eq 1, with *B* representing the *y* intercept, can be modified to eq 2 with the point at which SI > 3. Eq 3 is a rearranged version of eq 2 solved for unknown x_3. Substituting y_3 with the value 3 and eq 2 for the *y* intercept yields eq 4. Applying eq 5 into eq 4 results in the estimated concentration needed to reach a SI of 3 (eq 6). For example, using reported data on diethyl maleate,[15] 1% (SI = 2.1) and 2.5% (SI = 3.3), in this linear extrapolation results in a calculated concentration of 2.1% for SI = 3.

* To whom correspondence should be addressed. E-mail: holdera@umkc.edu.

[†] Department of Chemistry, University of Missouri−Kansas City.

[‡] Schools of Pharmacy and Medicine, University of Missouri−Kansas City.

[§] School of Dentistry, University of Missouri−Kansas City.

[||] Midwest Research Institute.

10.1021/ci050018z CCC: $30.25
Published on Web 05/10/2005

$$y = mx + B \quad (1)$$

$$B = y_1 - mx_1 \quad (2)$$

$$x_3 = \frac{y_3 - B}{m} = \frac{y_3}{m} - \frac{B}{m} \quad (3)$$

$$x_3 = \frac{3}{m} - \frac{y_1 - mx_1}{m} = \frac{3 - y_1}{m} + x_1 \quad (4)$$

$$m = \frac{y_2 - y_1}{x_2 - x_1} \quad (5)$$

$$x_3 = \frac{3 - y_1}{y_2 - y_1}(x_2 - x_1) + x_1 \quad (6)$$

The calculated or *estimated concentration* at which SI = 3 (EC3) has become a common method to report LLNA data in the literature rather than reporting the experimental concentrations and isotope readings. Basketter and co-workers[15] compared linear extrapolation with more complex methods (quadratic regression and Richard's model) to determine the best technique. The study concluded that there was a high correlation between the more complex methods and simple linear interpolation, and as such, the simpler linear method adequately represented the extrapolation of EC3.

Additionally, Basketter and co-workers[15] suggested that it is unreasonable to assume that small differences in EC3 values should be interpreted as real variance in sensitization potentials. Other studies have reached similar conclusions concerning the lack of significance in small EC3 variations for biological significance, as it is not uncommon for biological experiments to have some unexplainable variation.[1,16,17] For example, the reproducibility of isoeugenol is quite acceptable with a range of 0.5−2.6 for 29 separate assays over a 4 year period. For this reason, classification of sensitization potency is usually based on ranges to separate strong, moderate, and weak sensitization potentials. One of the more commonly used ranges is defined by separations based on orders of magnitude, such as <0.1% (extreme), 0.1−1% (strong), 1−10% (moderate), 10−100% (weak), and >100% (nonsensitizing).[1] There is still debate about the most appropriate classification ranges,[1,3] and the ranges from the study were selected for simplicity. Other concentration ranges for EC3% values also in use include <0.1, <6.5, <30, >30;[18] <0.1, <10, <30, >30;[19] <0.1, <6, <20, >20;[20] and <10, <30, >30.[21]

In this study, the skin sensitization potentials for a variety of molecules available in the literature reported from LLNA experiments were used to develop a predictive quantitative structure−activity relationship (QSAR) model based on geometric, topological, electrostatic, constitutional, thermodynamic, and quantum chemical descriptors. This is the first QSAR based on quantum mechanical descriptors to predict sensitization potentials utilizing the LLNA for experimental data. It was decided in early stages of development that the QSAR should cover a broad class of compounds for predictive screening. Testing the model on an external set allowed evaluation of the ability of the model to successfully predict sensitization potency.

It was determined that the use of EC3 data would be the most efficient method for determining sensitization potentials. Although conversion of mass concentrations to molar concentrations was considered, the success of this model allowed the omission of this conversion process for direct prediction of sensitization potentials in a standard reportable format.

