
Brogan & Partners

Prioritizing Testing of Organic Compounds Detected as Gas Phase Air Pollutants:
Structure-Activity Study for Human Contact Allergens

Author(s): Robert Johnson, Orest T. Macina, Cynthia Graham, Herbert S. Rosenkranz,
Glen R. Cass and Meryl H. Karol

Source: *Environmental Health Perspectives*, Vol. 105, No. 9 (Sep., 1997), pp. 986-992

Published by: The National Institute of Environmental Health Sciences

Stable URL: <http://www.jstor.org/stable/3433880>

Accessed: 10-08-2016 16:15 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://about.jstor.org/terms>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Brogan & Partners, The National Institute of Environmental Health Sciences are collaborating with JSTOR to digitize, preserve and extend access to *Environmental Health Perspectives*

Prioritizing Testing of Organic Compounds Detected as Gas Phase Air Pollutants: Structure–Activity Study for Human Contact Allergens

Robert Johnson,¹ Orest T. Macina,² Cynthia Graham,² Herbert S. Rosenkranz,² Glen R. Cass,¹ and Meryl H. Karol²

¹Environmental Engineering Science Department, California Institute of Technology, Pasadena, CA 91125 USA; ²University of Pittsburgh, Graduate School of Public Health, Center for Environmental and Occupational Health and Toxicology, Pittsburgh, PA 15238 USA

Organic compounds that are used or generated anthropogenically in large quantities in cities can be identified through their presence in the urban atmosphere and in air pollutant source emissions. Compounds identified by this method were screened to evaluate their potential to act as contact allergens. The CASE and MULTICASE computer programs, which are based on the detection of structure–activity relationships (SAR), were used to evaluate this potential. These relationships first are determined by comparing chemical structures to biological activity within a learning set comprised of 458 compounds, each of which had been tested experimentally in human trials for its sensitization potential. Using the information contained in this learning set, CASE and MULTICASE predicted the activity of 238 compounds found in the atmosphere for their ability to act as contact allergens. The analysis finds that 21 of 238 compounds are predicted to be active contact allergens (probability >0.5), with potencies ranging from mild to very strong. The compounds come from chemical classes that include chlorinated aromatics and chlorinated hydrocarbons, *N*-containing compounds, phenols, alkenes, and an *S*-containing compound. Using the measured airborne concentrations or emission rates of these compounds as an indication of the extent of their use, together with their predicted potencies, provides an efficient method to prioritize the experimental assessment of contact sensitization of untested organic compounds that can be detected as air pollutants. *Key words:* atmospheric, emissions, contact allergen, pollutants, sensitization, structure–activity relationship.

Environ Health Perspect 105:986–992 (1997). <http://ehp.niehs.nih.gov>

Skin sensitization or allergic contact dermatitis (ACD) is a common condition associated with immunologically mediated dermal inflammation in response to contact with certain compounds. Identification and regulation of allergens has taken place historically in the occupational setting. Since ACD is a frequent condition in the general population, there is a need to identify allergens that are present in environmental settings.

One particularly efficient method that can be used to conduct a systematic search for organic chemicals that are in widespread use is to examine their concentration in the atmosphere of cities. The atmosphere acts as the reservoir for the organics that evaporate from solvents, paint strippers, household cleaning chemicals, and fuels, as well as industrial chemicals. By examining those organic chemicals that are measurable in the atmosphere or in air pollutant source emissions, one can quickly obtain a broad overview of the entirety of the volatile organic compound use in the city as a whole. Even when the atmospheric concentration of a particular chemical is too low to directly represent a significant exposure to the skin, the presence of the volatile chemical often signals that there was the potential for skin contact with the liquid phase of that chemical at the place where it originally was in use.

Traditional methods for identifying sensitizers include animal and human response

tests, which are often time consuming and expensive. Various structure–activity relationship (SAR) models have been developed to evaluate the sensitization potential of compounds that have not been tested experimentally. These models involve the identification of structural alerts that are responsible for allergic sensitization. One such model is based on the CASE/MULTICASE system developed by Rosenkranz and Klopman (1,2).

Most current SAR models of allergic contact dermatitis are based on an assumed mechanism of activity. It is generally accepted that ACD involves the initial penetration of the allergen through the skin followed by its binding to proteins to initiate an immune response. It can be inferred that characteristics that enhance the ability of the chemical to penetrate the skin (such as lipophilicity) and react with proteins (such as electrophilicity) might be found in allergens. In fact, most SAR models of ACD incorporate these assumptions (3).

The CASE/MULTICASE system (C/MC) is different from many other SAR models in that it operates independently of an assumed mechanism of activity, i.e., it is knowledge based. C/MC analyzes compounds of experimentally determined activity, looking for structural features that are statistically associated with active chemicals. The C/MC system offers a fast

and reliable method for evaluating a large number of structurally diverse compounds. Its current limitations are that metallic species cannot be analyzed, and chemicals must contain at least two bonded nonhydrogen atoms. The C/MC system has been applied to a number of biological endpoints including ACD (3), mutagenicity (4), and carcinogenicity (5,6). It is ideal for analyzing the sensitization potential of the numerous organic compounds that are ubiquitous as a result of their presence in polluted air.

The aim of this study is to apply the predictive capabilities of C/MC to the gas phase organic compounds identified in urban air and in air pollution source emissions in order to prioritize compounds for further testing.

