Halogenated Aliphatic Toxicity QSARs Employing Metabolite Descriptors

Steven Trohalaki* and Ruth Pachter

Materials & Manufacturing Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio 45433-7702

Kevin T. Geiss and John M. Frazier

Human Effectiveness Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio 45433

Received November 12, 2003

The toxic effects from exposure to halogenated hydrocarbons (HAs), which are produced in large amounts and used in a variety of applications, are well-known. Previously, QSARs for the toxicity of a series of HAs in vitro have been studied extensively. In this work, using a composite toxicity metric calculated from a set of five in vitro hepatotoxicity endpoints determined for 20 HAs, we find that QSARs derived using quantum descriptors calculated from the neutral HA species are statistically similar to QSARs calculated from HA metabolites. In most cases, QSARs derived using descriptors calculated from both neutral HAs and metabolites are statistically superior to those derived using either neutral-HA descriptors or metabolite descriptors. However, to properly utilize metabolite descriptors, multiple QSARs, each of which utilizes a set of HAs that form unique metabolites, must be derived and toxicity values calculated therefrom must be averaged. These average toxicity values agree better with experiment than those calculated from the neutral-HA QSARs.

1. INTRODUCTION

Halogenated aliphatics (HAs) are produced in large amounts each year and are used in a variety of industrial and domestic applications.[1] The health hazards of several HAs have been well documented.[2−8] The primary mechanism of toxic response to HAs has been shown to involve reductive dehalogenation mediated by cytochrome P450.[9,10] Previously, we investigated the toxicity of HAs and described the application of QSARs for toxicity prediction, showing that a descriptor that measures the propensity of a HA to accept an electron from a specific donor is more appropriate and produces a better QSAR than a descriptor that is a general measure of a compound's affinity for an electron.[8] In addition, we previously showed that, although adequate QSARs were obtained using descriptors calculated with semiempirical molecular orbital methods, superior QSARs were obtained using descriptors calculated with either ab initio molecular orbital theory or density functional theory (DFT) and split-valence basis sets with polarization functions.[7]

We previously employed primary rat hepatocytes and a novel exposure system (the VITROBOX)[11] to study the metabolism of 10 halogenated methanes,[12] which represent nearly the full complement of brominated and chlorinated methanes (14 are possible). In this previous work, endpoints representing lipid peroxidation, levels of reactive oxygen species, and cytotoxicity were measured. Using semiempirical molecular orbital theory to calculate descriptors from the neutral species, 2-parameter QSARs were found to correlate with the data with r^2 values ranging from 0.81 to 0.89.[12]

The free radical produced upon acceptance by a HA of an electron from P450 reacts very rapidly with molecular oxygen to form a peroxyl radical,[9,10] although the free radical may also react with other molecules within the cell, e.g., oxidative metabolism of HAs may result in the formation of haloacetic acids.[12] However, to the best of our knowledge, no HA QSARs have been derived from descriptors that distinguish the reactivity of HA metabolites. Specifically, the purpose of this investigation is to ascertain whether improved QSARs result from incorporation of descriptors calculated from HA free radicals and peroxyl radicals.

For a given population of HAs, there may exist two or more neutral species that may form identical free radicals and identical peroxyl radicals. Of the 20 HAs used in this work (see Table 1), four pairs of such HAs exist, i.e.: CH_2CCl_2 and CH_2CBrCl both form $\dot{C}H_2Cl$; $CHCl_3$ and $CHBrCl_2$ both form $\dot{C}HCl_2$; CCl_4 and $CBrCl_3$ both form $\dot{C}Cl_3$; and CH_2ClCH_2Cl and CH_2BrCH_2Cl both form $\dot{C}H_2CH_2Cl$. A complication thereby arises if radical descriptors are to be used to derive a QSAR because two identical sets of descriptors will be mapped to two different toxicities (unless, of course, two compounds that yield identical radicals also have identical toxicities). One approach is to delete from the data set all HAs that yield identical radicals before deriving the QSAR. Another approach is to systematically include only one member from each pair and employ what amounts to a composite QSAR. Both approaches are explored here. Toward this end, we employed the descriptor-mining techniques and multilinear regression analyses as developed within the framework of the CODESSA software program.[13,14]

* Corresponding author phone: (937)255-6671-x3118; fax: (937)255-3377; e-mail: steven.trohalaki@wpafb.af.mil. Corresponding author address: AFRL/MLPJ Building 651, 3005 Hobson Way, Suite 1, Wright-Patterson Air Force Base, OH 45433-7702.

10.1021/ci0342627 CCC: $27.50 © 2004 American Chemical Society
Published on Web 03/02/2004

Table 1. Halogenated Aliphatics[a]

CH_2Cl_2	$CHBr_2Cl$
CH_2BrCl	CBr_2Cl_2
	CH_2Br_2
$CHCl_3$	$CHBr_3$
$CHBrCl_2$	CBr_4
	CH_3CCl_3
CCl_4	$CH_2ClCHCl_2$
$CBrCl_3$	$CHCl_2CHCl_2$
	CH_2BrCH_2Br
CH_2ClCH_2Cl	$CHBr_2CH_2Br$
CH_2BrCH_2Cl	$CHClCCl_2$
	CCl_2CCl_2

[a] The HAs in the second column form unique free radicals and peroxyl radicals (according to the accepted biophysical mechanism), but the HAs in the first column do not, i.e., the four pairs of HAs in the first column will form four unique metabolites.

2. MATERIALS AND METHODS

2.1. Experimental Methods. Details on the preparation of primary hepatocytes, chemical dosing, and biological response assays have been published previously.[11,12,15] The following toxicity endpoints were evaluated: $EC50_{MTT}$ is the effective HA concentration that results in a 50% reduction in mitochondrial function within rat hepatocytes; $EC50_{LDH}$ is the effective HA concentration that results in half of the total observed lactate dehydrogenase leakage from hepatocytes into the cell media, which is an indicator of cell-membrane integrity; $EC20_{SH}$ is the effective HA concentration resulting in a 20% decrease in the level of total protein thiols in hepatocytes; LEC_{LP} is the lowest effective HA concentration resulting in a statistically significant increase in the level of thiobarbituric acid reactive species, which is a measure of the lipid-peroxidation level in hepatocytes; LEC_{ROS} is the lowest effective HA concentration resulting in an increase in the level of reactive oxygen species in hepatocytes; LEC_{CAT} is the lowest effective HA concentration resulting in a decrease in catalase enzyme activity in hepatocytes.

