Membrane-Interaction QSAR Analysis: Application to the Estimation of Eye Irritation by Organic Compounds

Amit S. Kulkarni[1] and A. J. Hopfinger[1,2]

Received October 26, 1998; accepted April 18, 1999

Purpose. The purpose of this study was to explore a potential mechanism of eye irritation, and to construct a corresponding general quantitative structure-activity relationship (QSAR) model, in terms of diversity of irritant chemical structure, based on the Draize eye irritation ECETOC data set.

Methods. Molecular dynamic simulation (MDS) was used to generate intermolecular membrane-solute interaction properties. These intermolecular properties were combined with intramolecular physicochemical properties and features of the solute (irritant) to construct QSAR models using multi-dimensional linear regression and the Genetic Function Approximation (GFA) algorithm.

Results. Significant QSAR models for estimating eye irritation potential were constructed in which solute aqueous solvation free energy and solute-membrane interaction energies are the principle correlation descriptors. These physicochemical descriptors were selected from a trial set of 95 descriptors for 18 structurally diverse compounds fully representative of the ECETOC set of 38 compounds.

Conclusions. Combining intermolecular solute-membrane interaction descriptors with intramolecular solute descriptors yields statistically significant eye irritation QSAR models. The resultant QSAR models support an eye irritation mechanism of the action in which increased aqueous solubility of the irritant and its strength of binding to the membrane both increase eye irritation.

KEY WORDS: molecular dynamics simulations; molar adjusted eye scores; partial least-squares regression; genetic function approximation; quantitative structure-activity relationship.

INTRODUCTION

Eye irritation potency has traditionally being scored using the Draize rabbit eye test (1). Other experimental approaches, such as the use of histology, modern methods of pathology and cell culture, have also been successfully employed in assessing ocular toxicity. The computational study reported here explores the possibility of estimating eye irritation potency for chemicals of reasonably diverse structures without using animals. A subset of the "standard" data set for eye irritation potency established by the European Center for Ecotoxicology & Toxicology of Chemicals (ECETOC) (2), see Table I, was used in this study. Recently Abraham and coworkers (3) have also suggested a predictive eye irritation model from the ECETOC data set based upon concentration corrections for irritant liquid vapor pressure, and descriptors reflecting among other features acidity and basicity of the irritant.

The quantitative structure-activity relationship (QSAR) paradigm, using *intramolecular* physicochemical descriptors, has been used to predict several toxicological endpoints (4). We first tried this approach by correlating intramolecular physicochemical properties of the molecules in the ECETOC data set with the corresponding reported eye irritation scores in the hope of discovering a significant QSAR model. Like all other QSAR studies on this data set only using intramolecular descriptors, see references 3 and 5 and discussions and references therein, our analysis was not successful in generating a significant statistical QSAR model. It subsequently occurred to us that "traditional" QSAR modeling does, in fact, only consider intramolecular features of the molecules in training set such as lipophilicity, dipole moment, molecular volume etc. In principle, progress might be made in the QSAR analysis of any chemically diverse data set, such as the ECETOC eye irritation data set, if the biochemical mechanism of action and/or "receptor" were known and could be included in constructing the QSARs.

In the case of eye irritation, uptake and diffusion into the keratocytes comprising the outer seven or so layers of the corneal epithelium of the eye is a significant event. At the corresponding molecular level eye irritation may involve transfer to, and penetration into, corresponding cellular membranes of the keratocytes. We have thus hypothesized that interactions of test organic molecules with a cell membrane are, at least, partly responsible for eye irritation. Moreover, the phospholipid regions of a membrane bilayer of the cell might contain "general binding sites" at least, in part, responsible for eye irritation. Thus, we decided to include simulations of the uptake and interaction of the compounds from ECETOC subset (the solutes) with model monolayer membranes as part of our QSAR analysis. The estimated membrane-solute interaction properties from the simulations were added to the intramolecular solute physicochemical property descriptors to provide an extended set of trial descriptors for building eye irritation QSAR models. This overall methodology is called *membrane-interaction QSAR analysis*, (MI-QSAR analysis).

MATERIALS AND METHODS

Eye Irritation Potency and the Training Set

Eye irritation potency dependent variables are the molar adjusted eye scores from the Draize Rabbit Eye Irritation Test. These scores were determined as follows: The molarity of the solution was calculated using molarity = (density × 1000)/ relative molecular mass. Density values were obtained from a standard source (6). Molar adjusted eye scores were then calculated as the raw eye irritation scores divided by the molarity of the solution. The values of molar adjusted eye scores (MES) for the compounds in data set are given in Table I. There is a considerable concern that the MES values lack sufficient quality and reproducibility to justify usage in constructing QSAR models. An analysis of the raw MES data, involving the use of multiple experiments (animals) for a single test compound, reveals that the standard deviations of fit for the compounds that are "safe" (low eye irritation) have highly reproducible

[1] Laboratory of Molecular Modeling and Design (M/C 781), College of Pharmacy, The University of Illinois at Chicago, 833 South Wood Street, Chicago, Illinois 60612-7231.

[2] To whom correspondence should be addressed. (e-mail: hopfingr@uic.edu)

0724-8741/99/0800-1245$16.00/0

Table I. The ECETOC Draize Eye Irritation Training and Test Subset Used in this Study Along with MI-QSAR (Eqs. 11–13) Descriptor Values for Each Solute Molecule