Methods

Classification of organic air pollutants. The organic compounds that are the focus of this study were obtained from two separate sets of data collected in the Los Angeles, California, area. On 8–9 September 1993, extensive measurements of both gas-phase and particle-phase organics were made at five field monitoring sites during a severe photochemical smog event (7). From this study, 156 different gas-phase organic compounds present in outdoor air were quantified and 48-hr average concentrations were calculated for each site. The second study consists of compounds identified in the emissions from pollution sources. The 11 largest source types that contribute to nonmethane anthropogenic gas-phase organic air pollutant emissions in Southern California during August 1987 were resurveyed (8). These are (in decreasing order of importance) noncatalyst-equipped motor vehicle exhaust, catalyst-equipped motor vehicle exhaust, industrial surface coatings, solvent-borne architectural surface coatings, gasoline headspace vapors, whole gasoline vapors, domestic solvents, industrial adhesives, diesel engine exhaust, composite thinning solvents, and waterborne architectural surface coatings. The organics

Address correspondence to M.H. Karol, University of Pittsburgh, Graduate School of Public Health, Center for Environmental and Occupational Health and Toxicology, 260 Kappa Drive, Pittsburgh, PA 15238 USA.

This study was supported by NIEHS grant ES 05651. Received 4 April 1997; accepted 22 May 1997.

present in these top source emitters, as well as in a few other sources, were grouped into one master list that consisted of 82 compounds in addition to those measured in the atmospheric samples. A composite list of 238 organic compounds that represent all volatile organic compounds found either in the source or ambient data sets was generated from the two studies and was submitted to the C/MC system for analysis.

CASE and MULTICASE methodology.

A learning set containing compounds that have been tested for human ACD activity was compiled from a literature survey. Chemicals were selected if standard human maximization tests (HMT) had been performed or if three or greater case reports of positive patch tests were cited. (The literature citations are available upon request.) The result was a database composed of 458 compounds of which 229 were active con-

tact allergens and 229 were inactive. In fact, a total of 695 inactive chemicals was identified. Since previous studies had indicated that the best SAR model was obtained with a learning set composed of an equal number of active and inactive chemicals (9), three test models were developed, each containing the 229 active sensitizers but coupled with different nonoverlapping sets of 229 inactive chemicals. No differences were found between the three models regarding the structural alerts identified. Accordingly, one model was randomly selected for further evaluation. The list of chemicals composing the model is available upon request. Each chemical in the model was assigned a potency based on the concentration of chemical used for the challenge dose, and in HMTs, on the sensitization rate (3). Potencies were designated as follows: 10, inactive; 25, marginally active; 39, weakly

active; 49, moderately active; 59, strongly active; and 69, extremely active. Two chemicals, i.e., α -pinene and 1,2-propanediol, detected as gas-phase air pollutants, were present in the learning set (Table 1).

The C/MC systems used in this study were developed as tools for predicting the activity of chemicals with respect to specific biological and/or toxicological endpoints (4). CASE and MULTICASE (MULTICASE Inc., Beachwood, OH) examine all possible molecular fragments containing 2–10 contiguous heavy atoms that can be formed from decomposition of an organic molecule. (Accordingly, a molecule must contain at least three nonhydrogen atoms to be processed by the C/MC system.) Fragments that are statistically associated with biological activity or inactivity are identified. The fragments are generated from chemicals in a learning set composed of biologically active

Table 1. Organic chemicals detected as gas-phase air pollutants that are predicted to be active contact allergens

No.	Species measured in the ambient air	Experimental potency	Predicted potency (MC units)	Bayes' probability that compound is active	Atmospheric concentration (ppb) ^a			Air basin-wide mass emission ^b	
					High	Low	Rank ^c	kg/day	Rank ^c
Aromatic compounds									
1	<i>p</i> -Dichlorobenzene		58	0.999	0.19	0.05	6	— ^d	— ^e
2	Chlorobenzene		57	0.999				59	9
Alkanes, alkenes, and alkynes									
3	1-Hexene		91	0.999	0.19	0.03	3	1,573	3
4	α -Pinene ^f	39	30	0.483	0.20	ND	7	40,910	2
Chloro alkanes, and chloro alkenes									
5	Dichloromethane		61	0.999	4.56	0.90	1	12,708	1
6	Chloroethane		61	0.999				— ^d	— ^e
7	1-Chlorobutane		67	0.999				— ^d	— ^e
8	3-Chloromethyl heptane		60	0.999				— ^d	— ^e
9	Trichloroethylene		63	0.603	0.45	0.02	4	786	7
10	Tetrachloroethylene		65	0.603	1.06	0.25	2	— ^d	— ^e
11	1,1 Dichloroethene		61	0.603	1.04	0.38	5	— ^d	— ^e
Alcohols and aldehydes ^g									
12	2 Methyl 1-propanol		39	0.999				— ^d	— ^e
13	1,2-Propanediol	49	49	0.999				— ^d	— ^e
<i>N</i> -Containing compounds									
14	Ethanamine		61	0.975				1,136	4
15	<i>N,N</i> -Dimethyl methanamine		58	0.603				1,136	5
16	<i>N,N</i> -Dimethylformamide		58	0.603				952	6
Phenols									
17	Phenol		56	0.999				42	10
18	4-Butyl phenol		58	0.999				79 ^h	8
19	4-Pentyl phenol		58	0.999				37 ^h	11
20	2,4-Dimethyl phenol		59	0.999				33 ^h	12
<i>S</i> -Containing compounds									
21	Tetramethyl thiourea		58	0.999				13	13

Abbreviations: ND, below detection limits; C/MC, CASE/MULTICASE; MC, MULTICASE.