2.2. Computational Methods. All ab initio quantum mechanical calculations were performed with Gaussian98, version A.9.[16] The optimal geometry for the neutral species of each compound was obtained using HF/6-31G** (6-31G** a split-valence basis set with polarization functions on all atoms). Radicals were optimized with unrestricted open-shell calculations, i.e., UHF/6-31G**. Frequencies were calculated for all species using the same level of theory as the optimizations.

Codessa version 2.61[14] was employed for a principal component analysis, to derive correlations between each descriptor and the toxicity data, to derive the QSARs, and to calculate the statistics for the QSARs, including the correleation coefficient, *r*, the Fisher significance parameter, *F*, and *q*, the cross-validated correlation coefficient calculated using a leave-one-out method. CODESSA's best multilinear regression method was used to obtain the best 2- and 3-parameter QSARs, while the heuristic method was used to obtain the best single-parameter QSARs as well as a check on the multilinear regression method's results. Both methods limit the collinearity of the selected descriptors and utilize statistical significance to derive QSARs.

CODESSA categorizes descriptors as constitutional, topological, geometrical, electrostatic, quantum mechanical, and thermodynamic, although both electrostatic and thermodynamic descriptors are obtained from quantum mechanics calculations. Constitutional, topological, and geometrical descriptors were seldom included in the best descriptors and rarely improved QSARs that were derived without them. In this work, we therefore employed only electrostatic, quantum mechanical, and thermodynamic descriptors to derive QSARs. The descriptor pool consisted of about 250 descriptors.

Table 2. In Vitro Toxicity Endpoints

HA	$EC50_{MTT}$ (mM)	$EC50_{LDH}$ (mM)	$EC20_{SH}$ (mM)	LEC_{LP} (mM)	LEC_{ROS} (mM)	LEC_{CAT} (mM)
CH_2Cl_2	32	50	25	25	25	50
CH_2BrCl	127	35	100	10	100	50
$CHCl_3$	1.2	5.4	2.0	2.0	2.0	5.0
$CHBrCl_2$	2.2	3.8	1.0	1.0	5.0	5.0
CCl_4	0.78	0.30	0.17	0.78	0.43	0.55
$CBrCl_3$	0.97	1.0	0.31	0.63	1.3	1.3
CH_2ClCH_2Cl	4.4	50	3.2	9.5	9.5	38
CH_2BrCH_2Cl	0.49	1.4	0.42	0.42	3.4	1.7
CH_3CCl_3	0.23	0.44	0.12	0.25	0.06	0.12
$CH_2ClCHCl_2$	2.54	5.00	1.08	1.08	2.16	4.32
$CHCl_2CHCl_2$	1.73	3.31	1.19	2.38	1.19	4.76
$CHClCCl_2$	1.05	1.50	0.83	1.66	0.83	1.66
CCl_2CCl_2	0.19	0.43	0.50	1.00	1.00	1.00
$CHBr_3$	0.60	2.72	0.29	0.29	0.29	1.15
$CHBr_2Cl$	0.56	1.18	0.59	0.59	0.59	1.18
CBr_2Cl_2	0.30	0.84	0.50	0.50	0.50	0.50
CH_2Br_2	6.40	14.70	5.00	25.00	5.00	25.00
CBr_4	0.18	0.11	0.15	0.10	0.10	0.40
CH_2BrCH_2Br	0.31	1.35	0.07	0.36	0.07	8.63
$CHBr_2CH_2Br$	0.38	1.11	0.31	0.31	0.61	0.61

Table 3. Correlation Matrix for In Vitro Toxicity Endpoints

	$EC50_{MTT}$	$EC50_{LDH}$	$EC20_{SH}$	LEC_{LP}	LEC_{ROS}	LEC_{CAT}
$EC50_{MTT}$	1.0000					
$EC50_{LDH}$	0.9054	1.0000				
$EC20_{SH}$	0.9406	0.9822	1.0000			
LEC_{LP}	0.9624	0.8997	0.9635	1.0000		
LEC_{ROS}	0.9627	0.9266	0.9576	0.9524	1.0000	
LEC_{CAT}	−0.0928	0.0351	−0.1136	−0.2631	−0.2413	1.0000

3. RESULTS AND DISCUSSION

Toxicity endpoints representative of dose−response curves for all six assays are presented in Table 2. In two-thirds of the cases, HAs that form identical radicals display similar toxicities. The correlation matrix for these endpoints, displayed in Table 3, shows that all toxicity endpoints except LEC_{CAT} correlate well with one another. For simplicity, we therefore calculated a composite toxicity metric (CT) for each HA by dividing by 5 the sum of the endpoint values for $EC50_{MTT}$, $EC50_{LDH}$, $EC20_{SH}$, LEC_{LP}, and LEC_{ROS}. QSARs were derived using −log CT. A loadings plot generated from a principal component regression, however, shows that the six toxicity endpoints, the composite toxicity metric, and a composite toxicity calculated from all six endpoints are all closely clustered (see Figure 1).

We found that QSARs of the form that describe well other HA toxicity data[8] correlate poorly with −log CT, i.e., the r^2 value for a QSAR with the polarizability and reciprocal LUMO (lowest unoccupied molecular orbital) energy as descriptors is only 0.624, which indicates that additional mechanisms of toxic response might be significant. Employing the large molecular descriptor space described in the

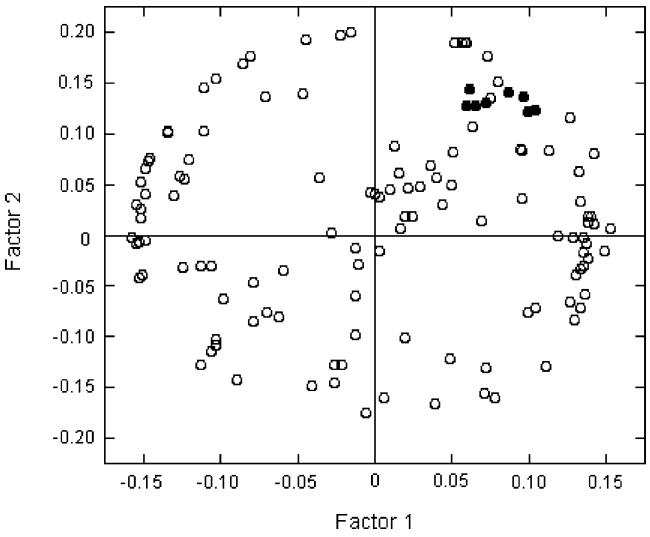

Figure 1. Loadings plot from a principal component analysis. Descriptors correspond to open circles and toxicity endpoints and composite metrics correspond to filled circles.