METHODS

Model Development. Calculations were performed on sensitizing agent molecules using the SAM1[22] semiempirical method implemented in the AMPAC 8.0 with Graphical User Interface software.[23] A conformational analysis was performed to identify the lowest energy conformation for use in the model. The descriptors were calculated using Codessa[24] from the AMPAC output files. Several hundred possible descriptors result from Codessa and may be categorized as follows: constitutional (number of atoms, bonds, and rings); topological (atomic connectivity indices); geometric (molecular volume and surface area); electrostatic (charge distribution); quantum chemical (orbital energies, charge distribution based on quantum chemical calculations); and thermodynamic (enthalpies and entropies). A heuristic algorithm was used to derive several correlations from these descriptors. The best model was selected based on R^2, adjusted R^2, cross-validated R^2_{CV} (leave-one-out method), F test, t-test, and chemical sense of the descriptors for the understood mechanism of sensitization. Statistical analyses were performed using SPSS 11.5 (SPSS Inc.). High intercorrelation between descriptors (VIF) and p significance were also considered.

Test Molecules. A total of 87 structures with reported literature values were divided into a training set (65 structures) and a validation set (22 structures). The compounds were separated into the two sets based on similar functionalities (aromaticity, halogen type, Michael reactant, and aldehyde/ketone) with random selection for each category. An attempt was also made to maintain an even distribution of EC3 values for the training and test sets.

A total of 67 compounds (Table 1) remained for the training set (n = 50) and validation set (n =17) following the removal of 20 outlier compounds. The basis for removal of the 20 compounds was a poor fit between calculated and experimental EC3 values. These compounds were later retested in the model (Table 2). Several possible causes exist that lead to an inaccurate prediction of certain functionalities. Compounds **1**−**3** have unique functionalities that may require metabolic activation prior to exhibiting the sensitization effect. In such cases, the use of a modified structure may enable better prediction. However, one benefit of the current model is its simplicity. Consideration of activated forms of the compounds adds complexity to the model and shall be investigated in a future study. Iodoalkane compounds (**4**−**7**), although properly categorized, resulted in calculated EC3 values exceeding 90 in each case, with large error. Inadequate representation of large EC3 values (**10**−**14)** and functionalities (**8**−**11**, **15**, and **16**) in the training set also limit the use of this model. These types of molecules as well as other unrepresented compounds may not be predicted well or may at least require caution and interpretation. The final four outliers listed simply did not fit in the model (compound **17** was properly categorized with large error). The original source for the azlactone EC3 values (**18**−**20**) indicated

Table 1. Data Used in the Model, Including Calculated and Experimental EC3 Values, Corresponding Classifications, and Calculated Descriptor Values