^aRange of average ambient concentration of gas-phase organic compounds measured during a Los Angeles smog event on 8–9 September 1993; high and low numbers are range of 2-day average concentrations measured among four urban sites during that particular event (7).

^bAir basin-wide mass emission rates in the South Coast Air Basin that surrounds Los Angeles for 27 August 1987 (8).

^cRank is determined from the product of Bayes' probability value (see Table 2) \times MC units (potency) \times atmospheric concentration. The chemical yielding the highest product receives the rank of 1.

^dCompound is present in the source emission library, but the basin-wide emission rate is not calculated.

^eRank cannot be calculated because the measured mass emission rate is not known.

^fChemical was included in the table because the probability of activity (0.48) was very close to the cut-off value of 0.50.

^gGlyoxal is known experimentally to be a strong contact sensitizer, but is too small to be fragmented by C/MC.

^hEmission rate is given is for all isomers with the same molecular formula, not just the compound shown.

and inactive compounds. Fragments significantly enriched in the active group of chemicals are considered activating fragments (biophores), while those associated predominantly with inactive compounds are considered deactivating (biophobes).

CASE uses the significant fragments to classify a chemical submitted for analysis. Classification is expressed as the probability of being active or inactive. CASE also derives a quantitative estimation of potency by generating a global multivariate linear regression equation using structural determinants as well as the log P (octanol:water partition coefficient).

The MULTICASE program differs in that it uses a hierarchical approach to identify biophores that account for activity. MULTICASE uses molecular fragments as well as 2-dimensional distance descriptors to find the biophore accounting for the majority of the active class of chemicals. Chemicals containing the biophore are removed from further consideration, and subsequent biophores are identified that explain the activity of the remaining compounds. MULTICASE then groups compounds that contain each biophore and examines them for additional descriptors (modulators) responsible for increasing or decreasing the potency of molecules containing the biophore. Modulators may be additional structural fragments, distance descriptors, or physical-chemical properties such as electronic energies calculated by the Hückel method (HOMO, highest occupied molecular orbital, or LUMO, lowest unoccupied molecular orbital) or transport indices (lipophilicity, water solubility). The modulators are considered only in chemicals that contain the biophore under consideration. MULTICASE predicts the probability of activity as well as the potency of untested chemicals based on the presence of biophores and modulators.

A chemical submitted for testing to C/MC will be assigned a predicted probability of activity and a predicted potency. Any fragment present in a chemical being examined that is not encountered in the learning set is identified by a warning of uncertainty as to its relevance for biological activity. Test chemicals that contain no biophores are presumed inactive.

The C/MC system generates four individual predictions of activity for each test compound: the CASE potency estimate (in CASE units), the CASE-generated probability that the compound is active, the MULTICASE potency estimate (in MULTICASE units), and the MULTICASE-generated probability that the compound is active. The sensitivity, specificity, and concordance for each parameter, as defined

below and in Table 2, are then combined using Bayes' theorem (5,6) to yield an overall prediction of activity.

Results

A validation study was performed to determine the predictive capability of the C/MC system when applied to human contact allergens. Chemicals were removed from the learning set (5% each time); the model was then reestablished using this reduced learning set and tested for its ability to predict the activity of the removed chemicals. The results of this *n*-fold validation study are shown in Table 2. The cutoff value for each parameter was modeled for optimal predictivity by varying the four indices within reasonable ranges (C/MC units between 17 and 45; C/MC probabilities between 0.45 and 0.75). The cutoffs selected (see Table 2) gave the best overall concordance between experimental and predicted results.

Comparison of the predictions with the known activities for each of the selected

chemicals yielded the sensitivity, specificity, and concordance for each parameter. The sensitivity (number of correctly predicted actives divided by the total number of experimentally known actives) ranged from 0.67 for CASE units to 0.81 for MULTICASE units. The specificity (number of correctly predicted inactive chemicals divided by the known inactives) ranged from 0.88 for CASE probability to 0.93 for CASE units. The concordance (number of accurately predicted active and inactive compounds divided by the total number of active and inactive chemicals evaluated) ranged from 0.80 to 0.86.

An overall probability that a chemical is active was obtained by application of Bayes' theorem. The method utilizes the sensitivity and specificity of each of the parameters described above together with a prior estimate of the probability of the chemical being active. For the current study, the prior probability was arbitrarily estimated to be 0.20 based upon the number of chemicals that

Table 2. Optimum threshold values and predictive ability of CASE/MULTICASE parameters used in the analysis of the human allergic contact dermatitis (ACD) learning set

Parameter	Cutoff ^a	Sensitivity ^b	Specificity ^c	Concordance ^d
MULTICASE units ^e	18	0.81	0.90	0.86
MULTICASE probability ^f	0.65	0.80	0.91	0.86
CASE units ^e	19	0.67	0.93	0.80
CASE probability ^f	0.69	0.80	0.88	0.84

^aThe threshold (cutoff) value is the parameter value at or above which the compound is considered to be positive for ACD activity.

^bSensitivity is the number of correctly predicted active compounds divided by the total number of active compounds.