Methods section, the best 2- and 3-parameter QSARs are

$$-\log \mathrm{CT} = 0.436 - 0.122\ \mathrm{WPSA\text{-}3} + 7.35\ (I_A/N_{atoms})$$

$$N = 20,\ r^2 = 0.834,\ F = 42.8,\ s^2 = 0.0947,\ q^2 = 0.782 \quad (1)$$

$$-\log \mathrm{CT} = -1.81 - 27.9\ \mathrm{FNSA\text{-}3} + (I_A/N_{atoms}) + 2.15\ \mathrm{NRI}_{min,C}$$

$$N = 20,\ r^2 = 0.876,\ F = 37.8,\ s^2 = 0.0750,\ q^2 = 0.773 \quad (2)$$

where I_A is the moment of inertia about the principal axis A, N_{atoms} is the number of atoms in the HA, WPSA-3 the partial positive surface area weighted by the total molecular surface area (TMSA) divided by 1000,[17] $\mathrm{NRI}_{min,C}$ is the minimum nucleophilic reactivity index for a carbon atom,[18] and FNSA-3 is the fraction of the TMSA that has a partial negative charge.[17] We follow Stanton and Jurs' conventions[17] for labeling charged partial surface area (CPSA) descriptors:[17] "-1" is appended when the CPSA is not weighted by charge; "-2" when the CPSA is weighted by the sum total positive or negative charges for the molecule, and "-3" when weighted by the partial atomic charges. We interpret I_A/N_{atoms} as a shape descriptor normalized by the number of atoms in molecule. NRI for a particular atom, which is essentially a Fukui function[19] calculated from the atomic orbital coefficients and energy of the HA's highest occupied molecular orbital, measures the HA's propensity to react with a nucleophile at the particular atom.

Improvements to eqs 1 and 2 should result if the descriptor pool used to derive the QSARs is expanded to include descriptors calculated from the free radicals and peroxyl radicals. However, as noted above, HAs that form an identical radical should not be included in such a data set. One approach is to exclude all HAs that form identical radicals from the data set, which leaves twelve compounds. To have a basis for comparison, we first derived QSARs using only neutral-HA descriptors for the set of twelve HAs. The best 2- and 3-parameter QSARs that result are

$$-\log \mathrm{CT} = -1.35 - 21.8\ \mathrm{FNSA\text{-}3} + 7.70\ (I_A/N_{atoms})$$

$$N = 12,\ r^2 = 0.816,\ F = 20.0,\ s^2 = 0.0574,\ q^2 = 0.189 \quad (3)$$

$$-\log \mathrm{CT} = -2.11 - 32.4\ \mathrm{FNSA\text{-}3} + 6.28\ (I_A/N_{atoms}) + 2.91\ \mathrm{NRI}_{min,C}$$

$$N = 12,\ r^2 = 0.908,\ F = 26.2,\ s^2 = 0.0325,\ q^2 = 0.364 \quad (4)$$

The descriptors in eqs 3 and 4 are identical to those in eqs 1 and 2 except that first descriptors in eqs 1 and 3 are different CPSAs. Equations 3 and 4 predict toxicity values for the eight excluded compounds remarkably well, taking into consideration their poor q^2 values. −Log CT values calculated according to eqs 3 and 4 are plotted vs experimental values in Figures 2 and 3; for the linear least-squares fit for all 20 HAs, the r^2 is 0.812 for Figure 2 and 0.872 for Figure 3. Following the same approach but including both neutral-HA and metabolite descriptors results in QSARs with improved statistics

$$-\log \mathrm{CT} = 1.53 - 0.128\ \mathrm{WPSA\text{-}3} - 2.77\ (\mathrm{FPSA\text{-}2})_{PR}$$

$$N = 12,\ r^2 = 0.894,\ F = 38.0,\ s^2 = 0.0331,\ q^2 = 0.605 \quad (5)$$

$$-\log \mathrm{CT} = 2.50 + 0.898\ (\mu_{hyb})_{PR} - 4.42\ (\mathrm{FPSA\text{-}2})_{PR} + 16.5\ (\mathrm{NAC}_{max,O})_{PR}$$

$$N = 12,\ r^2 = 0.972,\ F = 93.4,\ s^2 = 0.0098,\ q^2 = 0.566 \quad (6)$$

where FPSA-2 is the fraction of the TMSA that has a partial positive charge, μ_{hyb} is the total hybridization component of the molecular dipole,[18] and $\mathrm{NAC}_{max,O}$ is the maximum value of the net atomic charge for an oxygen atom (the PR subscript signifies that the descriptor was calculated from the peroxyl radical). However, the calculated −log CT values for all 20 HAs (and especially for several of the eight excluded HAs) are poor; a least-squares fit of −log CT values calculated for all 20 HAs according to eq 5 as a function of the experimental values has an r^2 value of only 0.677. Similarly, r^2 from a least-squares fit of −log CT values calculated for all 20 HAs according to eq 6 is 0.600. Equation 6 performs worse than eq 5, which is expected because all of the descriptors in eq 6 are calculated from peroxyl radicals (for a pair of HAs that form the same radical but which display different toxicities, identical sets of descriptors will predict identical toxicities), whereas eq 5 has both neutral and radical descriptors. Predictions are poor for all of the 3-parameter QSARs we derived (not shown) using both neutral and radical descriptors calculated from the twelve HAs that form unique radicals. Clearly, information from the eight HAs excluded in the derivations of eqs 5 and 6 is necessary to optimize the QSARs utilizing both neutral and radical HA descriptors.

Another approach is to incorporate into the derivation of QSARs only one member of each of the four pairs of HAs that form identical radicals. Sixteen combinations of HAs that form sixteen unique radicals are possible—the twelve HAs that form unique radicals (Table 1, second column) plus

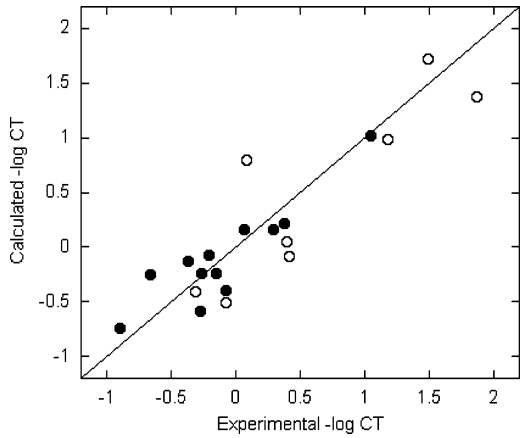

Figure 2. −Log CT values calculated according to eq 1 are plotted versus experimental values. The closed circles correspond to HAs that were used to derive eq 1, while the open circles correspond to the HAs in the first column in Table 1, i.e., those HAs that do not form unique metabolites. The solid line with a slope of unity was drawn to represent a perfect correlation between theory and experiment.