$EC3_{calc}$	$EC3_{lit}$	compound	$class_{calc}$	$class_{lit}$[a]	HOMO	ESP
		Training Set				
−2.45	0.1[21]	hydroquinone	strong	strong	8.83	0.60
−0.83	2.0[25]	cinnamal	strong	strong	8.65	0.87
−0.16	2.2[21]	2-amino-6-chloro-4-nitro phenol	strong	strong	7.64	1.42
−0.08	0.002[15]	oxazolone	strong	strong	9.71	0.45
0.92	1.3[21]	naphthol	strong	strong	8.23	1.26
2.58	0.78[26]	metol	strong	strong	8.43	1.34
3.27	0.05[26]	*N*-methyl-*N*-nitroso urea	strong	strong	10.00	0.68
3.52	4.7[25]	phenyl acetaldehyde	strong	strong	9.89	0.76
3.66	5.5[21]	*trans*-2-hexenal	strong	**moderate**	10.08	0.69
3.98	3.5[21]	isoeugenol	strong	strong	8.28	1.57
4.08	2.1[15]	diethyl maleate	strong	strong	9.95	0.80
4.10	0.5[21]	2-amino phenol	strong	strong	8.71	1.38
4.16	0.4[25]	formaldehyde	strong	strong	11.44	0.11
5.02	1.3[15]	phenyl benzoate	strong	strong	8.70	1.48
5.26	0.85[27]	benzylidene acetone	strong	strong	8.73	1.50
5.29	0.2[25]	glutaraldehyde	strong	strong	10.83	0.51
5.34	2.5[21]	*trans*-2-decenal	strong	strong	10.01	0.91
5.56	5.8[21]	2-methoxy-4-methyl phenol	strong	**moderate**	8.96	1.42
6.52	8.8[26]	hexadecanoyl chloride	moderate	moderate	10.37	0.87
7.12	4.5[21]	methyl cinnamic aldehyde	moderate	**strong**	8.99	1.58
7.15	12.9[15]	eugenol	moderate	moderate	8.97	1.59
7.33	10[21]	camphorquinone	moderate	moderate	8.95	1.62
7.82	13.2[21]	5-methyl eugenol	moderate	moderate	8.85	1.72
8.21	6.5[21]	4-chloro aniline	moderate	moderate	8.74	1.81
8.31	7.95[21]	1-(−)-perillaldehyde	moderate	moderate	9.82	1.32
9.38	12.05[21]	pentyl cinnamic aldehyde	moderate	moderate	8.98	1.83
9.52	10.2[21]	bromo tridecane	moderate	moderate	10.90	0.95
9.55	5.2[21]	bromo pentadecane	moderate	moderate	10.80	1.00
9.90	8[28]	hexyl cinnamic aldehyde	moderate	moderate	8.98	1.89
10.04	15.5[21]	bromo dodecane	moderate	moderate	10.96	0.97
10.35	16.6[21]	bromo octadecane	moderate	**weak**	10.68	1.14
10.73	9.2[21]	bromo tetradecane	moderate	moderate	10.84	1.10
10.95	10.3[21]	bromo hexane	moderate	moderate	11.49	0.82
11.57	14.3[21]	lactic acid	moderate	moderate	11.86	0.72
11.76	13[25]	citral	moderate	moderate	9.47	1.86
11.92	10[21]	2-methylundecanal	moderate	moderate	11.05	1.14
11.95	7.5[21]	safranal	moderate	moderate	8.55	2.31
12.84	9.1[21]	chloro hexadecane	moderate	moderate	11.79	0.89
12.86	20.2[21]	chloro tetradecane	moderate	**weak**	11.89	0.85
13.11	16.9[21]	6-methyl eugenol	weak	weak	8.94	2.25
15.82	11.7[21]	farnesal	weak	**moderate**	9.16	2.45
18.25	16.1[21]	C11-azlactone	weak	weak	10.81	1.94
18.48	22.3[19]	cyclamen aldehyde	weak	weak	9.73	2.47
18.98	19[21]	C17-azlactone	weak	weak	10.79	2.03
18.98	26.4[21]	C19-azlactone	weak	weak	10.79	2.03
20.52	17.1[21]	lyral	weak	weak	10.36	2.40
20.56	20[25]	hydroxycitronellal	weak	weak	11.00	2.10
23.11	17[27]	benzyl benzoate	weak	weak	11.85	1.98
23.73	19[25]	*p-tert*-butyl cinnamic aldehyde	weak	weak	9.72	3.05
23.77	30[21]	linalool	weak	weak	10.49	2.69
		Validation Set				
−3.76	0.6[21]	glyoxal	strong	strong	9.46	0.28
−1.54	4[21]	2,4-heptadienal	strong	strong	8.88	0.87
−1.13	0.05[25]	diphencyclopropenone	strong	strong	7.78	1.03
0.81	1.1[26]	*N*-ethyl-*N*-nitroso urea	strong	strong	10.00	1.10
2.20	2.8[26]	3,4-dihydro coumarin	strong	strong	9.40	1.05
3.67	0.02[26]	2,4-dinitro chlorobenzene	strong	strong	8.67	1.03
4.13	2.2[21]	HC red no. 3	strong	strong	6.82	1.53
6.55	6.3[21]	2-phenyl propional	moderate	moderate	9.89	1.03
6.63	13.7[27]	4-methyl hydrocinnamic aldehyde	moderate	moderate	9.65	1.24
7.12	3.5[21]	3-methyl isoeugenol	moderate	**strong**	8.88	1.59
8.07	3.2[26]	3-amino phenol	moderate	**strong**	9.14	0.90
8.44	13.7[21]	butyl cinnamic aldehyde	moderate	moderate	8.98	1.65
9.36	11[27]	2-phenyl methylene heptanal	moderate	moderate	8.98	1.86
12.59	16.3[21]	chloro octadecane	moderate	**weak**	11.72	3.62
14.08	21.3[21]	7-bromo tetradecane	weak	weak	11.06	2.83
18.48	17.8[21]	C15-azlactone	weak	weak	10.79	4.09
24.76	30.9[21]	butyl glycidyl ether	weak	weak	12.49	1.60

[a] Boldface indicates discrepant classification between calculated and literature values.