^cSpecificity is the number of correctly predicted inactive compounds divided by the total number of inactive compounds.

^dConcordance is the number of correct predictions (both active and inactive) divided by the total number of compounds.

^eMULTICASE units and CASE units are a measure of predicted compound potency.

^fMULTICASE and CASE probabilities are the probabilities that the compound is an active human ACD sensitizer.

Table 3. Patterns of predictions and Bayes' probability of activity obtained from the validation study^a

MULTICASE units	Pattern			Bayes' probability that compound is active
	MULTICASE probability	CASE units	CASE probability	
+	+	+	+	0.999
+	+	-	+	0.978
+	+	+	-	0.975
-	+	+	+	0.968
+	-	+	+	0.965
+	+	-	-	0.603
-	+	-	+	0.534
+	-	-	+	0.515
-	+	+	-	0.501
+	-	+	-	0.483
-	-	+	+	0.415
-	+	-	-	0.037
+	-	-	-	0.034
-	-	-	+	0.026
-	-	+	-	0.023
-	-	-	-	0.001

^aNumber of chemicals tested = 440.

^bValues are assigned when the result for a compound falls at or above (+) or below (-) the cutoff values given in Table 2.

have been documented to be human contact allergens. Bayes' theorem was used to update the probability estimate using each of the prediction models. Table 3 lists the Bayesian probabilities for each of the 16 possible patterns of predictions obtained from the validation study. The Bayesian probabilities were then used to classify the activity of each of the chemicals detected in the atmospheric monitoring experiments or in source emissions. Those chemicals having a probability >0.50 were listed as potentially active contact sensitizers. It should be noted that only 1 of the 238 chemicals examined yielded a Bayesian probability close to the 0.50 value. For this reason, α -pinene (probability of activity = 0.48) is included as a putative positive in Table 1. The probability of activity of each test chemical, multiplied by the predicted potency (estimated by MULTICASE units) and by either the airborne concentration or mass emission rate into the atmosphere, was used to rank the chemicals predicted to be active in order of greatest priority for future testing (see Table 1).

A typical prediction generated by C/MC is presented in Figure 1. For *p*-dichlorobenzene, MULTICASE identifies the chloroarene moiety as a biophore. The biophore is found in 19 compounds in the learning set, 17 of which are contact allergens. Also shown is the electronegativity characteristics of the compound that modulate the activity to increase slightly its potency. MULTICASE calculates a probability that this compound is active (0.857) and has a potency of 58 units. It concludes that the compound is active. CASE analysis identifies two further refined chloroarene biophores and a third disubstituted arene fragment. CASE predicts that the compound is active (0.997 probability) with a predicted potency of 43 units. Based on this pattern of predictions (++++), Bayes' probability of ACD activity for *p*-dichlorobenzene is 0.999 (see Table 3).

Table 1 provides detailed information for the 21 vapor phase organic compounds present in ambient air or in air pollutant source emissions that are either known to be active or are predicted by C/MC to be contact allergens. Data include the experimental activities of those chemicals that have been examined in human testing protocols, the MC potency units, the Bayesian probability associated with each chemical, and the range of the 48-hr average ambient concentration during a Los Angeles photochemical smog episode for those compounds that were identified in the ambient air. The highest concentration shown represents the urban air monitoring site that reported the highest average concentration, while the lower value represents the urban monitoring site that reported the lowest average concentration