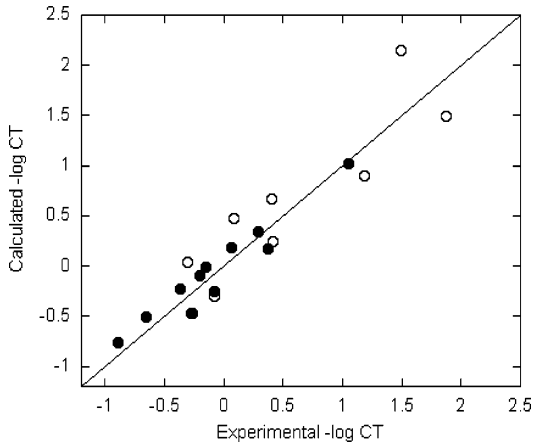

Figure 3. −Log CT values calculated according to eq 2 are plotted versus experimental values. The closed circles correspond to HAs that were used to derive eq 1, while the open circles correspond to the HAs in the first column in Table 1, i.e., those HAs that do not form unique metabolites. The solid line with a slope of unity was drawn to represent a perfect correlation between theory and experiment.

four HAs in each row of Table 4. Using these sixteen sets of HAs, the best 1-, 2-, and 3-parameter QSARs were derived using either neutral-HA descriptors, free-radical descriptors, peroxyl radical descriptors, or all three descriptor types. The statistics for all of these QSARs are presented in Table 5. For the 2- and 3-parameter QSARs, the statistics are generally superior when all three classes of descriptors are included in the descriptor pool (this obviously cannot be true for the 1-parameter QSARs). In some instances, the predicted toxicities for the HAs left out of the data set for a particular QSAR are poor but this is minimized by taking averages of the predicted toxicities. For each of the twenty HAs, we averaged the toxicities calculated from the sixteen QSARs, plotted them versus the experimental values (see Figures 4 and 5), and performed a linear least-squares analysis. For the 2- and 3-parameter QSARs, the r^2 values are 0.881 and 0.914, respectively, which are superior to the r^2 values for the linear least-squares fits of −log CT values calculated according eqs 1 and 2 (derived using only neutral HA descriptors) to the experimental values (see Figures 2 and 3).

Table 4. Sixteen Sets of HAs Used To Derive Eqs 7−22[a]

set 1	CH_2BrCl	$CBrCl_3$	$CHBrCl_2$	CH_2BrCH_2Cl
set 2	CH_2Cl_2	$CHCl_3$	CCl_4	CH_2ClCH_2Cl
set 3	CH_2Cl_2	$CBrCl_3$	$CHBrCl_2$	CH_2BrCH_2Cl
set 4	$CHCl_3$	CH_2BrCl	$CBrCl_3$	CH_2BrCH_2Cl
set 5	CCl_4	CH_2BrCl	$CHBrCl_2$	CH_2BrCH_2Cl
set 6	CH_2ClCH_2Cl	CH_2BrCl	$CBrCl_3$	$CHBrCl_2$
set 7	CH_2Cl_2	$CHCl_3$	$CBrCl_3$	CH_2BrCH_2Cl
set 8	$CHCl_3$	CCl_4	CH_2BrCl	CH_2BrCH_2Cl
set 9	CCl_4	CH_2ClCH_2Cl	CH_2BrCl	$CHBrCl_2$
set 10	$CHCl_3$	CH_2ClCH_2Cl	CH_2BrCl	$CBrCl_3$
set 11	CH_2Cl_2	CH_2ClCH_2Cl	$CBrCl_3$	$CHBrCl_2$
set 12	CH_2Cl_2	CCl_4	$CHBrCl_2$	CH_2BrCH_2Cl
set 13	CH_2Cl_2	$CHCl_3$	CCl_4	CH_2BrCH_2Cl
set 14	CH_2Cl_2	$CHCl_3$	CH_2ClCH_2Cl	$CBrCl_3$
set 15	CH_2Cl_2	CCl_4	CH_2ClCH_2Cl	$CHBrCl_2$
set 16	$CHCl_3$	CCl_4	CH_2ClCH_2Cl	CH_2BrCl

[a] Each set is composed of the twelve HAs in the second column of Table 1 plus four HAs in each row above.

The sixteen 3-parameter QSARs derived from all three classes of descriptors and used to calculate the toxicities in Figure 5 are presented below

Set 1

$$-\log CT = -2.37 - 0.123\ \text{PNSA-3} + 17.1\ (I_A/N_{atoms})_{PR} + 4.04\ (ERI_{min,C})_{PR} \quad (7)$$

Set 2

$$-\log CT = -2.10 - 0.141\ (\text{PNSA-3})_{PR} + 1.43\ I_A + 15.5\ (NRI_{min,C})_{FR} \quad (8)$$

Set 3

$$-\log CT = -1.95 + 0.393\ (\text{WNSA-3})_{PR} + 16.5\ (I_A/N_{atoms})_{PR} + 3.04\ (ERI_{min,C})_{FR} \quad (9)$$

Set 4

$$-\log CT = -2.31 - 0.119\ \text{PNSA-3} + 17.3\ (I_A/N_{atoms})_{PR} + 3.82\ (ERI_{min,C})_{PR} \quad (10)$$

Set 5

$$-\log CT = -2.38 - 0.124\ \text{PNSA-3} + 17.8\ (I_A/N_{atoms})_{PR} + 3.65\ (ERI_{min,C})_{PR} \quad (11)$$

Set 6

$$-\log CT = -2.20 - 0.117\ \text{PNSA-3} + 1.60\ I_A + 4.14\ (ERI_{min,C})_{PR} \quad (12)$$

Set 7

$$-\log CT = -1.94 - 0.392\ (\text{WNSA-3})_{PR} + 16.5\ (I_A/N_{atoms})_{PR} + 3.02\ (ERI_{min,C})_{FR} \quad (13)$$

Set 8

$$-\log CT = -2.33 - 0.121\ \text{PNSA-3} + 18.0\ (I_A/N_{atoms})_{PR} + 3.44\ (ERI_{min,C})_{PR} \quad (14)$$

Set 9

$$-\log CT = -2.78 + 0.139\ \text{DPSA-3} + 1.38\ I_A + 27.3\ (NRI_{min,C})_{FR} \quad (15)$$