Solutes	F(H2O)	E(chg + vdw)	E(chg)	Kappa-3-AM	Molar adjusted eye score (MES)
Training set					
Hydrocarbons					
3 Methyl hexane	2.75	−17.31	−13.22	3.84	0.10
2 Methyl pentane	2.54	−16.22	−12.15	5.33	0.26
Methylcyclopentane	1.36	−7	−10.69	0.98	0.41
*1,9−Decadiene	1.64	−14.36	−13.16	8.63	0.37
*Dodecane	3.62	−24.48	−23.65	11.11	0.45
1,5-hexadiene	0.83	−9.04	−10.49	4.88	0.55
Aromatic					
4-fluoroaniline	−13.68	−4.04	−10.54	1.29	6.62
*Xylene	−0.85	−12.83	−12.13	1.35	1.10
Toluene	−0.93	−4.58	−5.38	1.04	0.96
Styrene	−1.53	−9.65	−9.35	1.21	0.77
1-Methylpropylbenzene	−0.26	−5.36	−5	1.86	0.31
Ketones					
Methyl amyl ketone	1.75	−9.47	−13.12	6.88	2.26
Methyl isobutyl ketone	1.8	−3.92	−3.41	2.37	0.59
Acetone	0.94	−2.06	−1.46	0.00	4.83
Alcohols					
n-Butanol	−7.45	−18.37	−8.64	3.96	5.47
Isopropanol	−7.51	−1.69	−5.63	0.00	2.34
Propylene glycol	−16.95	−15.04	−7.91	3.92	0.10
Butyl cellsolve	−11.01	−18.3	−18.45	7.12	8.99
*Cyclohexanol	−8.08	−10.74	−11.5	1.48	8.29
Acetates					
Ethyl acetate	−2.82	6.59	−11.17	5.01	1.47
Methyl acetate	−3.02	−15.79	−8.97	3.63	3.14
Methyl trimethyl acetate	−2.16	−3.31	−4.88	2.52	0.36
*Ethyl-2-methyl acetoacetate	−2.92	−5.59	1.18	6.30	2.55
Test set					
1,9-Decadiene	1.64	−14.36	−13.16	8.63	0.37 0.94 (p)
Dodecane	3.62	−24.48	−23.65	11.11	0.45 1.24 (p)
Xylene	−0.85	−12.83	−12.13	1.35	1.10 1.89 (p)
Ethyl-2-methyl acetoacetate	−2.92	−5.59	1.18	6.30	2.55 1.98 (p)
Cyclohexanol	−8.08	−10.74	−11.5	1.48	8.29 4.98 (p)

* Solute molecules used in test set, (p) predicted MES values using eq. 11

eye irritation measures (small standard deviations) while strong eye irritants have large standard deviations in their eye irritation measures. Thus, highly irritating compounds are reproducibly measured to be irritants, but the extent of high irritation varies among animals. This translates into QSAR models that are reliable except, perhaps, in predicting how highly irritating a predicted high eye irritant might be. The 18 compounds in Table I are a subset of the 38 compounds in the ECETOC eye irritation data set for which MES values can be obtained. The 18 compounds in our training set were chosen to test the simplifying assumption that if a model can explain the behavior of any *representative* subset of compounds of a data set, the model can explain the behavior of *all* compounds of the data set and "equivalent" compounds *outside* the data set. The 18 training set compounds and the five test compounds (not in the training set) of Table I were selected according to the following criteria to achieve a composite representative subset; a) Span the entire range in eye irritation potency for the composite ECETOC data set. b) Include representative chemical structures from each of the analog subsets composing the composite ECETOC set. c) Span the range of eye irritation potency within each analog subset. "Charged" molecules were not included in the training set because it is not clear if they are actually charged when at, or in, the membrane. Both neutral and charged forms of an ionizable compound could be considered in this approach, but was not done in this initial application of MI-QSAR analysis. The goal of this work was to apply and evaluate the MI-QSAR method on a data set, which would permit us to analyze and explore the results and findings of the analysis.

Building Solute Molecules and the DMPC Monolayer

The solute molecules, see Table I, were built using the Chemlab-II (7) molecular modeling package. A single dimyristoylphosphatidylcholine (DMPC) molecule was built in HyperChem (8) using available crystal structure data (9). Partial atomic charges were assigned to solute molecules, and the DMPC molecule, using semi-empirical molecular orbital calculations. The AM1 Hamaltonian in Mopac 6.0 (10) was used for the partial atomic charge estimation.

The DMPC molecule was selected as the model phospholipid in this study. The structure of a DMPC molecule is shown

in Fig 1. An assembly of 16 DMPC molecules (4*4*1) in (x, y, z) directions, respectively, was used as a model membrane monolayer. The size of the monolayer simulation system was selected based on the work done by van der Ploeg and Berendsen (11). Other researchers have obtained similar geometric and energetic equilibrium property values with regard to the size of the simulation system (12) permitting a minimum effective size (phospholipids) of the monolayer to be defined.

The unit cell parameters used for building the DMPC monolayer were a = 8 A, b = 8 A, c = 32 A and $\gamma = 96.0^0$. These unit cell parameters yield an average surface area per phospholipid of 64 A^2. The experimental reported value for average surface area per phospholipid is 66 A^2 for a fully hydrated fluid lamellar (Lα) phase of DMPC (12). An alternate initial set of a and b unit cell parameters (a = 8.5 A, b = 8.5 A), corresponding to surface area per phospholipid of 72.3 A^2, was used to optimize the DMPC monolayer structure without any solute present. The resulting optimized structure is virtually identical to that found using the "standard" a = 8 A and b = 8 A. dimensions. Both initial trial unit cell parameter sets optimize to (a = 8.15 A, b = 8.20 A, c = 32 A, $\gamma = 94.5^0$).

Each of the test solute molecules of the data set was inserted at three different positions (depths) in the DMPC monolayer with the most polar group of the solute molecule "facing" toward the head group region of the monolayer. Three corresponding MDS models were generated for each solute molecule with regard to the trial positions of the solute molecule in the monolayer. The three trial positions were,

1. Solute molecule in the head group region.
2. Solute molecule in between the headgroup region and the aliphatic chains.
3. Solute molecule in the tail region of the aliphatic chains.

The energetically favorable geometry of the solute molecule in the monolayer was sought using each of these trial positions. The three different positions of isopropanol (one of the test solute molecules) are shown in Fig. 2 (top) to illustrate this modeling procedure. The energetically most favorable geometry of butyl cellosolve (the solute molecule having highest MES value) is shown in Fig. 2 (bottom).

Molecular Dynamic Simulation

MDS was carried out using Molsim (14) with an extended MM2 force field. The selection of the simulation temperature was based on the phase transition temperature for DMPC which is 297°K (15). A simulation temperature of 311°K was selected since it is body temperature and is also above the DMPC phase transition temperature. Simulation temperature was held constant by coupling the system to an external constant temperature bath (16). The trajectory step size was 0.001 ps with a total simulation time of 20 ps. Two dimensional periodic boundary conditions were employed (a = 32 A, b = 32 A, c = 80 A and $\gamma = 96.0^0$) for the DMPC molecules of the monolayer model, but not the test solute molecule. Only a single solute molecule is explicitly considered in each MDS. The angle γ is the angle an extended DMPC molecule makes with the "planar surface" of the monolayer.