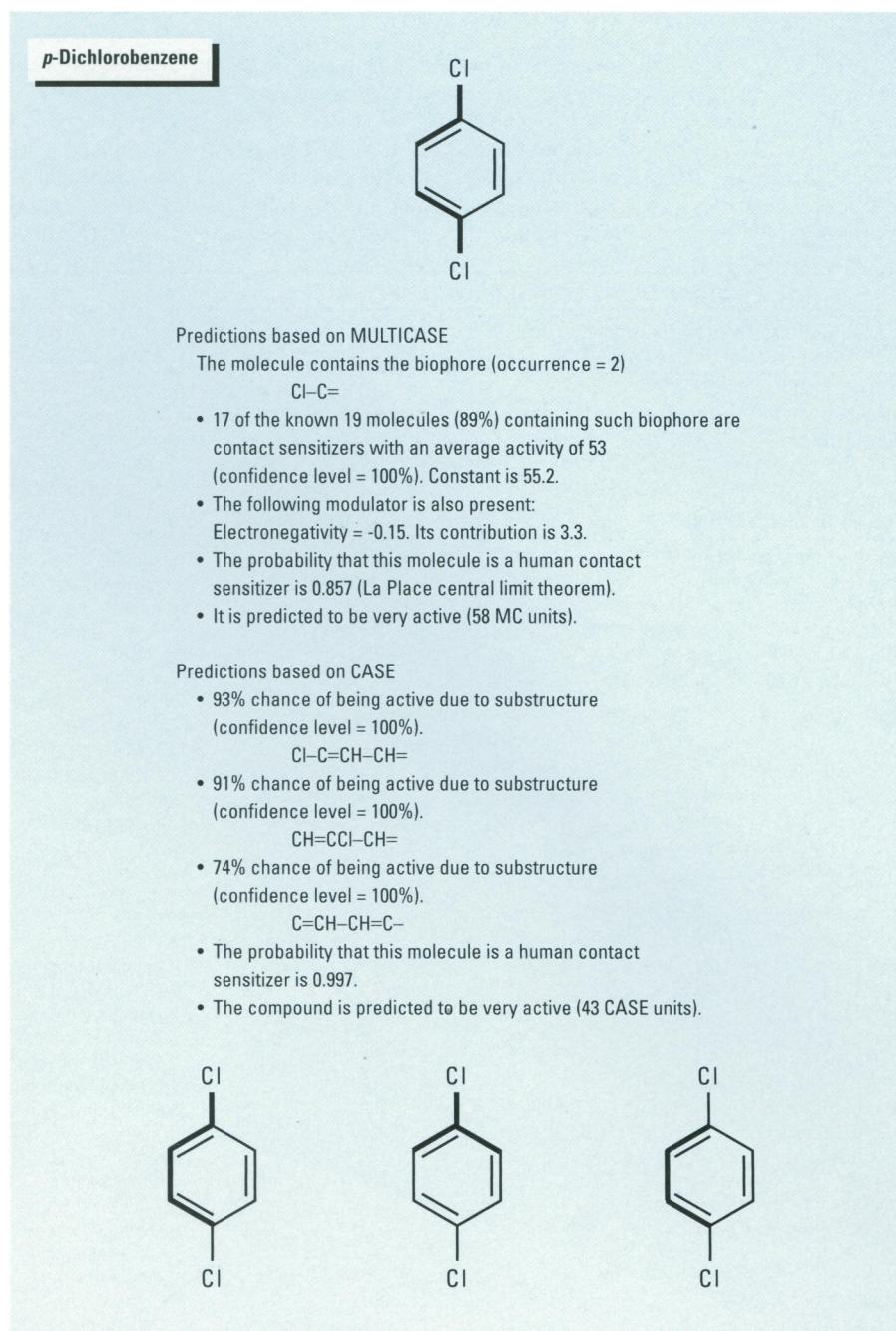


Figure 1. Example of a CASE/MULTICASE prediction.

[not including the offshore island background monitoring site studied during the experiments of Fraser et al. (7)]. The active compounds identified in the ambient air are prioritized (rank) for further testing. Also shown in Table 1 is the air-basin wide mass emission rate in kilograms per day within the South Coast Air Basin that surrounds Los Angeles for each compound listed, followed by a rank ordering of the compounds based on the product of the probability that the chemical is active times its predicted

potency and times its air basin-wide mass emissions rate in the Los Angeles area.

Of the 238 chemicals measured in the ambient air or in air pollutant source emissions, 7 had been tested experimentally. The C/MC system could not evaluate one of the chemicals (glyoxal) because it is too small to be fragmented by C/MC. However, the system accurately predicted the activity of each of the other 6 chemicals. Benzene, ethylbenzene, 2-heptanone, and carvone were each correctly predicted to be inactive; α -pinene

and 1,2-propanediol were predicted to be active (Table 1). MULTICASE identifies seven different portions of the α -pinene ring system as structural alerts, but notes that the fragments are present only in a small number of recognized sensitizers in the learning set. For this reason, MC predicts that the chemical is active but has marginal/weak activity (30 MC units). The Bayesian proba-

bility of activity for α -pinene is 0.48, which is close to the cutoff of 0.50. Accordingly, α -pinene is listed among the chemicals considered to be active.

In Table 1, the five compounds ranked highest in priority for further testing according to activity-weighted atmospheric concentration are dichloromethane, tetrachloroethylene, 1-hexene, trichloroethyl-

ene, and 1,1-dichloroethylene. The five compounds ranked highest in priority according to activity-weighted mass emission rate are dichloromethane, α -pinene, 1-hexene, ethanamine, and *N,N*-dimethylmethanamine.

Table 4 lists the compounds that have been tested experimentally and found to be inactive, and compounds that are presumed