Table 5. Statistics for the QSARs Derived for Sets 1−16

	neutrals				free radical				peroxyl radical				all descriptors			
set no.	r^2	q^2	F	s^2	r^2	q^2	F	s^2	r^2	q^2	F	s^2	r^2	q^2	F	s^2
(a) 1-Parameter QSARs																
1	0.646	0.479	25.6	0.165	0.758	0.703	43.8	0.113	0.684	0.544	30.4	0.147	0.758	0.703	43.7	0.113
2	0.746	0.699	41.1	0.121	0.643	0.556	25.2	0.169	0.643	0.555	25.2	0.169	0.746	0.699	41.1	0.120
3	0.633	0.533	24.2	0.139	0.706	0.655	33.6	0.111	0.667	0.513	28.0	0.126	0.705	0.655	33.5	0.112
4	0.529	0.446	15.7	0.110	0.592	0.497	20.3	0.0949	0.599	0.450	20.9	0.0933	0.599	0.450	20.9	0.0933
5	0.687	0.526	30.7	0.146	0.768	0.711	46.4	0.108	0.714	0.575	35.0	0.575	0.768	0.711	46.3	0.108
6	0.757	0.676	43.6	0.133	0.697	0.643	32.2	0.166	0.665	0.555	27.8	0.184	0.757	0.676	43.6	0.133
7	0.635	0.535	24.4	0.138	0.707	0.656	33.8	0.111	0.667	0.512	28.0	0.126	0.707	0.656	33.7	0.111
8	0.654	0.491	26.5	0.164	0.757	0.703	43.6	0.115	0.693	0.556	31.6	0.145	0.757	0.703	43.6	0.703
9	0.760	0.681	44.3	0.134	0.698	0.644	32.4	0.169	0.672	0.559	28.6	0.184	0.760	0.681	44.3	0.134
10	0.621	0.501	22.9	0.146	0.669	0.572	28.3	0.572	0.672	0.525	28.7	0.126	0.692	0.638	31.5	0.118
11	0.739	0.690	39.5	0.122	0.639	0.550	24.8	0.168	0.633	0.544	24.2	0.544	0.739	0.690	39.5	0.122
12	0.639	0.542	24.8	0.139	0.704	0.653	33.3	0.114	0.664	0.577	27.6	0.130	0.704	0.653	33.3	0.114
13	0.642	0.545	25.1	0.138	0.706	0.655	33.5	0.114	0.666	0.579	27.9	0.129	0.705	0.655	33.5	0.114
14	0.704	0.651	33.4	0.651	0.641	0.554	25.0	0.168	0.637	0.551	24.5	0.170	0.705	0.651	33.4	0.138
15	0.744	0.697	40.6	0.697	0.636	0.566	24.4	0.173	0.641	0.553	25.1	0.170	0.744	0.697	40.6	0.122
16	0.748	0.665	41.7	0.140	0.699	0.645	32.5	0.168	0.673	0.561	28.8	0.183	0.762	0.683	44.7	0.133
av	0.680	0.585	31.5	0.205	0.689	0.623	32.1	0.165	0.662	0.544	27.6	0.200	0.726	0.663	38.0	0.156
(b) Best 2-Parameter QSARs																
1	0.859	0.769	39.7	0.0706	0.910	0.880	66.1	0.0449	0.872	0.826	44.4	0.826	0.913	0.874	68.1	0.0436
2	0.894	0.848	54.7	0.0542	0.806	0.719	26.9	0.0993	0.819	0.775	29.5	0.0922	0.899	0.856	58.1	0.0514
3	0.861	0.820	40.1	0.0569	0.883	0.839	49.0	0.0477	0.866	0.804	41.8	0.0548	0.892	0.855	53.5	0.0442
4	0.787	0.658	24.0	0.0533	0.865	0.812	41.8	0.0337	0.830	0.764	31.8	0.763	0.886	0.820	50.5	0.0286
5	0.835	0.725	33.0	0.083	0.911	0.881	66.9	0.0444	0.857	0.794	38.9	0.0716	0.913	0.883	68.5	0.0434
6	0.879	0.811	47.1	0.0715	0.825	0.762	30.5	0.104	0.820	0.751	29.7	0.106	0.908	0.858	63.9	0.0544
7	0.868	0.828	42.6	0.0539	0.884	0.840	49.4	0.0473	0.865	0.803	41.7	0.0549	0.894	0.839	54.9	0.0431
8	0.838	0.736	33.7	0.0824	0.913	0.883	68.0	0.0444	0.856	0.787	38.7	0.787	0.919	0.889	73.2	0.0415
9	0.891	0.839	53.2	0.0656	0.839	0.782	33.8	0.0971	0.837	0.774	33.4	0.0979	0.903	0.859	60.8	0.0581
10	0.862	0.821	40.7	0.0571	0.906	0.867	62.8	0.0389	0.868	0.810	42.8	0.0547	0.906	0.867	62.8	0.0389
11	0.877	0.845	46.3	0.0617	0.803	0.718	26.6	0.0984	0.812	0.718	28.1	0.0939	0.907	0.863	63.7	0.0463
12	0.826	0.768	30.9	0.0722	0.897	0.858	56.6	0.0429	0.867	0.795	42.5	0.0552	0.908	0.866	64.5	0.0381
13	0.839	0.786	33.8	0.0670	0.898	0.859	57.2	0.0424	0.868	0.795	42.6	0.0550	0.915	0.883	69.6	0.0354
14	0.878	0.847	46.6	0.0616	0.800	0.711	26.0	0.101	0.811	0.764	27.9	0.0950	0.908	0.866	64.1	0.0463
15	0.862	0.825	40.8	0.0703	0.807	0.721	27.1	0.0989	0.819	0.773	29.3	0.0927	0.899	0.855	57.9	0.0516
16	0.907	0.857	63.2	0.0561	0.839	0.783	34.0	0.0965	0.838	0.775	33.6	0.0974	0.907	0.857	63.2	0.0561
av	0.860	0.799	41.9	0.0648	0.862	0.807	45.2	0.0676	0.844	0.782	36.0	0.212	0.905	0.862	62.3	0.0450
(c) Best 3-Parameter QSARs																
1	0.927	0.848	50.6	0.0397	0.968	0.944	119	0.0176	0.971	0.947	133	0.0159	0.982	0.970	217	0.0098
2	0.940	0.877	63.1	0.0330	0.915	0.852	43.1	0.0470	0.912	0.863	41.5	0.0487	0.959	0.925	93.2	0.0228
3	0.910	error	40.4	0.0398	0.963	0.944	103	0.0165	0.968	0.943	119	0.0143	0.969	0.948	123	0.0139
4	0.885	0.814	30.6	0.0313	0.939	0.909	61.4	0.0166	0.947	0.868	71.5	0.0144	0.944	0.895	67.0	0.0153
5	0.893	0.765	33.5	0.0579	0.960	0.931	96.3	0.0216	0.963	0.935	105	0.935	0.971	0.950	134	0.0158
6	0.950	0.909	75.2	0.0323	0.935	0.891	57.2	0.0419	0.931	0.891	53.9	0.0441	0.979	0.966	183	0.0137
7	0.910	0.846	40.3	0.0398	0.962	0.944	102	0.0166	0.967	0.942	118	0.0145	0.968	0.948	122	0.0139
8	0.897	0.780	35.0	0.0567	0.961	0.919	98.9	0.0214	0.962	0.931	102	0.931	0.971	0.947	133	0.0161
9	0.942	0.892	65.1	0.0377	0.921	0.857	46.5	0.0516	0.919	0.872	45.5	0.0527	0.976	0.959	162	0.0157
10	0.909	0.848	39.7	0.0411	0.964	0.944	108	0.0160	0.967	0.942	118	0.0148	0.968	0.944	119	0.0146
11	0.937	error	59.0	0.0344	0.915	0.849	43.1	0.0460	0.925	0.881	49.1	0.0409	0.961	0.920	99.7	0.0209
12	0.907	0.848	38.8	0.0421	0.962	0.940	102	0.0170	0.960	0.931	95.1	0.0182	0.964	0.944	107	0.0163
13	0.914	0.859	42.5	0.0387	0.962	0.940	102	0.0170	0.959	0.931	94.6	0.0183	0.964	0.943	106	0.0164
14	0.932	0.873	54.8	0.0370	0.941	0.897	63.5	0.0323	0.923	0.878	47.6	0.0422	0.962	0.921	100	0.0209
15	0.927	0.877	50.7	0.0405	0.916	0.853	43.5	0.0467	0.913	0.864	41.9	0.0483	0.959	0.924	92.4	0.0230
16	0.948	0.906	72.6	0.0340	0.920	0.856	46.0	0.0521	0.918	0.871	44.9	0.0532	0.972	0.948	141	0.0180
av	0.921	0.746	49.5	0.0398	0.944	0.904	77.2	0.0299	0.944	0.906	80.0	0.144	0.967	0.941	125	0.0166