Table 2. Outlier Compounds with Calculated EC3 Values and Possible Causes for Poor Prediction

$EC3_{calc}$	$EC3_{lit}$		compound	possible cause
0.06	9.2[21]	**1**	trimelletic anhydride	anhydride ring
2.1	11.3[21]	**2**	2,3-butanedione	adjacent ketones
7.3	26[27]	**3**	5-methyl-hexane-2,3-dione	adjacent ketones
109	13.1[21]	**4**	1-iododecane	iodine
136	24.2[21]	**5**	1-iodononane	iodine
93.7	19.1[21]	**6**	1-iodohexadecane	iodine
98.1	13.8[21]	**7**	tetradecyl iodide	iodine
19.1	5.2[21]	**8**	tetramethylthiuram disulfide	sulfur
20.2	8.8[21]	**9**	dodecyl methane sulfonate	sulfur
7.0	71.9[21]	**10**	dimethyl sulfoxide	sulfur
21.6	98[21]	**11**	oleyl methane sulfonate	sulfur
6.19	68.2[21]	**12**	2-ethyl butraldehyde	large literature value
9.2	95.8[21]	**13**	xylene	large literature value
6.2	68.2[21]	**14**	2-ethyl butyraldehyde	large literature value
9.7	23.1[21]	**15**	*cis*-6-nonenal	internal alkene (not $\alpha\beta$ unsaturated)
10.0	6.8[5]	**16**	undec-10-enal	internal alkene (not $\alpha\beta$ unsaturated)
18.9	35[25]	**17**	ethylene glycol dimethacrylate	outlier
15.7	2.1[21]	**18**	c4-azlactone	outlier
16.5	1.3[21]	**19**	c6-azlactone	outlier
16.9	2.8[21]	**20**	c9-azlactone	outlier

similar difficulty predicting sensitization potentials, which may indicate difficulties with the experimental measurements.[21]

RESULTS

The training set statistics ($n = 50$) include $R^2 = 0.773$, adjusted $R^2 = 0.763$, $R^2_{CV} = 0.738$, $F = 79.9$, VIF = 1.07 for each descriptor, p significance < 0.001, t-test values (intercept, −8.8; HOMO−LUMO, 8.9; $FPSA2_{ESP,}$ 11.0), and $\chi^2 = 87.0$ ($\alpha < 0.001$) using eq 7.

$$EC3 = 9.16FPSA2_{ESP} + 4.29E_{HOMO-LUMO} - 45.89 \quad (7)$$

A plot of experimental versus calculated values for the training set is found in Figure 1A. This plot also includes a categorization scheme for the molecules, represented by the area rectangles. As previously discussed, small differences in EC3 values may not necessarily predict varying biological activity. For this reason, a *bin selection* method was used to classify the molecules into three different groups (bins) based on an EC3 classification of strong, moderate, and weak. Rather than creating different category ranges arbitrarily, three training set ranges were created that would give the strong and moderate groups similar quantities of compounds ($n_{strong} = 18$, $n_{moderate} = 21$) with fewer compounds in the weak range ($n_{weak} = 11$). The lack of quantitative data in the weak range is due to the reporting style of weak sensitizers. Weak sensitizers with large EC3 values are frequently reported as exceeding a value, such as EC3 > 30%, with accurate measurement considered unnecessary. Ideally, the compound would be retested for more quantitative data, but as discussed earlier, EC3 = 30% versus EC3 = 40% will not likely indicate a substantial difference in sensitization potency.

Strong sensitizers ($EC3_{calc}$ < 6%) had experimental values of $EC3_{exp}$ < 5%. This is propitious in that the molecules predicted to be strongly sensitizing would have experimental values even stronger than predicted. For the weak range with calculated $EC3_{calc}$ > 13%, $EC3_{exp}$ values were larger than 16%. This is again a favorable situation, as predicted weak molecules have even weaker experimental EC3 values.

