Predicting Caco-2 Cell Permeation Coefficients of Organic Molecules Using Membrane-Interaction QSAR Analysis

Amit Kulkarni,† Yi Han, and A. J. Hopfinger*

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

Received October 15, 2001

A methodology termed *membrane−interaction QSAR* (MI-QSAR) *analysis* has been developed in order to predict the behavior of organic compounds interacting with the phopspholipid-rich regions of biological membranes. One important application of MI-QSAR analysis is to estimate ADME properties including the transport of organic solutes through biological membranes as a computational approach to forecasting drug intestinal absorption. A training set of 30 structurally diverse drugs, whose permeability coefficients across the cellular membranes of Caco-2 cells were measured, was used to construct significant MI-QSAR models of Caco-2 cell permeation. Cellular permeation is found to depend primarily upon aqueous solvation free energy (solubility) of the drug, the extent of drug interaction with a model phospholipid (DMPC) monolayer, and the conformational flexibility of the solute within the model membrane. A test set of eight drugs was used to evaluate the predictivity of the MI-QSAR models. The permeation coefficients of the test set compounds were predicted with the same accuracy as the compounds of the training set.

INTRODUCTION

Drug discovery programs generally focus on the development of orally active drugs. This is the least bothersome and, therefore, the preferred route of administration and often an absolute requirement from a marketing perspective. The majority of low molecular weight drugs are absorbed by passive diffusion from the gut. The extent of absorption is mainly dependent on dose, solubility/dissolution rate, and gut membrane permeability. Solubility/dissolution rate and permeability can compensate for one another in the overall bioavailability process. An unfavorable action of one mechanism may be compensated by a favorable behavior of the other.

An important and specific parameter for oral bioavailability is the transport of the drug across the intestinal epithelial cell barrier. One of the in vitro models, which has been shown to mimic this process, is a Caco-2 cell monolayer.[1] Caco-2 cells, a well-differentiated intestinal cell line derived from human colorectal carcinoma, display many of the morphological and functional properties of the in vivo intestinal epithelial cell barrier. (e.g. polarization and tight cellular junctions). Hence Caco-2 cells provide a tool that is potentially useful for rapidly evaluating the degree of transport of drugs across a cell barrier, for elucidating the mechanisms of drug transport, and for improving formulation strategies to enhance drug transport. Caco-2 cell models are used with regularity for determination of cellular transport properties, in both industry and academia, as a surrogate marker for in vivo intestinal permeability in humans.[2]

The increasing use of Caco-2 cell screening for oral bioavailability has led to a corresponding increased interest in understanding the mechanism of cellular permeability in this screen. While it seems relatively clear that passive diffusion usually underlies transport, understanding more details of the Caco-2 cell transport would (1) define the appropriate range of applications and solute chemistries of this screen, (2) identify the key properties of the organic solutes (drugs) responsible for cellular permeation behavior, and (3) permit construction of possible quantitative structure−activity relationship, QSAR, models to use as virtual screens to predict Caco-2 cell permeation and, more generally, oral bioavailability.

The ability of a molecule to permeate cell membranes by passive diffusion is primarily dependent on its partitioning into the membrane bilayer. The most frequently used physicochemical property to represent this partitioning, and the prediction of cellular permeability, is the Log of the (1-octanol/water) partition coefficient, LogP.[3,4] Efforts to correlate LogP to Caco-2 cell permeability coefficients have yielded mixed results. For example, Hilgers et al. have reported a sigmoidal relationship, whereas Artursson and Karlson observed a poor linear correlation.[2,4] Still, it is widely accepted that lipophilic drugs have better cellular permeability and absorption profiles than hydrophilic drugs. Yamashita and co-workers have shown that for Caco-2 cells drug permeability increases with increasing drug lipophilicity and that drug permeation has a good linear relationship with in vivo permeability.[5]

Other descriptors and modeling/QSAR work has been done to better understand Caco-2 cell permeability. The hydrogen bonding capacity and molecular surface properties of the solute have been used to construct correlation models for Caco-2 cell permeability.[6−8] The oral absorption potential of a drug can be predicted by in situ perfusion in a rat model with good results.[9] Unfortunately, such experimental ap-

* Corresponding author phone: (312)996-4816; fax: (312)413-3479; e-mail: hopfingr@uic.edu.