Set 10

$$-\log \mathrm{CT} = -1.97 - 0.120\,(\mathrm{PNSA\text{-}3})_{\mathrm{PR}} + 1.59\,I_{\mathrm{A}} + 3.52\,(\mathrm{ERI}_{\mathrm{min,C}})_{\mathrm{PR}} \quad (16)$$

Set 11

$$-\log \mathrm{CT} = -1.82 - 0.115\,(\mathrm{PNSA\text{-}3})_{\mathrm{PR}} + 1.41\,I_{\mathrm{A}} + 2.53\,(\mathrm{ERI}_{\mathrm{min,C}})_{\mathrm{FR}} \quad (17)$$

Set 12

$$-\log \mathrm{CT} = -2.20 + 0.127\,(\mathrm{DPSA\text{-}3})_{\mathrm{PR}} + 0.438\,(I_{\mathrm{A}}/N_{\mathrm{atoms}})_{\mathrm{FR}} + 19.9\,(\mathrm{NRI}_{\mathrm{min,C}})_{\mathrm{FR}} \quad (18)$$

Set 13

$$-\log \mathrm{CT} = -0.877 + 0.335\,(\mathrm{WPSA\text{-}2})_{\mathrm{FR}} + 17.4\,(I_{\mathrm{A}}/N_{\mathrm{atoms}})_{\mathrm{PR}} + 1.41\,(\mathrm{NAC}_{\mathrm{min,C}})_{\mathrm{FR}} \quad (19)$$

Set 14

$$-\log \mathrm{CT} = -1.82 - 0.114\,(\mathrm{PNSA\text{-}3})_{\mathrm{PR}} + 1.41\,I_{\mathrm{A}} + 2.51\,(\mathrm{ERI}_{\mathrm{min,C}})_{\mathrm{FR}} \quad (20)$$

Set 15

$$-\log \mathrm{CT} = -2.11 - 0.114\,(\mathrm{PNSA\text{-}3})_{\mathrm{PR}} + 1.42\,I_{\mathrm{A}} + 15.6\,(\mathrm{NRI}_{\mathrm{min,C}})_{\mathrm{FR}} \quad (21)$$

Table 6. Descriptor Values

HA	DPSA-3 Å^2	$(DPSA\text{-}3)_{PR}$ Å^2	PNSA-3 Å^2	$(PNSA\text{-}3)_{PR}$ Å^2	$(WNSA\text{-}3)_{PR}$ Å^4	$(WPSA\text{-}2)_{FR}$ Å^4	WPSA-3 Å^4	$(FNSA\text{-}2)_{PR}$ Å^4	FNSA-3 Å^2
CH_2Cl_2	17.0385	14.5569	−13.6430	−10.9726	−2.5423	2.3144	0.7228	0.0477	−0.0641
CH_2BrCl	15.3503	13.1967	−13.2932	−11.6945	−3.0251	2.2497	0.4805	0.0137	−0.0569
$CHCl_3$	9.8204	8.3617	−9.8204	−8.3617	−2.2745	0.7735	0.0000	0.0000	−0.0392
$CHBrCl_2$	19.0097	16.7723	−15.2022	−12.5100	−3.3238	2.1697	0.9065	0.0898	−0.0639
CCl_4	14.3567	11.8798	−12.1612	−9.4313	−2.5317	3.4133	0.5683	0.0658	−0.0470
$CBrCl_3$	21.0299	18.6328	−17.6702	−15.0821	−4.3128	5.0048	0.8873	0.0674	−0.0669
CH_2ClCH_2Cl	21.2150	19.5371	−18.8042	−17.1267	−5.3081	4.1599	0.6761	0.0348	−0.0670
CH_2BrCH_2Cl	18.1942	16.7806	−15.4938	−13.2315	−3.6782	3.2062	0.6742	0.0435	−0.0621
CH_3CCl_3	17.5194	15.3117	−15.6259	−13.5198	−4.0474	1.8176	0.5331	0.0120	−0.0555
$CH_2ClCHCl_2$	11.7076	10.0775	−10.3382	−8.8184	−2.3557	1.4165	0.3472	0.0106	−0.0408
$CHCl_2CHCl_2$	14.6686	14.5569	−11.8094	−10.9726	−2.5423	2.3144	0.6104	0.0477	−0.0553
$CHClCCl_2$	9.1129	8.3617	−9.1129	−8.3617	−2.2745	0.7735	0.0000	0.0000	−0.0349
CCl_2CCl_2	14.2695	13.1967	−12.4437	−11.6945	−3.0251	2.2497	0.4514	0.0137	−0.0503
$CHBr_3$	13.0055	11.4807	−11.3319	−10.1615	−2.6362	1.7490	0.4186	0.0118	−0.0453
$CHBr_2Cl$	17.3103	16.7723	−13.9112	−12.5100	−3.3238	2.1697	0.8460	0.0898	−0.0559
CBr_2Cl_2	8.6945	7.2867	−8.6945	−7.2867	−2.0464	0.6374	0.0000	0.0000	−0.0321
CH_2Br_2	12.7454	12.4238	−10.2987	−9.3886	−2.2233	1.6115	0.5352	0.0388	−0.0471
CBr_4	7.4709	5.6532	−7.4709	−5.6532	−1.6352	0.4213	0.0000	0.0000	−0.0265
CH_2BrCH_2Br	14.8966	14.8468	−11.5860	−10.9643	−3.0038	1.7668	0.8538	0.0769	−0.0449
$CHBr_2CH_2Br$	16.5027	14.9133	−13.9807	−11.9213	−3.4867	3.7141	0.7188	0.0533	−0.0491