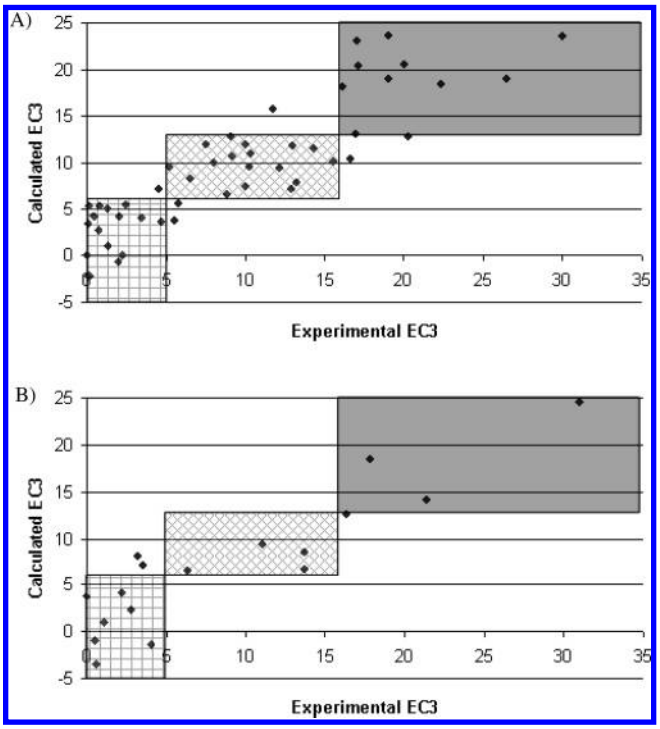

Figure 1. Training set (A) and validation set (B) for the sensitization model. Regions are coded according to classification of sensitization: squares = strong, diamonds = moderate, gray = weak.

The training set does contain several negative values. Although a negative EC3 has no physical meaning, the values fall into the "strong" category and have been interpreted as such. This *bin selection* method to categorize the sensitizing potency successfully predicted 44 out of 50 training set molecules, with the results plotted in Figure 1A.

The external validation set included 17 molecules and is shown in Figure 1B. In this series, 14 molecules are correctly classified, with an R^2 slightly better than the training set ($R^2 = 0.877$). Incorrect predictions from both the training and validation sets were only off by a single category, with no strong members being predicted as weak, or vice versa.

The model has been further tested on compounds *3M/ESPE−1* and *3M/ESPE−2*, in which the sensitization

potential was unknown.[29] The model predicted EC3 = 27 for *3M/ESPE-1* and EC3 > 100 for *3M/ESPE-2*. Subsequent experimental results indicated successful categorization with EC3 = 35 for *3M/ESPE-1* and EC3 > 100 for *3M/ESPE-2*. The predictions also reduced the number of necessary animal tests for each compound by providing the critical concentrations for SI = 3. Serendipitous insight from this test resulted because the two compounds contained an element not represented in the training set. Although caution should be taken for prediction under these circumstances with such a small data set, the tests indicate that the element should not impart substantial effects on sensitization potentials, which may be useful for structural modifications in molecular design.

DISCUSSION

Structure−activity correlations based on the quantitative EC3 values have been developed, but many are limited to small sets of molecules.[5,8,30,31] Larger models have successfully predicted sensitization with external validation test comparisons, [21,32] but only for classification into categorical strength rather than specific EC3 values. Unfortunately, interpretation of the descriptors for chemical sense and for the structural design of compounds with low sensitization potentials is difficult for the latter models.

The model in the present study involves only two descriptors given in eq 7. This avoids one of the pitfalls found in many of the other QSAR models described that have low observation-to-data ratios, ideally at least 1:5.[33] In addition, the descriptors are quantum mechanical based on molecular orbital calculations. When describing a process such as protein interaction or skin permeation (related to hydrophobicity), these descriptors provide more relevant information for molecular interactions than geometric or topological indices descriptors.

The fractional positively charged surface area descriptor based on the electrostatic potential charge ($FPSA2_{ESP}$)[34] may be relevant to the skin permeation of a substance. In agreement with our finding, a recent computational model to predict blood−brain barrier permeation also included a similar FPSA descriptor.[35] Also, other studies for the prediction of hydrophobic indices and octanol/water partition coefficient values have been shown to correlate with the combination of calculated electrostatic potentials and molecular surface areas.[36−39] A larger FPSA2 leads to a higher required threshold for the compound to permeate the skin and induce sensitization. In short, increased positively charged surface area seems to reduce sensitization potential.