† Current address: Accelrys, Inc., 200 Wheeler Rd., Suite 2, Burlington, MA 01803.

10.1021/ci010108d CCC: $22.00
Published on Web 02/09/2002

proaches are costly and time-consuming and require the use of animals.

Pade and Stavchansky performed a study in establishing a relationship between drug permeability and solubility in vitro and the extent of drug absorption in vivo in humans.[10] These workers selected drugs with varying permeability and solubility measures. Effective permeability coefficients of the model drugs at 37 °C and pH of 7.2 were estimated using the Caco-2 cell line. It was seen that drugs with high permeability and solubility are completely absorbed (90% or higher). The findings from this study indicate that there is a strong link between permeability measured in Caco-2 cells, solubility, and fraction absorbed in humans.

The permeability of a solute to the phospholipid-rich regions of a cellular membrane has a nonspecific component with regard to the chemical structure of the solute that is largely reflected in its LogP. However, there is also a specific component to solute permeability as can be seen in the limitations of using LogP in constructing permeability QSAR models and the need for hydrogen bonding and surface features of the solute which are structure-specific. Structural specificity, in turn, usually limits the range of applicability of QSAR modeling to analogue training sets. The chemical and structural diversity in most training sets of measured Caco-2 cell permeability coefficients is likely the major source of the limited success in the QSAR analysis of such data sets.

Recently there has been a surge in computational efforts to compute ADME properties including Caco-2 cell permeation of structurally diverse compounds including drugs.[11] These new computational approaches remain focused on modeling structurally diverse solute data sets by dealing only with the properties of the solutes. A 'philosophy' has been adopted to get around the limitations of performing a QSAR analysis on a structurally diverse data set. The number of intramolecular solute properties computed is made as large as possible, and then some type of data reduction method is employed as part of the data fitting process in constructing the QSAR model. The idea behind this philosophy is that if enough solute features are included, the key intramolecular solute properties for describing multiple mechanisms of action (permeation) will be captured and built into the QSAR model without doing data overfitting. Of course, there is no way to know (1) if the right set of intramolecular solute features are included in the QSAR descriptor pool, and, if indeed, (2) any set of intramolecular solute descriptors exist that can capture the requisite mechanistic information by themselves. Moreover, once data reduction is performed it becomes exceedingly difficult to interpret the resultant QSAR model and to gain insight into mechanisms of action. The QSAR is in a particular data reduction space and not in actual physicochemical property space. Clearly, some type of structure-based design QSAR approach is needed in order to meaningfully handle the chemical and structural diversity of the solutes of the training sets encountered in constructing ADME property (Caco-2 cell permeation) QSAR models.

We have developed a methodology called *membrane-interaction, MI, QSAR* analysis where structure-based design methodology is combined with classic intramolecular QSAR analysis to model chemically and structurally diverse compounds interacting with cellular membranes.[12,13] In MI-QSAR analysis the assumption is made that the phospholipid regions of a cellular membrane constitute the "receptor" required in structure-based design that permits incorporation of structural and chemical diversity into a training set. A set of *membrane-solute intermolecular properties* are determined and added to the set of intramolecular solute QSAR descriptors to enlarge the trial QSAR descriptor pool and, ostensibly, to provide the information needed to incorporate chemical and structural diversity into the QSAR analysis.

More recently, we have augmented the descriptor pool used in MI-QSAR analysis to better model the general solute dissolution-aqueous solvation process. In fact, dissolution and aqueous solvation endpoint measures of test solute molecules can be studied independent of membrane uptake and transport. However, drug (solute) permeation of Caco-2 cells involves both dissolution-solvation as well as membrane interactions.

MI-QSAR analysis has been successfully applied to construct robust models of eye irritation for structurally diverse training sets.[12,13] It has subsequently occurred to us that MI-QSAR analysis is ideally suited to handle the construction of QSAR models for membrane uptake and transport measures such as Caco-2 cell permeability. The initial application of MI-QSAR analysis to this class of bioavailability problems is reported in this paper.