The model monolayer, without a solute molecule, was slowly heated starting at 20°K, then at 50°K, and from that point in increments of 50°K until a final step was used to achieve the simulation temperature of 311°K. At each temperature increment, 4 ps of MDS was carried out to allow for structural relaxation and the distribution of the kinetic energy throughout the simulation model. When 311°K was reached, 50 ps of MDS was performed to equilibrate the monolayer.

In order to prevent unfavorable van der Waals interactions between a solute molecule and the membrane DMPC molecules, one of the "center" DMPC molecules was removed from the equilibrated monolayer and a test solute molecule inserted in the space created by the missing DMPC molecule. Each of the solute molecules was placed at each of the three different positions in the monolayer described above with the polar portion of the solute "facing" towards the headgroup region. The same heating steps used to equilibrate the isolated monolayer were carried out with solute present. However, only a final 20 ps production run was carried out at 311°K. The MDS scheme for equilibration and exploration simulation is shown in Fig. 3.

Calculation of Descriptors

Both intramolecular physicochemical properties of the solute molecules and intermolecular solute-membrane interaction properties were calculated. "Properties" will be referred to as descriptors from this point forward as they constitute the independent variables in the QSAR models. The intramolecular descriptors that were considered are given in Table II. Log P, the water, 1-octanol partition coefficient, is *an intermolecular physicochemical property.* However, Log P has been computed using a group additive approach (17) which estimates this property as an *intramolecular descriptor.* Thus, Log P is placed in Table II. F(H_2O) and F(OCT), the aqueous and 1-octanol solvation free energies, respectively, are also intermolecular properties, and have been computed using an intermolecular method, a hydration shell model (7). Consequently, these properties are considered intermolecular descriptors and reported in Table II.

Some of the solute-membrane interaction descriptors extracted directly from the MDS trajectories are given in Table II. These intermolecular descriptors were calculated using the most stable (lowest total potential energy) solute-membrane geometry from the three positions sampled for each of the solutes. For example, Fig. 2b shows the lowest potential energy

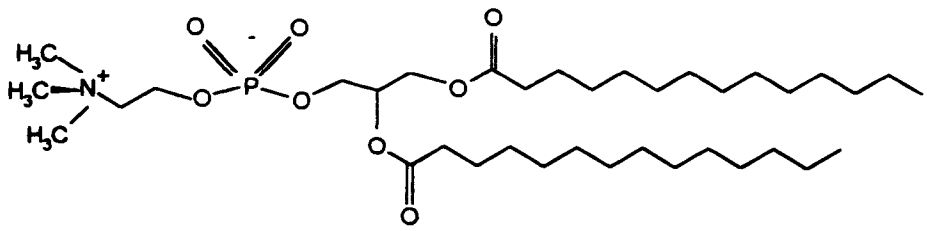

Fig. 1. The chemical structure of a DMPC molecule.

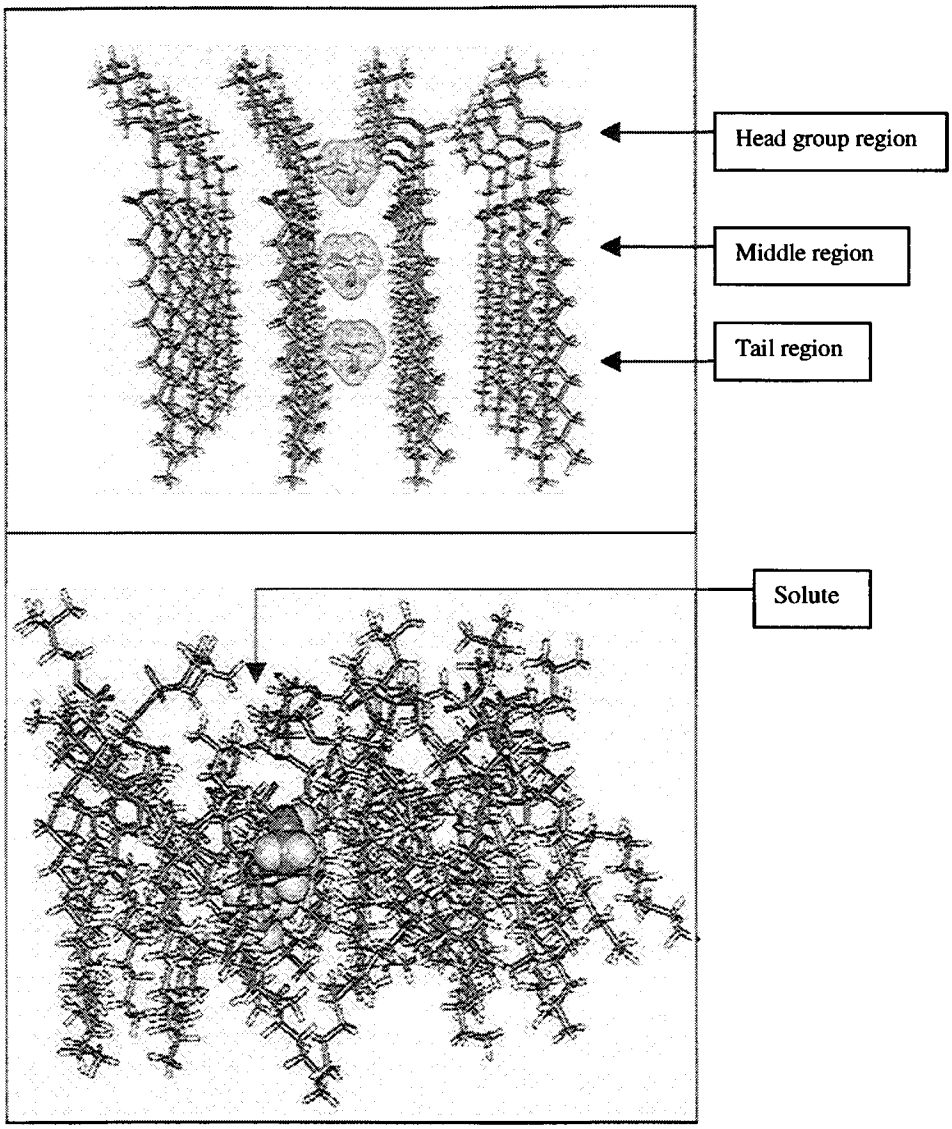

Fig. 2. (*top*) A "side" view of an isopropanol molecule inserted at three different positions in the DMPC model monolayer, prior to the start of each simulation; and (*bottom*) "side" view of the most energetically favorable geometry of butyl cellosolve (the solute having the highest MES value) in the DMPC model monolayer.