HA	$(NAC_{min,C})_{PR}$ au	$(ERI_{min,C})_{FR}$ au	$(ERI_{min,C})_{PR}$ au	$NRI_{min,C}$ au	$(NRI_{min,C})_{FR}$ au	I_A au	$(I_A)_{FR}$ au	$(I_A)_{PR}$ au	$(\mu_{hyb})_{PR}$ au
CH_2Cl_2	−0.0558	0.2835	0.2581	3.222×10^{-3}	0.0414	1.0902	9.2005	0.5933	−0.9381
CH_2BrCl	0.0730	0.2753	0.2423	0.0000	0.0399	0.1098	1.5483	0.1247	−1.2910
$CHCl_3$	0.1544	0.2642	0.2391	0.0000	0.0401	0.0579	0.1120	0.0663	−1.7865
$CHBrCl_2$	−0.4342	0.0103	0.0184	3.554×10^{-3}	8.319×10^{-4}	0.9853	1.1091	0.3640	−1.1356
CCl_4	−0.6834	5.169×10^{-3}	6.11×10^{-3}	0.0000	1.940×10^{-3}	0.0791	0.2181	0.0827	−0.8144
$CBrCl_3$	−0.4506	0.1285	0.1022	1.381×10^{-3}	6.576×10^{-4}	0.1173	0.4252	0.1044	−1.1426
CH_2ClCH_2Cl	−0.2841	8.409×10^{-3}	0.0335	8.178×10^{-5}	6.178×10^{-4}	0.0561	0.1127	0.0614	−0.7974
CH_2BrCH_2Cl	−0.2963	0.0329	0.0362	0.0144	4.939×10^{-3}	0.1298	0.4811	0.1566	−0.8780
CH_3CCl_3	−0.1596	0.0326	0.0368	0.0151	3.949×10^{-3}	0.0578	0.1259	0.0642	−1.5107
$CH_2ClCHCl_2$	−0.0704	0.1668	0.1478	0.0000	0.0322	0.0415	1.2365	0.0859	−1.6134
$CHCl_2CHCl_2$	−0.0558	0.2835	0.2581	4.054×10^{-4}	0.0414	0.9893	9.2006	0.5933	−0.9381
$CHClCCl_2$	0.1544	0.2642	0.2391	1.010×10^{-4}	0.0401	0.0579	0.1120	0.0663	−1.7865
CCl_2CCl_2	0.0730	0.2753	0.2423	1.200×10^{-4}	0.0399	0.1096	1.5482	0.1247	−1.2910
$CHBr_3$	5.780×10^{-3}	0.1983	0.1740	2.462×10^{-4}	0.0349	0.0833	1.4000	0.0967	−1.4390
$CHBr_2Cl$	−0.4342	0.0103	0.0184	6.236×10^{-5}	8.052×10^{-4}	0.9657	1.1092	0.3640	−1.1356
CBr_2Cl_2	0.0797	0.2023	0.1714	1.993×10^{-4}	0.0369	0.0456	0.1121	0.0656	−1.9681
CH_2Br_2	−0.1173	0.1833	0.1617	5.286×10^{-4}	0.0322	0.8760	9.1620	0.5405	−0.9628
CBr_4	−0.0734	0.1540	0.1417	0.0000	0.0332	0.0214	0.0416	0.0310	−2.3173
CH_2BrCH_2Br	−0.5023	9.084×10^{-3}	7.690×10^{-3}	7.400×10^{-4}	9.011×10^{-4}	0.2517	1.0460	0.3124	−0.9808
$CHBr_2CH_2Br$	−0.5257	0.0730	0.0651	1.130×10^{-4}	7.384×10^{-4}	0.0492	0.3057	0.0693	−0.9309

Set 16

$$-\log CT = -2.64 - 0.131\,(DPSA\text{-}3)_{PR} + 1.42\,I_A + 25.2\,(NRI_{min,C})_{FR} \quad (22)$$

where WNSA-3 is the negative-charge counterpart to WPSA-3, DPSA-3 is the difference between PNSA-3 and its positive-charge counterpart, PPSA-3, ERI for a particular atom, which, like NRI, is essentially a Fukui function[19] (the FR subscript signifies that the descriptor was calculated from the free radical). ERI measures the HA's propensity to react with an electrophile at the particular atom and is calculated from the atomic orbital coefficients and energy of the HA's LUMO. Although they do not consist of identical descriptors, eqs 7−22 are similar in that they all contain a CPSA descriptor, a shape descriptor, and a reactivity index. Relevant statistics for eqs 7−22 appear in Table 5c. Descriptor values are given in Table 6.

4. CONCLUSIONS

In summary, we have presented toxicity endpoints from six in vitro assays in which twenty volatile HAs were

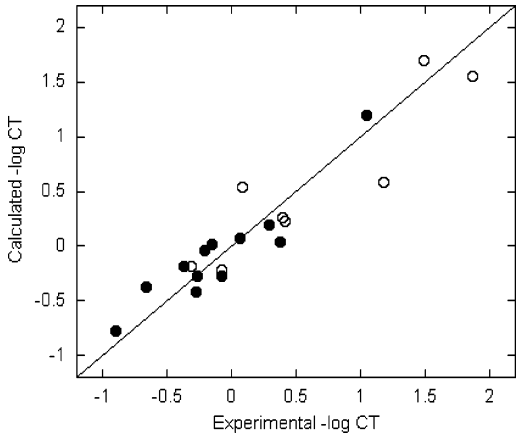

Figure 4. −Log CT values calculated as averages of the values computed from sixteen 2-parameter QSARs are plotted versus experimental values. The closed circles correspond to the twelve HAs that were common to each of the sixteen QSARs. The open circles correspond to the HAs in the first column in Table 1, i.e., those HAs that do not form unique metabolites. The solid line with a slope of unity was drawn to represent a perfect correlation between theory and experiment.