MATERIALS AND METHODS

A. Caco-2 Cell Permeation Coefficients. The dependent variable used in MI-QSAR analysis is the Caco-2 cell permeability coefficient, $P_{caco-2.}$ Yazdanian and co-workers performed permeability experiments on a data set of 38 structurally and chemically diverse drugs ranging in molecular weight from 60 to 515 amu and varying in net charge at pH 7.4.[14] Caco-2 cells, originating from human colorectal carcinoma, were obtained from American Tissue Culture Collection, Rockville, MD. The cells were grown at 37 °C in an atmosphere of 5% CO2 in Dulbecco's Modified Eagle Medium (DMEM). Caco-2 cells were then seeded at a density of 80 000 cells/cm^2 in 6-well plates on polycarbonate filters coated with rat tail collagen type I. All the permeability experiments were performed in an incubator at 37 °C and an atmosphere of 5% CO2. Experiments were performed under "sink" conditions where the concentration of the solute in the receiver side was less than 10% of the dose applied at all time points, hence minimizing the diffusion of the solute back from the receiving to donor side. The permeability coefficient was determined according to

$$P_{caco-2} = J(ACo)^{-1} \quad (1)$$

In eq 1 J is the rate of appearance of solute in the receiver chamber, Co is the initial concentration of the solute in the donor chamber, and A is the surface area of the filter.

Table 1 contains the P_{caco-2} values for 30 structurally diverse drugs used as the training set of compounds and eight drugs used as a test set. The construction of the training and test sets was accomplished by insisting that members of the test set be representative of all members of the training set in terms of the ranges of P_{caco-2} values, molecular weights, and structural and chemical diversities. Table 1 also contains a composite summary of the "% absorbed" of many of the drugs in the table. These data were compiled by search of

Table 1. Molecular Weight, Caco-2 Permeability Coefficient, and Corresponding Percent of Drug Absorbed (as Available) for the Drugs of the Training and Test Sets

drug	MW	permeability $\times 10^6$ (cm/s)	% absorbed
		Training Set	
diazepam	284.74	33.40	100
caffeine	194.19	30.80	100
phenytoin	252.27	26.70	90
alprenolol	249.35	25.30	93
testosterone	288.43	24.90	100
phencyclidine	243.39	24.70	
desipramine	266.39	24.20	95
metoprolol	267.37	23.70	95
progesterone	314.47	23.70	
salicylic acid	138.12	22.00	100
clonidine	230.10	21.80	100
corticosterone	346.47	21.20	100
indomethacin	357.79	20.40	100
chlorpromazine	318.86	19.90	90
nicotine	162.23	19.40	100
estradiol	272.39	16.90	
pindolol	248.32	16.70	95
hydrocortisone	362.47	14.00	89
timolol	316.42	12.80	72
dexamethasone	392.47	12.20	100
scopolamine	303.36	11.80	100
dopamine	153.18	9.33	
labetalol	328.41	9.31	90
bremazocine	315.45	8.02	
nadolol	309.40	3.88	
atenolol	266.34	0.53	50
terbutaline	225.29	0.47	73
ganciclovir	255.23	0.38	3
sulfasalazine	398.39	0.30	13
acyclovir	225.21	0.25	20
		Test Set	
aminopyrine	231.3	36.5	100
propranolol	259.35	21.80	90
warfarin	308.33	21.10	98
meloxicam	351.39	19.50	90
zidovudine	267.24	6.93	100
urea	60.06	4.56	
sucrose	342.30	1.71	
mannitol	182.17	0.38	16

the literature. It can be seen from a comparison of the P_{caco-2} and "% absorbed" that P_{caco-2} is indeed indicative of in vivo drug absorption/uptake.

B. Building Solute Molecules and a DMPC Monolayer. All the solute molecules of the training and test sets, see Table 1, were built using the Chemlab-II molecular modeling package.[15] A single dimyristoylphosphatidylcholine (DMPC) molecule was built using HyperChem from available crystal structure data.[16,17] The AM1 Hamiltonian in Mopac 6.0 was used for the estimation of partial atomic charges on all molecules.[18]

Dimyristoylphosphatidylcholine (DMPC) was selected as the model phospholipid in this study. The structure of a DMPC molecule is shown in Figure 1. An assembly of 25 DMPC molecules ($5 \times 5 \times 1$) in (x,y,z) directions, respectively, was used as the model membrane monolayer. The size of the monolayer simulation system was selected based on the work done by van der Ploeg and Berendsen.[19] These workers performed a molecular dynamic simulation (MDS) study for two decanoate bilayers having ($2 \times 8 \times 2$) and ($2 \times 16 \times 2$) phospholipid molecules. It was found that the estimated order parameters for these two model bilayers agree with one another suggesting that the smaller assembly is adequate for modeling short-range properties. Other researchers have obtained similar geometric and energetic equilibrium property values with regard to the size of model simulation system permitting a minimum effective size (number of phospholipids) of the monolayer to be defined.[20] Additional information regarding construction of the model monolayer used in this MI-QSAR analysis is given in refs 12 and 13.

