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

Amit S. Kulkarni, A. J. Hopfinger[*] and José S. Duca

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

Abstract

Purpose: The purpose of this study was to explore a possible mechanism of eye irritation by constructing a corresponding general quantitative structure-activity relationship (QSAR) model using a genetic algorithm. The model was derived from a subset of diverse chemical structures found in 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.

Conclusion: 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.

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

INTRODUCTION

Eye irritation potential is an important toxicological endpoint. It has traditionally being scored using the Draize rabbit eye test (Draize, Woodard and Calvery 1944). Exposure of the body's external tissues to chemicals or chemical formulations can produce irritation which might lead to permanent damage of the tissue. Thus, most commercial chemical products are evaluated for their ocular irritation potential. This study has explored possible alternatives to the Draize eye test, and has probed possible mechanisms of eye irritation by constructing a corresponding general quantitative structure -activity relationship (QSAR) model. In the Draize test the test compounds are placed in a rabbit eye, and the damage to various eye components evaluated. The mechanisms of eye irritation are still not understood, but it appears that there are no "irritation receptors" in the eye. Uptake and diffusion into the keratocytes of the corneal epithelium of the eye is a significant event. Molecules which are exposed to the eye must permeate through the cell membrane of the keratocytes. We have hypothesized that interactions of test organic molecules with a cell membrane are, at least, partly responsible for eye irritation. Moreover, the lipid regions of the membrane bilayer of the cell might, in composite, constitute a "**general receptor"** for eye irritation. Thus, we decided to simulate the uptake and interaction of test solute molecules with model monolayer membranes as a part of our QSAR analysis. Molecular dynamic simulation (MDS) was used to generate intermolecular membrane-solute interaction descriptors**.** The estimated membrane-solute interaction properties from the simulations were added to the intramolecular physicochemical property descriptors of the test solute molecules to provide an extended set of trial descriptors for building an eye irritation QSAR model. This overall methodology is called *membrane-interaction QSAR analysis*, (MI-QSAR analysis).

QSAR models were constructed using the genetic function approximation, GFA, which is a method based on the genetic algorithm paradigm. The GFA algorithm is coded in the program WOLF (Rogers 1994) . Statistical significance of a QSAR model is based on Friedman's lack of fit (LOF) measure (Friedman 1988) . The LOF measure is designed to resist overfitting which is a problem often encountered in constructing statistical models. Since number of descriptors normally available in MI-QSAR analysis exceeds the number of observations (test

compounds), the ability to prevent overfitting using GFA is critical to the successful construction of a QSAR model.

A smoothing factor of 1.8 and 70000 crossover operations were used to optimize the QSAR models using WOLF. Optimization of a QSAR model was considered to be realized when descriptor usage became constant and independent of increasing number of GFA 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 (Waterbeemd 1995) 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 non-scrambled 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.

METHODS AND MATERIALS

A. Eye Irritation Potency and the Training set

The eye irritation potency measures (dependent variables in the QSAR analysis) 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 x 1000)/relative molecular mass. Density values were obtained from a standard source (Aldrich 1988). Molar adjusted eye scores were then calculated as the raw eye irritation scores divided by the molarity of the solution. There is a considerable concern that the MES values lack sufficient quality and reproducibility to justify usage in constructing QSAR models. An analysis of raw MES data, involving the use of multiple experiments (animals) for a single test compound, reveals that the MES standard deviations of fit for the compounds that are "safe" (low eye irritation) have highly reproducible eye irritation measures (small standard deviations).   Conversely, 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 selected in this study are a subset of the 37 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 compounds of Table Ia and 5 compounds of Table V 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 to permit us to analyze the potential of the MI-QSAR approach to treat other toxicological applications.

B. Building Solute Molecules and the DMPC Monolayer:

The solute molecules, see Table Ia, were built using the Chemlab-II (Pearlstein 1988) molecular modeling package. A single dimyristoylphosphatidylcholine (DMPC) molecule was built in HyperChem (Hyperchem 1997) using available crystal structure data (Hauser et al. 1981). The aliphatic chains of DMPC molecule were placed in *trans* planar conformations. 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 (Stewart 1990) 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 (Ploeg and Berendsen 1982). They performed MDS on decanoate bilayers having (2x8x2) and (2x16x2) phospholipid molecules. It was found that the estimated order parameter values of the two model assemblies agree with each other suggesting that the smaller assembly is adequate for modeling. Other researchers have obtained similar geometric and energetic equilibrium property values with regard to the size of the simulation system (Stouch 1993)

permitting a minimum effective size (phospholipids) of the monolayer to be defined. A DPMC molecule is shown below in Fig.1.
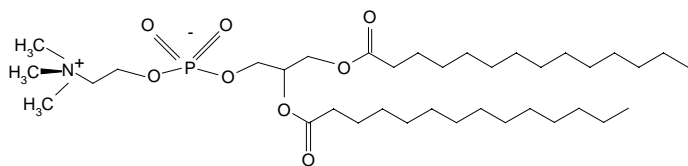

Fig 1. The chemical structure of a DMPC molecule.