Table 4. Compounds predicted to be inactive as contact sensitizers

Compounds determined experimentally to be inactive contact sensitizers (<i>n</i> = 4)	Bayesian probability of activity	Compounds predicted by C/MC to be inactive contact sensitizers ^a (<i>n</i> = 201)	Bayesian probability of activity
2-Heptanone	0.001	<i>n</i> -Dodecanal	0.001
Benzene	0.001	<i>n</i> -Tridecanal	0.001
Ethylbenzene	0.001	<i>n</i> -Tetradecanal	0.001
Carvone	0.001	3-Pentanone, 2,2,4,4-tetramethyl	0.001
Compounds predicted by C/MC to be inactive contact sensitizers ^a (<i>n</i> = 201)		2-Methyl-3-hexanone	0.001
Aromatic compounds (<i>n</i> = 33)		2-Hexanone	0.001
Benzene, 1,2-diethyl	0.001	6-Methyl-2-heptanone	0.001
Benzene, 1,3-diethyl	0.42	Nopinone	0.001
Benzene, 1,4-diethyl	0.42	Cyclohexanone	0.001
Benzene, 1-methyl,3-propyl	0.42	Cyclohexanone, 2-propyl	0.001
Benzene, 1-methyl,3-isopropyl	0.42	Cyclopentanone, 2,4,4-trimethyl	0.001
Toluene	0.001	Acetone	0.001
Styrene	0.001	2-Butanone	0.001
2-Ethyltoluene	0.001	3-Methyl-2-butanone	0.001
3-Ethyltoluene	0.42	4-Methyl-2-pentanone	0.001
4-Ethyltoluene	0.42	Alkanes (<i>n</i> = 37)	
1,2,3-Trimethylbenzene	0.42	2,2-Dimethylpropane	0.001
1,2,4-Trimethylbenzene	0.42	<i>n</i> -Butane	0.001
<i>i</i> -Propylbenzene	0.001	2,2-Dimethylbutane	0.001
<i>n</i> -Propylbenzene	0.001	2,3-Dimethylbutane	0.001
1,3,5-Trimethylbenzene	0.001	Isobutane	0.001
<i>o</i> -Xylene	0.001	Isopentane	0.001
<i>m</i> -Xylene	0.42	<i>n</i> -Pentane	0.001
<i>p</i> -Xylene	0.42	2-Methylpentane	0.001
Benzene, butyl	0.001	3-Methylpentane	0.001
Benzene, (2-methylpropyl)	0.001	2,2-Dimethylpentane	0.001
<i>t</i> -Butylbenzene	0.001	2,3-Dimethylpentane	0.001
<i>sec</i> -Butylbenzene	0.001	2,4-Dimethylpentane	0.001
Benzene, 1-methyl-2-propyl	0.001	3,3-Dimethylpentane	0.001
Benzene, pentyl	0.001	2,2,4-Trimethylpentane	0.001
1 <i>H</i> -Indene	0.42	2,3,4-Trimethylpentane	0.001
2,3-Dihydro-1 <i>H</i> -indene (indane)	0.42	<i>n</i> -Hexane	0.001
2,3-Dihydro-1 methyl-1 <i>H</i> -indene	0.42	2-Methylhexane	0.001
2,3-Dihydro-4 methyl-1 <i>H</i> -Indene	0.42	3-Methylhexane	0.001
2,3-Dihydro-4,7-dimethyl-1 <i>H</i> -Indene	0.42	2,3-Dimethylhexane	0.001
<i>p</i> -cymene	0.42	2,4-Dimethylhexane	0.001
Naphthalene	0.42	2,5-Dimethylhexane	0.001
1-Methylnaphthalene	0.42	2,2,4-Trimethylhexane	0.001
2-Methylnaphthalene	0.42	<i>n</i> -Heptane	0.001
Carbonyls (<i>n</i> = 31)		3-Methylheptane	0.001
<i>m</i> -Tolualdehyde	0.42	2-Methylheptane	0.001
Propanal ^b	0.001	2,2-Dimethylheptane	0.001
Propanal, 2-methyl	0.001	2,4-Dimethylheptane	0.001
Methacrolein (2-methyl-2-propanal) ^b	0.001	2,6-Dimethylheptane	0.001
Methylglyoxal (2-oxopropanal) ^b	0.001	<i>n</i> -Octane	0.001
Benzaldehyde	0.001	3-Methyloctane	0.001
Crotonaldehyde	0.001	2,6-Dimethyloctane	0.001
<i>n</i> -Butanal	0.001	<i>n</i> -Nonane	0.001
Butanal, 3-methyl	0.001	2,6-Dimethylnonane	0.001
<i>n</i> -Pentanal	0.001	<i>n</i> -Decane	0.001
<i>n</i> -Hexanal	0.001	<i>n</i> -Undecane	0.001
<i>n</i> -Heptanal	0.001	<i>n</i> -Dodecane	0.001
<i>n</i> -Nonanal	0.001	<i>n</i> -Tridecane	0.001
<i>n</i> -Octanal	0.001	Ethers (<i>n</i> = 4)	
<i>n</i> -Undecanal	0.001	MTBE	0.001
<i>n</i> -Decanal	0.001	Butyl cellosolve (ethanol, 2-butoxy)	0.001
		Ethyl isopropyl ether	0.001
		Dibutyl ether (butane, 1,1'-oxybis)	0.001

(continued, next page)

to be inactive. As indicated previously, α -pinene had a probability of activity of 0.48 and, accordingly, was included as a putative positive chemical.

Discussion

Contact sensitivity is a delayed-onset allergic inflammation of the skin. Numerous chemicals have been reported to cause this disorder in humans, and many more chemicals have been identified from animal tests.

Structural alerts associated with contact sensitizers are electrophilic and nucleophilic moieties, as well as some aromatic fragments (3).

Several factors have been shown to influence the development of sensitivity.

Table 4. (continued)