To prevent unfavorable van der Waals interactions between a solute molecule and the membrane DMPC molecules, the "center" DMPC molecule, located at position (x,y) = (3,3) of the 5×5 DMPC monolayer model, was removed from the monolayer, and a test solute moleculewas inserted in the space created by the missing DMPC molecule. Each of the test solute molecules of the permeation 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 headgroup 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 headgroup region, (2) solute molecule between the headgroup region and the aliphatic chains, and (3) solute molecule in the tail region of the aliphatic chains.

The lowest energy geometry of the solute molecule in the monolayer was sought using each of the three trial solute positions. The three different initial MDS positions of urea, one of the training set solute molecules, are shown in Figure 2a to illustrate this modeling procedure. The energetically most favorable geometry of this solute molecule in the model DMPC monolayer is shown in Figure 2b.

C. Molecular Dynamic Simulations. MDS were carried out using the Molsim package with an extended MM2 force field.[21] The selection of the simulation temperature was based on the phase transition temperature for DMPC, which is 297 K.[22] A simulation temperature of 311 K was selected since it is body temperature, and it is also above the DMPC phase transition temperature. Temperature was held constant in the MDS by coupling the system to an external fixed temperature bath.[23] The trajectory step size was 0.001 ps over a total simulation time of 20 ps for each solute of the training and test sets. Every membrane-solute system, for the solutes of the training and test sets of this study, reached equilibrium by 1500 trajectory steps, that is 1.5 ps. The MDS trajectory for urea is shown in Figure 3, and the lowest energy equilibrium step on the trajectory, near 4500 steps, corresponds to the urea-membrane complex geometry shown in Figure 2b. Two-dimensional periodic boundary conditions, corresponding to the "surface plane" of the monolayer, were employed ($a = 32$ A^2, $b = 32$A^2, $c = 80$ A^2, and $\gamma = 96.0°$) for the DMPC molecules of the monolayer model but not the test solute molecule. The angle γ is the angle an extended DMPC molecule makes with the "planar surface" of the monolayer. Only a single solute molecule was explicitly considered in each MDS. Each of the solute molecules was placed at each of the three different positions in the monolayer, as described above, with the most polar portion of the solute "facing" toward the headgroup region. Additional details of the membrane-solute MDS can be found in refs 12 and 13.