The HOMO−LUMO energy gap also makes chemical sense for the process of skin sensitization. According to frontier molecular orbital theory, smaller gaps between HOMO and LUMO orbitals are a simple measure of kinetic stability.[40,41] Thus, independent of the compound's ability to penetrate the skin barrier, the reactivity of the compound toward a protein is also relevant to sensitization. As such, a higher $E_{HOMO-LUMO}$ is consistent with a higher concentration required for sensitization.

Many interlaboratory and intralaboratory studies have been performed, and the results are relatively consistent.[16,42−44] However, the actual values reported by different laboratories may vary by 3-fold or more. To test the impact of using different experimental EC3 values on the predictive ability of this QSAR, a second training set was created using the average of available reported values of four compounds. One such molecule, isoeugenol, was tested at the same laboratory 29 times with a range of EC3 values from 0.5 to 2.6 and an average value of 1.2.[45] The initial isoeugenol property value used in the training set had an EC3 experimental value of 3.5 (calculated from experimental SI readings and concentrations).[15] Although this did not change the category from a strong sensitizer, the new value is $^1/_3$ of the initially entered value. Similarly, dinitrochlorobenzene (initial property value of 0.016 calculated manually)[26] compared with an average over five separate trials of 0.05.[4] Again, this average would not change the category from a strong sensitizer, but the change was 3-fold. Two other molecules, hexyl cinnamal and eugenol, had interlaboratory comparisons for ranges described in Table 3, with even greater absolute variation.

Table 3. Molecules with Interlaboratory Values to Compare with Initial QSAR Model Values

structure[ref]	range	*n*	mean	standard deviation	initial value
isoeugenol[45]	0.5−2.6	29	1.2	0.6	3.5
dinitrochlorobenzene[4]	0.03−0.06	5	0.05	0.013	0.016
hexyl cinnamal[4,46]	7.0−12.2	15	9.5	2	8
eugenol[43,47]	5.1−14.5	10	9.8	3.8	12.9

After replacing the initial EC3 values with the new averaged values, the model was recalculated. The categorization by the bin selection method resulted in no changes, and all molecules remained in the same groups as the initial models. Statistics were similar for the new model, with $R^2 = 0.778$ and $F = 82.4$. The coefficients of eq 7 did change, but only to a slight degree. Thus, collecting and averaging all available experimental values should not be necessary, as the predictive ability of the model is quite robust.

CONCLUSIONS

The model described here successfully categorizes potential sensitizing compounds on the basis of LLNA experimental data using quantum mechanical data with descriptors that relate to the understood mechanism of sensitization. Categorizing each molecule into one of three ranges resulted in correct predictions for 44 of the 50 training set molecules and 14 of the 17 validation molecules for an overall successful classification of greater than 80%. Functionalities in the model include halogenated compounds, aromatic compounds, alcohols, aldehydes, and ketones. However, compounds with iodine, sulfur, anhydride rings, adjacent ketones, and alkene aldehydes that are not Michael reactants were not compatible. As such, this model should be restricted to functionalities in the training set or caution should be maintained when such functionalities are tested.

Reported values for LLNA experimental data vary in the literature, illustrating that small differences in EC3 values do not necessarily correspond to real differences in sensitization potency. This being the case, categorizing the data into groups allows for accurate prediction of relative potency. Collecting available data and using average values is probably not required, as this study indicates such steps do not substantially affect the model.

There are no in vitro methods recognized by regulatory authorities to replace animal testing for sensitization potential. The use of QSAR models (in silico) such as this one may assist in screening candidates by predicting likely sensitizing agents at early stages in the discovery cycle. In addition, critical concentrations provided by this model can reduce the number of animal tests necessary by eliminating ranging tests at concentrations that do not generate useful information. This model is currently being used to predict such critical concentrations within this research group.

Further development with additional structures and chemical functionalities may allow an expansion of this model to include a broader array of compounds such as iodine or sulfur compounds. This may also include activated or metabolically modified molecules, which are the actual sensitizing agents, in contrast to the parent structures.

ACKNOWLEDGMENT

This investigation was supported by Research Grants DE09696 and DE07294 from the National Institute of Dental and Craniofacial Research, Bethesda, Maryland 20892.