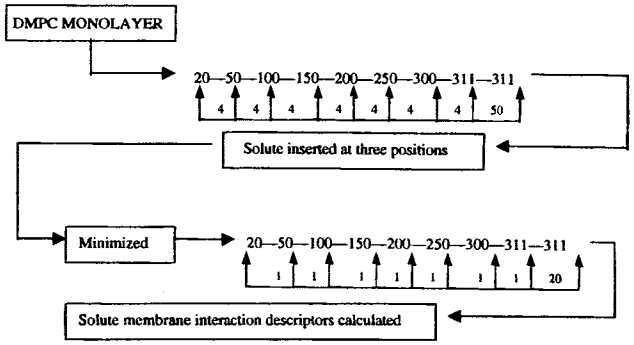

Fig. 3. The schedule of performing a membrane-solute MDS. See text for details.

state of butyl cellosolve in the membrane monolayer which was used to estimate the solute-membrane interaction descriptors. The remainder of the solute-membrane interaction descriptors used in the QSAR descriptor trial set were determined using data from the MDS trajectories.

D (Diffusion Coefficient of the Solute in the Membrane)

A diffusion coefficient can be calculated by (1) the mean-square displacement method or (2) by the force auto-correlation method. In this study the mean-square displacement method (18) was used.

From the MDS trajectory, each movement Δd_i of the solute molecule between two adjacent time steps was determined in the MDS. The mean-square displacement is correspondingly defined as,

Table II. Solute Descriptors Considered in the Trial QSAR Descriptor Set

Intramolecular Descriptors	Source
Log P (Logarithm of the partition coefficient)	*c*
HOMO (Highest occupied molecular orbital energy)	*a*
LUMO (Lowest occupied molecular orbital energy)	*a*
Dipole moment	*b*
Molecular Volume	*c*
SA (Molecular surface area)	*c*
Density	*a*
Molecular weight	*b*
Molecular refractivity	*a*
Number of hydrogen bond acceptors	*a*
Number of hydrogen bond donors	*a*
Number of rotatable bonds	*a*
Jurs- Stanton CSPA (charged partial surface area) descriptors	*a*
Kappa descriptors (topological descriptors)	*a*
Radius of Gyration	*a*
PM (Principle moment of inertia)	*c*

Symbol	Descriptions of the Intermolecular Descriptors
D	Diffusion coefficient (cm^2/sec).
⟨d⟩	Average depth of solute in membrane (A°).
S	Entropy of solute in the membrane (cal/mole/deg).
Dp	Change in membrane density upon uptake of a solute molecule (amu/A^3).
E(total)	Average total interaction energy of the solute and membrane (kcal/mole).
E_{inter}(total)	Interaction energy between the solute and the membrane at the total system minimum potential energy (sum of electrostatic, H-bonding and vdw energies) (kcal/mole).
E(chg)	Electrostatic interaction energy between the solute and the membrane at total system minimum potential energy (kcal/mole).
E(vdw)	Van der Waals interaction energy between the solute and the membrane at the total system minimun potential energy (kcal/mole).
E(chg + vdw)	Electrostatic plus van der Waals interaction energy between the solute and the membrane at the total system minimum potential energy (kcal/mole).
F(H_2O)	The aqueous solvation free energy computed using a hydration shell model (7,31).
F(OCT)	The 1-octanol solvation free energy computed using a hydration shell model (7,31).

[a] Computed using Cerius2 (20).
[b] Calculated using MOPAC 6.0 (10).
[c] Calculated using Chemlab II (7).

$$\overline{\chi}^2 = \frac{\sum_{i=1}^{n} \Delta d_i^2}{t} \quad (1)$$

where n is the total number of time steps in MDS and t denotes the total time of the MDS. From the Einstein diffusion equation (19), the diffusion coefficient, D, is given by

$$D = \frac{\overline{\chi}^2}{2t} \quad (2)$$

S(Alignment and Conformational Entropy)

The alignment and conformational entropy, S, is calculated directly from the partition function, Q, and the corresponding thermodynamic probability of each state, P_i. Q is assumed to be well-represented by the states of the MDS.

$$Q = \sum_{i=1}^{n} \exp(-E_i/RT) \quad (3)$$

$$P_i = \frac{\exp(-E_i/RT)}{Q} \quad (4)$$

$$S = -R \sum_{i=1}^{n} P_i \ln P_i \quad (5)$$

R is the gas constant, E_i is the total potential energy of the i^{th} state (time step) of the MDS and T is the Kelvin temperature.

⟨d⟩ (Average Depth of a Solute in a Membrane)

A "center" atom of the solute molecule and the closest phosphate atom of the nearby (adjacent) phospholipid are selected to define solute "depth" in the membrane. The distance, d_i, between these two atoms for each step i in the MDS trajectory was computed as,

$$\langle d \rangle = \sum_{i=1}^{n} P_i * d_i \quad (6)$$

Again, P_i is the thermodynamic probability [eq. (4)] of the i^{th} state (MDS time step). The average depth can be interpreted as defining where the solute has the most favorable thermodynamic interaction with the membrane monolayer model.

dρ(Change in Membrane Density Upon Uptake of the Solute)

The Cerius 2 (20) software was used to calculate the Connolly volume of the model membrane assembly. The lowest potential energy state for the isolated monolayer model, and a spherical probe with a radius of 1.2 A^0, was used in calculating the Connolly volume. The Connolly volumes were also determined for each of the solute-membrane systems, again using the lowest potential energy states. The change in density, dρ, due to each solute was then calculated as,

$$d\rho = \left(\frac{M_{(mem+sol)}}{V_{(mem+sol)}}\right) - \left(\frac{M_{(mem)}}{V_{(mem)}}\right) \quad (7)$$

$M_{(mem)}$ is the mass of the isolated model membrane without the solute, $M_{(mem+sol)}$ is the mass of the membrane model with the solute and $V_{(mem)}$ and $V_{(mem+sol)}$ are the corresponding Connolly volumes, respectively.