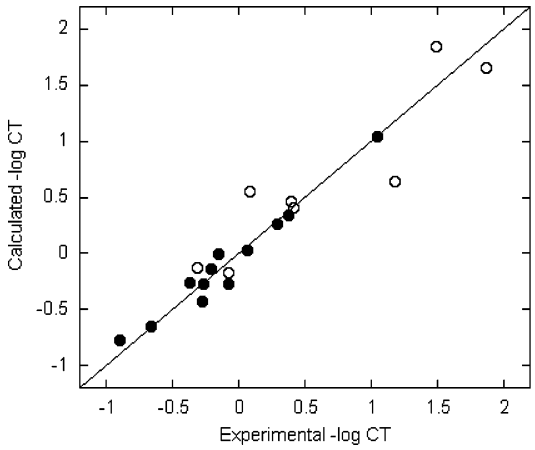

Figure 5. −Log CT values calculated as averages of the values computed from eqs 7−22 are plotted versus experimental values. The closed circles correspond to the twelve HAs that were common to each of the sixteen QSARs. The open circles correspond to the HAs in the first column in Table 1, i.e., those HAs that do not form unique metabolites. The solid line with a slope of unity was drawn to represent a perfect correlation between theory and experiment.

exposed to rat hepatocytes. Four pairs of HAs are expected to form the same metabolites, but the toxicity data are similar for only about two-thirds of the cases. Although good 2- and 3-parameter QSARs were derived using neutral-HA descriptors, improved QSARs generally result from incorporation of metabolite descriptors. However, to properly utilize metabolite descriptors, multiple QSARs, each of which utilizes a set of HAs that form unique metabolites, must be derived and toxicity values calculated therefrom must be averaged. We do not expressly address QSAR validation here primarily because of the small data set employed. However, our protocol can be easily used in conjunction with recently proposed validation strategies[20] for QSAR studies of larger populations of compounds.

ACKNOWLEDGMENT

We acknowledge funding from the U.S. Air Force Office of Scientific Research.

REFERENCES AND NOTES

(1) Fishbein, L. Halogenated aliphatic hydrocarbons: uses and environmental occurrence. In *Environmental Carcinogens − Selected Methods of Analysis, Some Volatile Halogenated Hydrocarbons and Environmental Carcinogens*; IARC Publication No. 68; Fishbein, L., O'Neill, I. K., Eds; International Agency for Research on Cancer: Lyon, 1985; Vol. 7, pp 47-67.

(2) MacFarland, H. N. Toxicology of Solvents. *Am. Ind. Hyg. Assoc.* **1986**, *47*, 704−707.

(3) IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. *Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs*; International Agency for Research on Cancer: Lyon, 1987; Vol. 1−42, Supplement 7.

(4) Ikeda, M. Public Health Problems of Organic Solvents. *Toxicol. Lett.* **1992**, *64/65*, 191−201.

(5) World Health Organization. *Organic Solvents and the Central Nervous System*; Joint WHO/Nordic Council Ministers Working Group: Copenhagen, WHO, 1985.

(6) Henschler, D. Specific Ccovalent Binding and Genotoxicity. In *Environmental Carcinogens − Selected Methods of Analysis, Some Volatile Halogenated Hydrocarbons and Environmental Carcinogens;* IARC Publication No. 68; Fishbein, L., O'Neill, I. K., Eds.; International Agency for Research on Cancer: Lyon, 1985; Vol. 7, pp 21−44.

(7) Trohalaki, S.; Gifford, E.; Pachter, R. Improved QSARs for Predictive Toxicology of Halogenated Alkanes. *Comput. Chem.* **2000**, *24*, 421−427.

(8) Trohalaki, S.; Pachter, R. Quantum Descriptors for Predictive Toxicology of Halogenated Aliphatic Hydrocarbons. *SAR QSAR Environ. Res.* **2003**, *14(2)*, 131−143.

(9) Slater, T. Free Radicals as Reactive Intermediates in Tissue Injury. In *Biological Reactive Intermediates, II*; Snyder, R., Parke, D. V., Eds.; Plenum Press: New York, 1982; pp 575−589.

(10) Waller, C.; Evans, M.; McKinney, J. Modeling the Cytochrome P450-Mediated Metabolism of Chlorinated Volatile Organic Compounds. *Drug Metab. Dispos.* **1996**, *24*, 203−210.

(11) Geiss, K. T.; Frazier, J. M. A Novel In Vitro System for Exposures of Cell Cultures to Volatile Chemicals. *Toxicologist* **2000**, *54*, 377.

(12) Geiss, K. T.; Frazier, J. M. QSAR Modeling of Oxidative Stress In Vitro Following Hepatocyte Exposures to Halogenated Methanes. *Toxicol. in Vitro* **2001**, *15*, 557−563.

(13) Karelson, M.; Maran, U.; Wang, Y.; Katritzky, A. R. QSAR and QSPR Models Devised Using Large Molecular Descriptor Spaces. A Review of CODESSA Applications. *Collect. Czech. Chem. Commun.* **1999**, *64*, 1551−1571.

(14) CODESSA, Version 2.061, Semichem, Inc., Shawnee, KS.

(15) Trohalaki, S.; Zellmer, R. J.; Pachter, R.; Hussain, S. M.; Frazier, J. M. Risk Assessment of High-Energy Chemicals by In Vitro Toxicity Screening and Quantitative Structure−Activity Relationships. *Toxicol. Sci*. **2002**, *68*, 498−507.

(16) Gaussian 98, Revision A.6. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 1998.

(17) Stanton, J. T.; Jurs, P. C. Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure−Property Relationship Studies. *Anal. Chem.* **1990**, *62*, 2, 2323−2329.

(18) Karelson, M.; Katritzky, A. R. 1995 *CODESSA Reference Manual*; Semichem and the University of Florida: pp 10−18.

(19) Pires, J. M.; Floriano, W. B.; Gaudio, A. C. Extension Of The Frontier Reactivity Indices to Groups of Atoms and Application to Quantitative Structure−Activity Relationship Studies. *J. Mol. Struct.* **1997**, *389*, 159−167.

(20) Tropsha, A.; Grammatica, P.; Gombar, V. K. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Intepretation of QSPR Models. *QSAR Comb. Sci*. **2003**, *22*, 69−77.

CI0342627