REFERENCES AND NOTES

(1) Kimber, I.; Dearman, R. J.; Basketter, D. A.; Ryan, C. A.; Gerberick, G. F. *Contact Dermatitis* **2002**, *47*, 315−328.
(2) Scheynius, A. In *Allergic Contact Dermatitis. The Molecular Basis*; Lepoittevin, J. P., Basketter, D. A., Goossens, A., Karlberg, A.-T., Eds.; Springer-Verlag: Heidelberg, Germany, 1997; pp 4−18.
(3) Kimber, I.; Basketter, D. A.; Berthold, K.; Butler, M.; Garrigue, J.-L.; Lea, L. J.; Newsome, C.; Roggeband, R.; Steiling, W.; Stropp, G.; Waterman, S.; Wiemann, C. *Toxicol. Sci.* **2001**, *59*, 198−208.
(4) Basketter, D. A.; Gerberick, G. F.; Kimber, I.; Loveless, S. E. *Food Chem. Toxicol.* **1996**, *34*, 985−997.
(5) Patlewicz, G. Y.; Wright, Z. M.; Basketter, D. A.; Pease, C. K.; Lepoittevin, J. P.; Arnau, E. G. *Contact Dermatitis* **2002**, *47*, 219−226.
(6) Smith Pease, C. K.; Basketter, D. A.; Patlewicz, G. *Clin. Exp. Dermatol.* **2003**, *28*, 177−183.
(7) Basketter, D. A.; Gerberick, G. F.; Kimber, I. *Food Chem. Toxicol.* **2001**, *39*, 621−627.
(8) Basketter, D. A.; Roberts, D. W.; Cronin, M.; Scholes, E. W. *Contact Dermatitis* **1992**, *27*, 137−142.
(9) Basketter, D. A.; Scholes, E. W.; Kimber, I.; Botham, P. A.; Hilton, J.; Miller, K.; Robbins, M. C.; Harrison, P. T. C.; Waite, S. J. *Toxicol. Methods* **1991**, *1*, 30−43.
(10) Haneke, K. E.; Tice, R. R.; Carson, B. L.; Margolin, B. H.; Stokes, W. S. *Regul. Toxicol. Pharmacol.* **2001**, *34*, 274−286.
(11) Frankild, S.; Volund, A.; Wahlberg, J. E.; Anderson, K. E. *Acta Derm. Venereol.* **2000**, *80*, 256−262.
(12) Basketter, D. A.; Scholes, E. W. *Food Chem. Toxicol.* **1992**, *30*, 65−69.
(13) Kimber, I.; Dearman, R. J.; Scholes, E. W.; Basketter, D. A. *Toxicology* **1994**, *93*, 13−31.
(14) Basketter, D. A.; Evans, P.; Fielder, R. J.; Gerberick, G. F.; Dearman, R. J.; Kimber, I. *Food Chem. Toxicol.* **2002**, *40*, 593−598.
(15) Basketter, D. A.; Lea, L. J.; Dickens, A.; Briggs, D.; Pate, I.; Dearman, R. J.; Kimber, I. *J. Appl. Toxicol.* **1999**, *19*, 261−266.
(16) Warbrick, E. V.; Dearman, R. J.; Lea, L. J.; Basketter, D. A.; Kimber, I. *J. Appl. Toxicol.* **1999**, *19*, 255−260.
(17) Basketter, D. A.; Gilmour, N. J.; Briggs, D.; Ullmann, L. G.; Gerberick, G. F.; Ryan, C. A.; Dearman, R. J.; Kimber, I. *Contact Dermatitis* **2003**, *49*, 37−41.
(18) Basketter, D. A.; Smith Pease, C. K.; Patlewicz, G. *Clin. Exp. Dermatol.* **2003**, *28*, 218−221.
(19) Basketter, D. A.; Wright, Z. M.; Warbrick, E. V.; Dearman, R. J.; Kimber, I.; Ryan, C. A.; Gerberick, G. F.; White, I. R. *Contact Dermatitis* **2001**, *45*, 89−94.
(20) Basketter, D. A.; Blaikie, L.; Dearman, R. J.; Kimber, I.; Ryan, C. A.; Gerberick, G. F.; Harvey, P.; Evans, P.; White, I. R.; Rycroft, R. J. G. *Contact Dermatitis* **2000**, *42*, 344−348.