Compounds determined experimentally to be inactive contact sensitizers (<i>n</i> = 4)	Bayesian probability of activity	Compounds predicted by C/MC to be inactive contact sensitizers ^a (<i>n</i> = 201)	Bayesian probability of activity
Esters (<i>n</i> = 9)		Cyclopentene	0.001
Ethyl acetate	0.001	<i>l</i> -Methylcyclopentene	0.001
Isoamyl isobutyrate	0.001	Cyclohexane	0.001
Isobutyl isobutyrate	0.001	Methylcyclohexane	0.001
Acetic acid, propyl ester	0.001	Ethylcyclohexane	0.001
Methyl amyl acetate	0.001	Propylcyclohexane	0.001
2-Ethoxy ethyl acetate	0.001	Butylcyclohexane	0.001
Butyl acetate	0.001	1,3-Dimethylcyclohexane	0.001
Acetic acid, 2 methylpropyl ester	0.001	1,4-Dimethylcyclohexane	0.001
Propanoic acid, 2-methyl-, methyl ester	0.001	1-Ethyl-4-methylcyclohexane	0.001
Alkenes (<i>n</i> = 34)		1,1,3-Trimethylcyclohexane	0.001
Isobutene	0.001	1,2,4-Trimethylcyclohexane	0.001
1-Butene	0.001	Cyclohexene	0.001
2-Methyl-1-butene	0.001	Cyclohexene,1-methyl	0.001
3-Methyl-1-butene	0.001	4-Acetyl- <i>l</i> -methylcyclohexene	0.001
<i>cis</i> -2-Butene	0.001	Octahydro- <i>1H</i> -indene	0.001
<i>trans</i> -2-Butene	0.001	1-Methyldecalin	0.001
2-Methyl-2-butene	0.001	2-Methyldecalin	0.001
1,3 Butadiene	0.001	Naphthalene, decahydro	0.001
1-Pentene	0.001	Octahydropentalene	0.001
<i>cis</i> -2-Pentene	0.001	2-Butyltetrahydrofuran	0.001
<i>trans</i> -2-Pentene	0.001	Carvomenthene	0.001
<i>cis</i> -2-Hexene	0.001	Isopulegone	0.001
<i>trans</i> -2-Hexene	0.001	β -Pinene	0.001
<i>cis</i> -3-Hexene	0.001	<i>d</i> -Limonene	0.001
2-Methyl-1-pentene	0.001	Alkynes (<i>n</i> = 5)	
4-Methyl-1-pentene	0.001	1-Butyne	0.001
2,4,4-Trimethyl-1-pentene	0.001	2-Butyne	0.001
2-Methyl-2-pentene	0.001	3,6-Dimethyl-1-octyne	0.001
<i>cis</i> -3-Methyl-2-pentene	0.001	1-Decyne	0.001
<i>cis</i> -4-Methyl-2-pentene	0.001	1-Undecene	0.001
<i>trans</i> -4-Methyl-2-pentene	0.001	Alcohols (<i>n</i> = 8)	
2,4,4-Trimethyl-2-pentene	0.001	2-Propanol	0.001
1-Hexene, 3-methyl	0.001	1-Ethoxy-2-propanol	0.001
1-Heptene	0.001	1-Hexanol, 2-ethyl	0.001
<i>trans</i> -3-Heptene	0.001	1-Heptanol	0.001
1-Octene	0.001	1-Octanol	0.001
2-Octene	0.001	1,2-Ethanediol	0.001
1-Nonene	0.001	2,4-Pentanediol, 2-methyl	0.001
1-Decene	0.001	Cyclohexanol, 3,3,5-trimethyl	0.001
1-Undecene	0.001	Others (<i>n</i> = 3)	
1,3-Butadiene, 2-methyl	0.001	Peroxyacetyl nitrate ^b	0.001
1,4-Pentadiene	0.001	Peroxypropionyl nitrate ^b	0.001
1,8-Nonadiene	0.001	Carbon disulfide	0.001
Isoprene	0.001	Compounds with unknown activity (<i>n</i> = 11)	
Halocarbons (<i>n</i> = 6)		(Chemicals too small to be analyzed by C/MC)	
Dichlorodifluoromethane ^b	0.001	Formaldehyde	
Trichlorofluoromethane ^b	0.001	Acetaldehyde	
1,1,1-Trichloro-2,2,2-trifluoroethane ^b	0.001	Acetylene	
Ethane, 1,1,1-trichloro ^b	0.001	Ethylene	
Carbon tetrachloride	0.001	Propene	
Chloroform	0.001	Ethane	
Cyclic compounds (<i>n</i> = 30)		<i>n</i> -Propane	
Cyclopentane	0.001	Ethanol ^b	
Methylcyclopentane	0.001	1,2-Propadiene	
1,3-Dimethylcyclopentane	0.001	Methanol	
1,2,4-Trimethylcyclopentane	0.001	Chloromethane ^b	
1,2,3-Trimethylcyclopentane	0.001		

C/MC, CASE/MULTICASE.

^aProbability of being active is <0.5.

^bWarning that the chemical contains a fragment that C/MC has not seen previously.

The dose of chemical applied to the skin, as well as the frequency of exposure, affect the sensitization rate. A dose-response relationship has been shown in animal and human testing, with high dose suppression noted in some animal systems. Adjuvants have been shown to enhance the sensitization rate. Compounds with adjuvant activity include detergents, fatty substances, aluminum hydroxide gel, and bacterial peptides. Because sensitization is believed to proceed through skin penetration, recruitment, and activation of inflammatory cells, irritants are also believed to enhance sensitization.

The CASE/MULTICASE system was able to evaluate 148 of the 156 chemicals identified in atmospheric samples and 78 of the additional 82 chemicals identified only in source emissions. Chemicals not evaluated because they were too small to be fragmented included formaldehyde, methanol, glyoxal, and chloromethane. C/MC also issued warnings associated with some predictions. In such cases, the system notes that the test chemical contains a fragment that is not found among the chemicals in the learning set and warns that the influence of this fragment on the biologic activity is uncertain. Such chemicals are noted in Table 4.

CASE/MC identified 21 chemicals with high probability of ACD activity. Of these, 2 have already tested positive in humans (i.e., α -pinene and 1,2-propanediol). The only active compound not listed in Table 1 is glyoxal. Glyoxal is known experimentally to be a strong contact allergen, but it is too small to be fragmented and analyzed by C/MC.