Construction and Testing of QSAR Models

QSAR models were constructed using the Genetic Function Approximation, GFA, (21) which is a method based on the genetic algorithm paradigm. The GFA algorithm is coded in the program WOLF (22). Statistical significance of a QSAR model is based on Friedman's lack of fit (LOF) measure (23). The LOF measure is designed to resist overfitting which is a problem often encountered in constructing statistical models.

Optimization of a QSAR model was considered to be realized when descriptor usage became constant and independent of increasing crossover operations. A crossover operation is the "birth" of a child model from its parent models. Both partial least-squares regression (PLS) and multi-dimensional linear regression (MLR) can be used in WOLF to establish functional data fits. MLR was used in this study.

In order to test and validate the MI-QSAR models, the dependent variable, MES, was randomly "scrambled" with respect to the independent variables (descriptor set) to see if meaningful correlations (QSARs) could be found (24) for the scrambled data sets. The loss of any significant correlation for each of the scrambled data sets is taken as evidence of the significance of the QSAR for the nonscrambled data set. The covariance among the descriptors in the optimized MI-QSAR models was evaluated by constructing the linear cross-correlation matrix of the descriptors, and by comparing relative descriptor usage in the crossover plots.

RESULTS

QSAR Analysis Using Intramolecular Physicochemical Descriptors

The best three QSAR models constructed by correlating MES only with the intramolecular solute descriptors are as follows:

$$\begin{aligned} \text{MES} &= 31.59 + 2.31 * \text{HOMO} \\ n &= 18;\ r^2 = 0.29;\ xv - r^2 = 0.16; \\ \text{LSE} &= 2.89;\ \text{LOF} = 3.51 \end{aligned} \quad (8)$$

$$\begin{aligned} \text{MES} &= 10.78 + 0.076 * \text{PM(Z)} - 0.09 * \text{SA} \\ n &= 18;\ r^2 = 0.42;\ xv - r^2 = 0.29; \\ \text{LSE} &= 2.37;\ \text{LOF} = 3.59 \end{aligned} \quad (9)$$

$$\begin{aligned} \text{MES} &= 2.82 - 0.56 * \text{Log P} \\ n &= 18;\ r^2 = 0.19;\ xv - r^2 = 0.09; \\ \text{LSE} &= 3.30;\ \text{LOF} = 4.02 \end{aligned} \quad (10)$$

The intramolecular solute descriptors in the QSAR models above (eqs. 8–10) are defined in Table II. Log P has often been found to be a significant correlation descriptor to biological activities when used in quadratic form (log P and (log P)^2) (25). Consequently Log P was considered in both linear and square forms in the trial set of QSAR descriptors in the intramolecular QSAR analysis and in the MI-QSAR analysis. However, no significant QSAR model could be constructed which included a (Log P)^2 term.

A QSAR model is usually considered significant if it has a correlation coefficient (r^2) greater than 0.7. None of the QSAR models above has an r^2 value greater than 0.5. Similar QSAR models, in terms of r^2 values, have been reported by Cornin, Basketter & York (26). These workers also report HOMO and LogP as important descriptors in their models. For both the intramolecular QSAR models reported in eqs. (8)–(10), as well as those of other workers, the descriptors found in the models can be used to speculate mechanisms of action. However, the correlation coefficients of all intramolecular MES QSAR models are so low that no plausible conclusion can be made regarding a biochemical mechanism of eye irritation.

QSAR Analysis Including Intermolecular Solute-Membrane Interaction Descriptors

Visual inspections of the MDS were done using the MOLSIM visualization tool MDMOVIE. It was observed that polar molecules, like alcohols and ketones, prefer to stay in the head group region of the membrane monolayer. Conversely, lipophilic molecules, like the hydrocarbons, locate in the tail region of the monolayer. Some of the solute molecules behave in distinct fashions. For example, toluene perfers a position "parallel" to the monolayer "plane", as the most stable intermolecular geometry.

Diffusion coefficients calculated for the ECETOC solutes are in the range of 10^{-6} to 10^{-7} cm^2/sec, which is consistent with reported experimental values for organics (10^{-5} to 10^{-9} cm^2/sec) in lipid membranes (27). The experimental value of the diffusion coefficient for propanol in a phospholipid membrane is 7.88×10^{-7} cm^2/sec (27). The diffusion coefficient for propanol was determined using the MDS scheme (see Fig. 3) for MI-QSAR analysis. The average value computed for the diffusion coefficient is 1.20×10^{-7} cm^2/sec. Thus, the MDS scheme used in this study would seem to be reasonable with respect to diffusion of the solute in the model DMPC monolayer.

Permeation and solubility coefficients have not been estimated, but the information they may provide to an MI-QSAR model should be captured among the diffusion coefficient, F(H_2O), F(OCT), log P and the membrane-interaction energy descriptors used in the trial MI-QSAR descriptor set.

The estimated average "depth" of a solute in the membrane exhibits a large variability over the ECETOC data set, ranging

from 5.5 to 14.5A. There is very little change in the conformational entropy of the membrane model upon uptake of solute molecules. This is likely due to the low "concentration" of solute molecules in the model (1solute per 15 DMPC molecules). Also, only small changes in monolayer density are seen upon solute uptake.

The most striking feature that could be discerned in the solute-membrane MDS modeling was the large variation in the interaction energies for the different solute molecules with the model monolayer. These variations in solute-membrane interaction energy can be related to size, shape, partial atomic charge distribution and net polarity of the solute. Location of the solute in the monolayer also impacts the interaction energetics.

The solute-membrane properties are not particularly sensitive to small (±10°K) changes in the simulation temperature. More generally, considerable work was done to develop the MI modeling method by Jin and Hopfinger (28–30) prior to the undertaking of this eye irritation MI-QSAR study. Jin and Hopfinger compared, for example, the computed order parameters and the trans/gauche ratios of the aliphatic chains of the DMPC monolayer to experimental values to validate the MDS membrane-solute interaction methodology.