The unit cell parameters used for building the DMPC monolayer were a = 8 Å , b= 8 Å, c = 32 Å and γ = 96.0 ^{0}. These unit cell parameters yield an average surface area per phospholipid of 64 Å$^{2.}$ The experimental reported value for average surface area per phospholipid is 66 Å^2 for a fully hydrated fluid lamellar (Lα) phase of DMPC (Lewis and Engelman 1983).

Each of the test solute molecules of the data set was inserted at three different positions (depths) in the DMPC monolayer model assembly 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 to illustrate this modeling procedure.

RESULTS

The best QSAR models obtained using the GFA algorithm are:

1. MES = − 0.03 – 0.46 * F(H$_2$O) – 0.12 * E(chg+vdw)

n = 16; r^2 = 0.87; xv - r^2 = 0.80; LSE = 0.83; LOF = 2.1

2. MES = − 0.81 – 0.07 * E(chg+vdw) – 0.48 * F(H$_2$O) + 0.35 * Kappa-3-AM

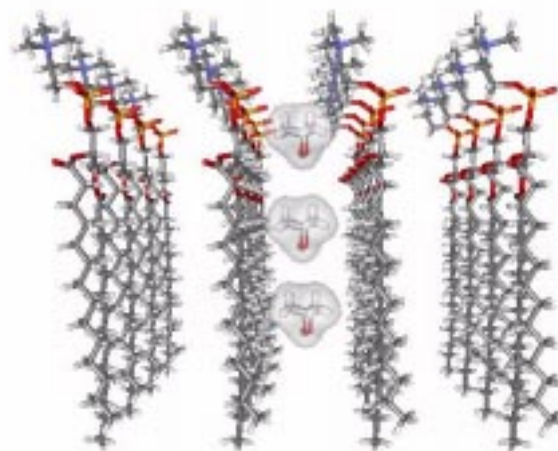n = 16 ; r^2 = 0.92; xv - r^2 = 0.88; LSE = 0.41; LOF = 1.94

Fig 2 A "side" view of an isopropanol molecule inserted at three different positions in the DMPC model monolayer, prior to the start of each MDS.

3. MES = − 1.06 – 0.23 * E(chg) – 0.43 * F(H$_2$O)

n = 16 ; r^2 = 0.87; xv - r^2 = 0.76; LSE = 0.82; LOF = 2.02

Where,

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$_2$O)   The aqueous solvation free energy computed using a hydration shell model.

Kappa-3-AM A topological connectivity descriptor.

CONCLUSION

Increasingly negative F(H$_2$O) values correspond to increasing aqueous solubility of a solute. In eqs. (1)-(3) 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. (1)-(3) 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. (1)-(3), 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 (McKarns et al 1997) which is consistent with this picture. The MES QSAR models given by eqs. (1-3) 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. A mechanistic generalization of eye irritation as scored by the Draize test can be made from the discussion above and eqs (1)-(3). 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.

ACKNOWLEDGEMENTS

REFERENCES

1. Draize, J. H, Woodard, G., and Calvery, H. O. 1944. Methods for the study of irritation and toxicity of substances applied to the skin and mucous membranes. *Journal of Pharmacology and Experimental Therapeutics.* 82: 377-390.
2. Rogers, D. and Molecular Simulations Inc. 1994, *WOLF 6.2 GFA Program.*
3. Friedman, J. 1988, *Multivariate Adaptive Regression Spline*, Technical Report No. 102, Laboratory for Computational Statistics, Dept. of Statistics, Stanford University.
4. Waterbeemd, H., 1995. *Chemometric Methods in Molecular Design*. VCH Publishers, Inc., New York.
5. Aldrich 1988, *Catalogue Handbook of Fine Chemicals,* Gillingham Dorset, UK.
6. Pearlstein, R. A, 1988. CHEMLAB-II *Users Guide*, CHEMLAB Inc. 1780 Wilson Dr., Lake Forest , IL 60045.
7. HyperChem 1997, *Release 4.5 for MS Windows.* Hypercube Inc, Waterloo, Ontario, Canada.
8. Hauser, H., Pascher, I., Pearson, R. H., and Sundell, S., 1981. *Biochim. Biophys. Acta.* 650: 21-51
9. Stewart, J. J. P., 1990, Mopac 6.0 release notes, Frank J. Seiler Research laboratory, United States Air Force Academy.
10. van der Ploeg, P., Berendsen, H. J. C., 1982. Molecular dynamics simulation of a bilayer membrane. *Journal Chem. Phys.* 76: 3271-3276.
11. Stouch, T. R, 1993. Lipid membrane structure and dynamics studied by all atom molecular dynamics simulations of hydrated phospholipid bilayer. *Molecular Simulation.* 1: 335-362.
12. Lewis, B. A. and Engelman, D. M., 1983. *Journal Mol. Bio.* 166: 211.
13. McKarns, S. C., Hansch, C., Caldwell, W. S., Morgan, W. T., Moore, S. K., and Doolittle, D. J., 1997. *Fundamental and Applied Toxicology.* 36**:** 62-70 (1997).