(21) Estrada, E.; Patlewicz, G.; Chamberlain, M.; Basketter, D. A.; Larbey, S. *Chem. Res. Toxicol.* **2003**, *16*, 1226−1235.
(22) Dewar, M.; Jie, C.; Yu, J. *Tetrahedron* **1993**, *49*, 5003−5038.
(23) *Ampac with Graphical User Interface*, version 8.0; Semichem, Inc.: Shawnee Mission, KS, 2004.
(24) *COmprehensive DEscriptors for Structural and Statistical Analysis (Codessa)*, Semichem, Inc.: Shawnee Mission, KS.
(25) Basketter, D. A.; Kimber, I. *Allergy* **2001**, *56*, 937−943.
(26) Ashby, J.; Basketter, D. A.; Paton, D.; Kimber, I. *Toxicology* **1995**, *103*, 177−194.
(27) Griem, P.; Goebel, C.; Scheffler, H. *Regul. Toxicol. Pharmacol.* **2003**, *38*, 269−290.
(28) Basketter, D. A.; Cookman, G.; Gerberick, G. F.; Hamaide, N.; Potokar, M. *Food Chem. Toxicol.* **1997**, *35*, 417−425.
(29) Kostoryz, E. et al. Manuscript in preparation.
(30) Roberts, D. W.; Basketter, D. A. *Contact Dermatitis* **2000**, *42*, 154−161.
(31) Patlewicz, G.; Basketter, D. A.; Smith, C. K.; Hotchkiss, S. A. M.; Roberts, D. W. *Contact Dermatitis* **2001**, *44*, 331−336.
(32) Fedorowicz, A.; Zheng, L.; Singh, H.; Demchuk, E. *Int. J. Mol. Sci.* **2004**, *5*, 56−66.
(33) Cronin, M. T. D.; Schultz, T. W. *THEOCHEM* **2003**, *622*, 39−51.
(34) Stanton, D. T.; Jurs, P. C. *Anal. Chem.* **1990**, *62*, 2323−2329.
(35) Subramanian, G.; Kitchen, D. B. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 643−664.
(36) Breindl, A.; Beck, B.; Clark, T. *J. Mol. Model.* **1997**, *3*, 142−155.
(37) Kantola, A.; Villar, H. O.; Loew, G. H. *J. Comput. Chem.* **1991**, *12*, 681−689.
(38) Alkorta, I.; Villar, H. O. *Int. J. Quantum Chem.* **1992**, *44*, 203-218.
(39) Brinck, T.; Murray, J. S.; Politzer, P. *J. Org. Chem.* **1993**, *58*, 7070−7073.
(40) Klopman, G. *J. Am. Chem. Soc.* **1968**, *90*, 223−234.
(41) Aihara, J. *Phys. Chem. Chem. Phys.* **2000**, *2*, 3121−3125.
(42) Kimber, I.; Basketter, D. A. *Food Chem. Toxicol.* **1992**, *30*, 165−169.
(43) Loveless, S. E.; Ladics, G. S.; Gerberick, G. F.; Ryan, C. A.; Basketter, D. A.; Scholes, E. W.; House, R. V.; Hilton, J.; Dearman, R. J.; Kimber, I. *Toxicology* **1996**, *108*, 141−152.
(44) Ryan, C. A.; Gerberick, G. F.; Cruse, L. W.; Basketter, D. A.; Lea, L. J.; Blaikie, L.; Dearman, R. J.; Warbrick, E. V.; Kimber, I. *Contact Dermatitis* **2000**, *43*, 95−102.
(45) Basketter, D. A.; Cadby, P. *Contact Dermatitis* **2004**, *50*, 15−17.
(46) Dearman, R. J.; Wright, Z. M.; Basketter, D. A.; Ryan, C. A.; Gerberick, G. F.; Kimber, I. *Contact Dermatitis* **2001**, *44*, 357−361.
(47) Gerberick, G. F.; Ryan, C. A.; Kimber, I.; Dearman, R. J.; Lea, L. J.; Basketter, D. A. *Am. J. Contact Dermatitis* **2000**, *11*, 3−18.

CI050018Z