Overall, the solute-membrane MDS seem to yield realistic, low energy, portrayals of this class of intermolecular interactions as judged by the analysis of a variety of properties of the model system as described above. Thus, the solute-membrane interaction physicochemical properties were included as members of the trial set of QSAR descriptors to construct a MES QSAR model. The best QSAR models obtained using the GFA algorithm are:

$$\begin{aligned} \text{MES} &= -0.03 - 0.46 * \text{F}(\text{H}_2\text{O}) - 0.12 * \text{E(chg + vdw)} \\ n &= 16;\ r^2 = 0.87;\ xv - r^2 = 0.80; \\ \text{LSE} &= 0.83;\ \text{LOF} = 2.10 \end{aligned} \quad (11)$$

$$\begin{aligned} \text{MES} &= -0.81 - 0.07 * \text{E(chg + vdw)} \\ &\quad - 0.48 * \text{F}(\text{H}_2\text{O}) + 0.35 * \text{Kappa-3-AM} \\ n &= 16;\ r^2 = 0.92;\ xv - r^2 = 0.88; \\ \text{LSE} &= 0.41;\ \text{LOF} = 1.94 \end{aligned} \quad (12)$$

$$\begin{aligned} \text{MES} &= -1.06 - 0.23 * \text{E(chg)} - 0.43 * \text{F}(\text{H}_2\text{O}) \\ n &= 16;\ r^2 = 0.87;\ xv - r^2 = 0.76; \\ \text{LSE} &= 0.82;\ \text{LOF} = 2.02 \end{aligned} \quad (13)$$

The independent variables (descriptors) in eqs. (11)–(13) are defined in Table II. Kappa-3-AM is a molecular connectivity descriptor (20) and reflects the size and shape of a molecule. The linear cross-correlation matrix for the descriptors in eqs. (11–13) and MES is presented in Table III. The scrambling experiments to test the validity of eqs. (11–13) led to models with $R^2 < 0.3$. These low R^2 values of the scrambling experiments suggest that eqs. (11–13) are significant MI-QSAR models and not the result of random correlations.

There are two outliers in the ECETOC data subset for all three of the membrane-interaction QSAR models given by eqs. (11)–(13). The predicted MES value of propylene glycol is significantly larger than is reported. Interestingly, propylene glycol is also an outlier in the model developed by Abraham and coworkers (3). The other outlier, acetone, can be marginally

Table III. The Linear Cross-Correlation Matrix for the MES and the MI-QSAR Descriptors in Eqs. (11–13)

	MES	F(H2O)	E(chg + vdw)	E(chg)	Kappa-3-AM
MES	1				
F(H2O)	−0.87	1			
E(chg + vdw)	−0.32	−0.01	1		
E(chg)	−0.49	0.18	0.50	1	
Kappa-3-AM	0.30	0.09	−0.43	−0.71	1

included in the modeling, and leads to a model with the same descriptors as in eq. (11), but with a lower r^2 value. The best MI-QSAR model with acetone included is,

$$\begin{aligned} \text{MES} &= 0.68 - 0.42* \text{F}(\text{H}_2\text{O}) - 0.091 * \text{E(chg + vdw)} \\ n &= 17;\ r^2 = 0.75;\ xv - r^2 = 0.65; \\ \text{LSE} &= 1.61;\ \text{LOF} = 3.73 \end{aligned} \quad (14)$$

$F(H_2O)$ is the dominant single descriptor in relation to the MES values, and its individual MES QSAR model, excluding acetone and propylene glycol, is,

$$\begin{aligned} \text{MES} &= 1.02 - 0.46 * \text{F}(\text{H}_2\text{O}) \\ n &= 16;\ r^2 = 0.76;\ xv - r^2 = 0.64; \\ \text{LSE} &= 1.51;\ \text{LOF} = 2.43 \end{aligned} \quad (15)$$

A comparison of eq. (15) to eqs. (11)–(13) reveals that the regression coefficients for $F(H_2O)$ are all nearly the same (0.43 to 0.48). This, in turn, indicates that the contribution to "explaining" the variation in MES by $F(H_2O)$ in eq. (15) is the same as in eqs. (11)–(13). Thus, while $F(H_2O)$ is statistically the dominant descriptor, E(chg + vdw) of eqs. (11)–(12), kappa-3-AM of eq. (12), and E(chg) of eq. (13) are each making separate (from $F(H_2O)$), and significant, contributions to explaining the variation in the MES of the training set.

DISCUSSION

Significant MES QSAR models were obtained for the structurally diverse subset of the ECETOC database by combining intramolecular physicochemical properties with intermolecular solute-membrane interaction properties. The most important type of descriptors used in the QSAR models are solute-membrane interaction energies and aqueous solvation free energies of the solute molecules. No statistically comprehensive QSAR model could be constructed unless both types of descriptors were employed in the trial descriptor set.

Aqueous solvation free energies were computed using a hydration shell model (31). Strictly speaking this is an intramolecular method of calculation, and the corresponding $F(H_2O)$ descriptor could be classified as an intramolecular property. However, aqueous solvation free energy is an intermolecular thermodynamic property arising from interactions between a solute molecule and the water molecules in which it is "embedded". Moreover, intermolecular simulations may be used to estimate $F(H_2O)$, although the computational time and effort to do so is substantially greater than the empirical hydration shell approach. Thus, we opted to estimate $F(H_2O)$ using the

hydration shell model, but classify this descriptor as an intermolecular property.

Increasingly negative $F(H_2O)$ values correspond to increasing aqueous solubility of a solute. In eqs. (11)–(13) it is seen that aqueous solvation free energy is negatively correlated with MES. This relationship suggests that water soluble compounds have a greater propensity to be eye irritants than hydrophobic compounds. The solute-membrane energy interaction descriptors in eqs. (11)–(13) are also negatively correlated with the MES. Thus, as the "binding energy" of a solute molecule to the membrane increases (a more negative descriptor value), it is going to be more of an eye irritant than solutes which do not bind as strongly to the membrane.

Combining the interpretations of the two types of descriptors in eqs. (11)–(13), leads to the following "picture". If a solute molecule is water soluble it possesses some polar moieties. These polar groups can also have favorable binding interactions with a membrane, probably involving the head group region. Polar alcohols are known to disturb membrane structure (32) which is consistent with this picture. The MES QSAR models given by eqs. (11–13) and (15) suggest that the eye irritation potency of a solute molecule, as measured by the Draize test, is mainly due to the aqueous solubility of the solute.