Two compounds, peroxyacetyl nitrate (PAN) and peroxypropionyl nitrate (PPN), that have an overall probability of activity below 0.50 (see Table 4) are worth mentioning. C/MC predicts these compounds to be inactive since neither contains a previously identified biophore. They both contain fragments ($\text{NO}_2\text{-O-}$, $\text{NO}_2\text{-O-O-}$) that are unknown to the learning set. Free radical generators, such as the peroxide structure bonded to the nitro- and carbonyl groups contained in PAN and PPN, are considered to be reactive with skin proteins and are thus suspected allergens (10).

In order to prioritize those compounds predicted to be active contact allergens, each was separately ranked according to probability of activity- and potency-weighted ambient concentration and according to activity- and

potency-weighted mass emission rate. The reason for two ranking schemes is that some compounds are identified in source emissions but not in the ambient air samples, and conversely some compounds are found in the ambient air but not in source emissions. Certain compounds are either emitted at a very low rate or undergo degradation due to atmospheric chemical reactions and are absent from the ambient database. Other compounds identified in the ambient air but not in any source emissions, most notably peroxyacetyl nitrate and peroxypropionyl nitrate, are formed in the atmosphere by atmospheric chemical reactions.

As mentioned, the top five compounds ranked according to activity-weighted mass emission rate are dichloromethane, α -pinene, 1-hexene, ethanamine, and *N,N*-dimethyl methanamine. Dichloromethane, the highest ranked compound, is an excellent solvent that is found in products as diverse as paint strippers, adhesives, and architectural surface coatings. It has a very high probability of activity (0.999) as well as of being highly potent (61 MC units). 1-Hexene also has a very high probability of being active (0.999) and extremely potent (91 MC units). This chemical is emitted directly to the atmosphere from noncatalyst-equipped gasoline-powered motor vehicle exhaust, catalyst-equipped gasoline-powered vehicle exhaust, whole gasoline evaporation, and gasoline headspace vapors. Skin contact can result from exposure to liquid gasoline. Ethanamine and *N,N*-dimethyl methanamine are two potentially active agents directly emitted from animal waste decomposition. α -Pinene, whose activity was discussed earlier, is thought to originate mainly from biogenic sources. Other prominent groups of compounds that are predicted to be active ACD sensitizers include a variety of chlorinated organic solvents (e.g., tetrachloroethylene, trichloroethylene, and 1,1-dichloroethylene) that are used in industrial cleaning operations, and substituted phenols that are emitted from motor vehicle exhaust and other sources.

In this study, 238 organic compounds present in the urban atmosphere in Southern California and in source emissions were narrowed to 21 compounds that are likely human contact allergens. It is recognized that many of the organic compounds found in polluted urban air are present in

very low concentrations and that these concentrations may be too low to cause initial sensitization or subsequent allergic reactions. However, pollutants present at low concentrations in urban air are present at much higher concentrations near their source of emission from the workplace or the home. By examining the polluted urban atmosphere, we are able to detect chemicals that are in widespread use and for which the probability of human contact at high concentrations elsewhere is substantial.

REFERENCES

1. Rosenkranz HS, Klopman G. The application of structural concepts to the prediction of the carcinogenicity of therapeutic agents. In: Burger's Medicinal Chemistry and Drug Discovery, vol 1: Principles and Practice (Wolff ME, ed). 5th ed. New York: John Wiley & Sons, 1995;223-249.
2. Klopman G. MULTICASE: A hierarchical computer automated structure evaluation program. *Quant Struct-Act Relat* 11:176-184 (1992).
3. Graham C, Gealy R, Macina OT, Karol MH, Rosenkranz HS. QSAR for allergic contact dermatitis. *Quant Struct-Act Relat* 15:1-7 (1996).
4. Rosenkranz HS, Klopman G. The structural basis of the mutagenicity of chemicals in *Salmonella typhimurium*: the National Toxicology Program Database. *Mutat Res* 228:51-80 (1990).
5. Zhang YP, Sussman N, Macina OT, Rosenkranz HS, Klopman G. Prediction of the carcinogenicity of a second group of chemicals undergoing carcinogenicity testing. *Environ Health Perspect* 104(suppl 5):1045-1050 (1996).
6. Macina OT, Zhang YP, Rosenkranz HS. Testing, predicting and integrating carcinogenicity. In: *Improved Predictivity of Carcinogens: The Use of a Battery of SAR Models* (Kitchin, ed). New York: Marcel Dekker, Inc. In press.
7. Fraser MP, Grosjean D, Grosjean E, Rasmussen RA, Cass GR. Air quality model evaluation data for organics: 1. Bulk chemical composition and gas/particle distribution factors. *Environ Sci Technol* 30:1731-1743 (1996).
8. Harley RA, Hannigan MP, Cass GR. Respeciation of organic gas emissions and the detection of excess unburned gasoline in the atmosphere. *Environ Sci Technol* 26:2395-2408 (1992).
9. Liu M, Sussman N, Klopman G, Rosenkranz HS. Estimation of the optimal data base size for structure-activity analyses: the *Salmonella* mutagenicity data base. *Mutat Res* 358:63-72 (1996).
10. Payne MP, Walsh PT. Structure-activity relationships for skin sensitization potential: development of structural alerts for use in knowledge based toxicity prediction systems. *J Chem Inf Comput Sci* 34:154-161 (1994).