The significance of $F(H_2O)$ in the proposed model for eye irritation gleaned from the MES QSAR models prompted an independent assessment of the reliability of the $F(H_2O)$ calculated values. This assessment consisted of correlating the calculated aqueous solvation free energies [$F(H_2O)$] of some of the solutes in the QSAR data subset with their respective experimental solubilities. Due to a lack of experimental solubility data for all of the solute molecules in the data subset, representative solute molecules, whose experimental solubility values are available, were selected for the particular chemical classes of the data subset. A reasonable correlation was established between calculated $F(H_2O)$ values and experimental solubility measures for the 20 solute molecules listed in Table IV. The correlation equation is,

$$\text{Solubility} = 1.53 * F(H_2O) - 5.33 \qquad (16)$$

$$n = 20;\ r^2 = 0.68$$

There are four outliers using eq. (16). These outliers probably arise because in the hydration shell calculation of $F(H_2O)$ dipole-dipole interactions between the solute and water are not taken into consideration. Removal of the two largest outliers (in bold in Table IV) results in the correlation given by eq. (17), and removal of all four outliers yields eq. (18).

$$\text{Solubility} = 0.41 * F(H_2O) - 2.71$$

$$n = 18;\ r^2 = 0.81 \qquad (17)$$

$$\text{Solubility} = 0.40 * F(H_2O) - 2.90 \qquad (18)$$

$$n = 16;\ r^2 = 0.90$$

Table IV also contains the experimental solubility values (33), the calculated $F(H_2O)$ values and the predicted solubilities, using eq. (16), for the test set of solute molecules.

The solute-membrane interaction energy terms in eqs. (11)–(13) suggest that eye irritation potency increases with increasing binding of the solute to the membrane. A straightforward interpretation of this type of descriptor term in eqs. (11)–(13) is that disruption of membrane structure and, likely, function promotes eye irritation.

Table IV. Calculated $F(H_2O)$ and Experimental Solubility Values

Solutes	Calculated $F(H_2O)$	Observed solubility	Predicted solubility[a]	Residuals
3 Methyl hexane	2.75	4.47	4.29	0.18
2 Methyl pentane	2.54	3.79	4.20	−0.41
Methylcyclopentane	1.36	3.30	3.67	−0.37
Xylene	−0.85	2.83	2.69	0.14
Toluene	−0.93	2.22	2.66	−0.44
Styrene	−1.53	2.58	2.39	0.19
Butanol	−7.45	0.02	−0.24	0.26
Isobutanol	−7.35	0.09	−0.20	0.28
Ethly acetate	−2.82	0.02	1.82	−1.79
Butylacetate	−2.42	1.18	2.00	−0.82
Dodecane	3.17	7.67	4.48	**3.19**
Hexadiene	0.83	2.69	3.44	−0.75
Cyclohexanol	−8.08	0.42	−0.52	0.94
2 Ethyl 1hexanol	−6.49	2.11	0.19	**1.92**
Propanol	−7.65	−0.62	−0.33	−0.29
Propylbenzene	−0.48	3.34	2.85	0.49
Isobutylacetate	−2.16	1.24	2.11	−0.87
Isoproplyacetate	−2.48	0.60	1.97	−1.37
Cyclooctane	1.62	4.15	3.79	0.36
Flourobenzene	−0.94	1.79	2.65	−0.86

[a] Using eq. (16).

A generalized view of eye irritation as scored by the Draize test can be made from the discussion above and eqs (11)–(13). The $F(H_2O)$ descriptor reflects the availability of a solute molecule to disrupt membrane structure. That is, $F(H_2O)$ is a solute concentration measure. The membrane-solute interaction energy descriptors provide measures of the intrinsic membrane disrupting potencies of each of the individual solute molecules. MES is thus controlled by an *effective solute concentration* coupled to the *intrinsic membrane disruption* propensity of the solute. This interpretation of the MI-QSARs is similar to the model of Abraham and coworkers (3) in terms of an effective solute *concentration*. Their model identifies the significance of transferring the solute from its application state (pure organic liquid or aqueous solution) to "an organic biophase" (the biological structure of the eye). In other words, the concentration of the solute in the "organic biophase" is crucial to eye irritation potency.

It is also important to point out two of the biochemical factors not considered in the MI-QSAR formalism. First, the possible interactions of a solute with membrane proteins are not considered. If this class of interactions is important to the expression of eye irritation for a compound, MI-QSAR analysis is not applicable and will fail. Secondly, at the current stage of development of MI-QSAR analysis, cellular membrane specificity, in terms of specific phospholipids, has not been considered. The MI-QSAR models are based solely on DMPC monolayer binding site models. However, there is no reason that other phospholipid membrane models cannot be considered in MI-QSAR analysis.

The membrane-interaction QSAR model given by eq. (11) was used to predict MES of ECETOC molecules that were not included in the training data subset during model development. The prediction of MES for compounds outside the training set constitutes a validation test of the QSAR model. Table I contains, in addition to the training set, the predicted and reported

experimental MES values for five compounds outside the original training data subset. Four of the five test compounds in Table I are well predicted by the QSAR model. Cyclohexanol is predicted to have a much lower MES value, corresponding to a 40% difference, than the reported value. Still, cyclohexanol is predicted to be an eye irritant.

The successful treatment of a representative set of structurally diverse compounds from the ECETOC eye irritation training set by including interactions of these compounds with membrane models is both novel and important. However, we do not wish to imply any specific single mechanism of action from this QSAR analysis. We have only developed a model consistent with the data. However, from a purely practical point of view, membrane-interaction QSAR analysis, may be a breakthrough method to reduce animal testing in several classes of current risk assessment screens. Some application areas that may involve membrane interactions in the biochemical mechanisms and, therefore, which could be evaluated using membrane-interaction QSAR analysis include:

1. Skin sensitivity and irritation.
2. Aquatic toxicity.
3. Membrane bound drug-receptor interactions.
4. General modeling of bioavailability.

In particular, there are some bioavailability measures, such as Caco-2 cell permeation coefficients (34), which are directly related to cell uptake and transport, and, therefore, well-suited for study using MI-QSAR analysis. Additional work on membrane-interaction QSAR analysis continues in our laboratory with the hope of learning more about both the reliability and general utility of the method.

ACKNOWLEDGMENTS

We are pleased to acknowledge the financial support of the Procter & Gamble Company. Resources of the Laboratory of Molecular Modeling and Design at UIC, and of The Chem21 Group, Inc. were used in performing this work. We had many enlightening discussions with Dr. Edward Thompson of the Procter & Gamble Co., and with Prabha Venketarangan, Chris Klein and Dick Wood, of our laboratory, over the course of this investigation.

REFERENCES

1. J. H. Draize, G. Woodard, and H. O. Calvery. Methods for the study of irritation and toxicity of substances applied to the skin and mucous membranes. *J. Pharmacol. Exp. Ther.* **82**:377–390 (1944).
2. D. M. Bagley, P. A. Botham, J. R. Gardner, G. Holland, R. Kreiling, R. W. Lewis, D. A. Stringer, and A. P. Walker. Eye irritation: reference chemical data bank. *Toxicol. In Vitro.* **6**:487–491 (1992).
3. M. H. Abraham, R. Kumarsigh, J. E. Cometto-Muniz, and W. S. Cain. A Quantitative Structure-Activity Relationship (QSAR) for a Draize Eye Irritation Database. *Toxicol. In Vitro.* **12**:201–207 (1998).
4. A. K. Debnath, G. Debanth, A. J. Shusterman, and C. Hansch. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. Correlation with molecular orbital energies and hydrophobicity. *J. Med. Chem.* **34**:786–797 (1991).
5. M. D. Barratt. The role of structure activity relationships and expert systems in alternative strategies for the determination of skin sensitization, skin corrosivity and eye irritation. *ATLA* **23**:111–122 (1995).
6. Aldrich. Aldrich *Catalogue Handbook of Fine Chemicals*, Gillingham Dorset, UK, 1988.
7. R. A. Pearlstein. CHEMLAB-II *Users Guide*, CHEMLAB Inc. 1780 Wilson Dr., Lake Forest, IL 60045, 1988.
8. HyperChem, *Release 4.5 for MS Windows*. Hypercube Inc, Waterloo, Ontario, Canada, 1997.
9. H. Hauser, I. Pascher, R. H. Pearson, and S. Sundell. Preferred conformation and molecular packing of phosphatidylethanolamine and phosphatidylcholine. *Biochim. Biophys. Acta.* **650**:21–51 (1981).
10. Mopac 6.0 release notes, Frank J. Seiler Research laboratory, United States Air Force Academy, 1990.
11. P. van der Ploeg and H. J. C. Berendsen. Molecular dynamics simulation of a bilayer membrane. *J. Chem. Phys.* **76**:327–3276 (1982).
12. T. R. Stouch. Lipid membrane structure and dynamics studied by all atom molecular dynamicssimulations of hydrated phospholipid bilayer. *Mol. Simulation.* **1**:335–362 (1993).
13. B. A. Lewis and D. M. Engelman. Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles. *J. Mol. Biol.* **166**:211–217 (1983).
14. D. C. Doherty. The Chem 21 Group Inc. *Molsim Version 3.0 User's Guide*. Lake Forest, IL 60045, 1994.
15. M. Bloom, E. Evans, and O. Mouritsen. Physical properties of the fluid lipid-bilayer component of cell membranes. a perspective. *Quart. Rev. Biophys.* **24**:293–397 (1991).
16. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling with to an external bath. *J. Phys. Chem.* **81**:3684–3590 (1984).
17. C. Hansch and A. Leo. *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Wiley-Interscience. New York, 1979.
18. I. Tinoco, K. Sauer, and J. Wang. *Physical Chemistry*. Prentice-Hall, NJ, 1985.
19. A. Einstein. *Investigations on the Theory of the Brownian Movement*. Methuen & Co. Ltd., London, 1926.
20. Cerius2. *V.3.0 Users Guide*. San Diego, Molecular Simulations Inc., 1997.
21. D. Rogers and A. J. Hopfinger. Applications of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.* **34**:854–862 (1994).
22. D. Rogers and Molecular Simulations Inc. *WOLF 6.2 GFA Program*, 1994.
23. J. Friedman. *Multivariate Adaptive Regression Spline*, Technical Report No. 102, Laboratory for Computational Statistics, Dept. of Statistics, Stanford University, Stanford, CA, 1988.
24. H. Waterbeemd. *Chemometric Methods in Molecular Design*. VCH Publishers, Inc., New York, 1995.
25. R. Franke. *Theoretical Drug Design Methods*. Akademie-Verlag, Berlin, 1984.
26. T. D. Cronin, D. A Basketter, and M. York. A quantitiative structure-activity relationship (QSAR) investigation of a draize eye irritation database. *J. Tox. In Vitro.* **8**:21–28 (1994).
27. W. D. Stein. *Transport and Diffusion across Cell Membranes*. Academic press, Inc., New York, 1986.
28. B. Jin and A. J. Hopfinger. Characterization of lipid membrane dynamics by simulation. 3. Probing molecular transport across the phospholipid bilayer. *Pharm. Res.* **13**:1786–1794 (1996).
29. B. Jin and A. J. Hopfinger. Characterization of lipid membrane dynamics by simulation 2. Quantitative representation of lipid fluidity. *Computational Polym. Sci.* **6**:95–101 (1996).
30. B. Jin and A. J. Hopfinger. Characterization of lipid membrane dynamics by simulation: I. Torsion angle motions of the linear chain. *Biopolymers.* **41**:37–50 (1997).
31. A. J. Hopfinger. *Conformational Properties of Macromolecules*. New York and London. 1973.
32. S. C. McKarns, C. Hansch, W. S. Caldwell, W. T. Morgan, S. K. Moore, and D. J. Doolittle. Correlation between hydrophobicity of short-chain aliphatic alcohols and their ability to alter plasma membrane integrity. *Fundamental and Applied Tox.* **36**:62–70 (1997).
33. T. M. Nelson and P. Jurs. Prediction of Aquoeus Solubility of Organic Compounds. *J. Chem. Inf. Comput. Sci.* **34**:601–609 (1994).