D. Calculation of Descriptors. Both *intramolecular* physicochemical properties and features of the solute mol-

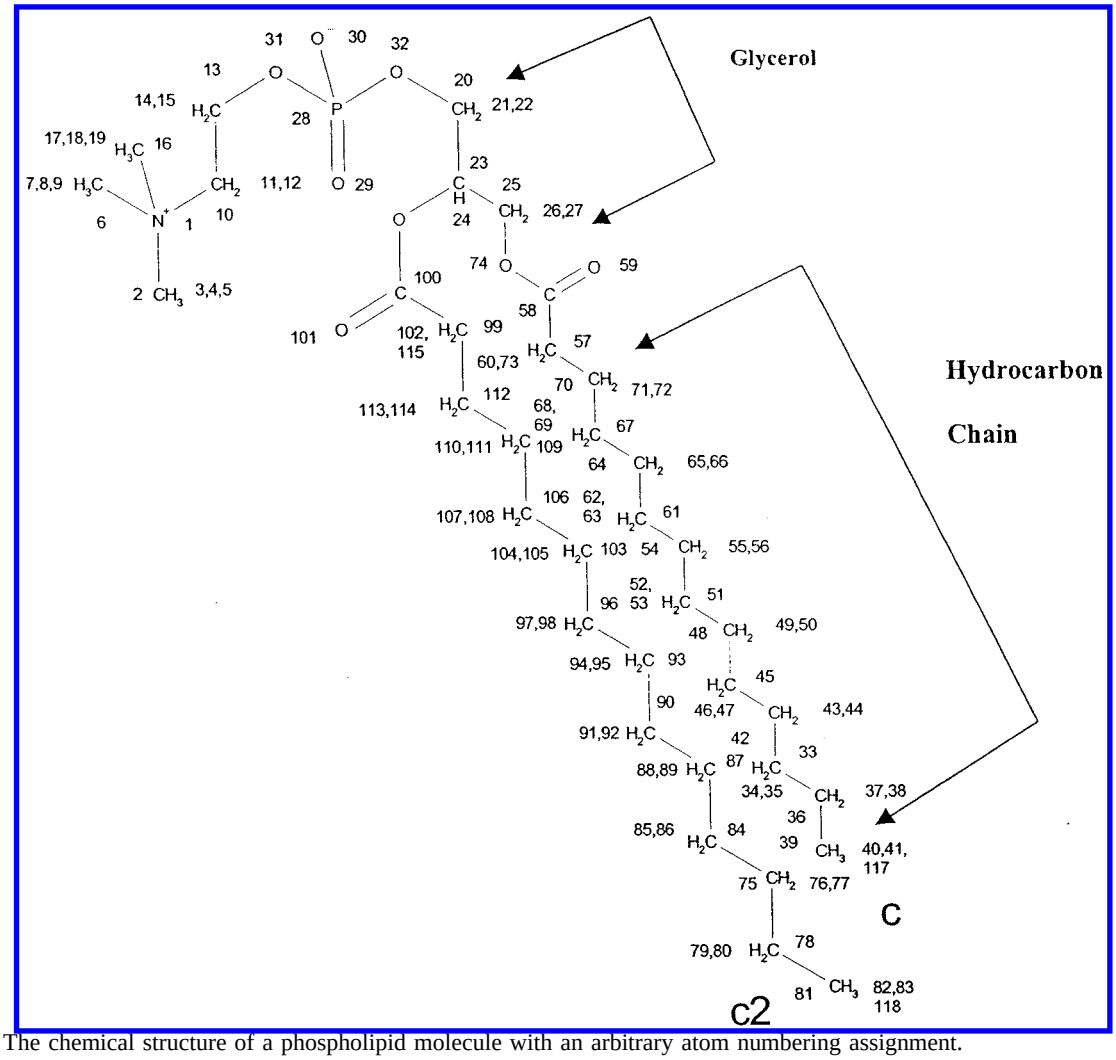

Figure 1. The chemical structure of a phospholipid molecule with an arbitrary atom numbering assignment.

ecules and *intermolecular* solute-membrane interaction properties were calculated. "Properties" and "features" will both be referred to as descriptors from this point forward as they constitute the trial pool of independent variables used to build the QSAR models. The descriptors used in the MI-QSAR analysis can also be divided into (a) *solute aqueous dissolution and solvation* descriptors, (b) *solute-membrane interaction* descriptors, and (c) *general intramolecular solute* descriptors. The tables reporting the trial pool of descriptors used in the Caco-2 cell permeation MI-QSAR modeling employ both classifications of the descriptors.

The *general intramolecular solute descriptors* included as part of the trial descriptor pool are defined in Table 2. The term *general* is used because solute descriptors in this class may be useful in describing different aspects of the bioavailability (in this case the Caco-2 cell permeation process) of a solute. It should noted that $F(H_2O)$, F(OCT), and LogP, the aqueous and 1-octanol solvation free energies of the solutes and the corresponding 1-octanol/water partition coefficient, respectively, are computed using intramolecular computational methods. This is also true for E(coh), T_M, and T_G, the cohesive energy and the hypothetical crystal-melt and glass transition temperatures of the solutes, which are used to estimate solute dissolution properties. However, all of these descriptors are intermolecular properties, the first three relating to solute solvation, and the last three to solute dissolution. Therefore, these descriptors are classified as *solvation and dissolution intermolecular descriptors* and are reported as Part B of Table 3.

E(coh) is a measure of the energy required to remove a molecule from being surrounded by other molecules identical to itself. TM measures the crystal packing strength of a molecule, and TG measures the amorphous packing strength of a molecule. The assumption is made here that E(coh), TM, and TG, taken in composite, can be used to describe the dissolution behavior of any solute when developing an MI-QSAR model.

Some of the *intermolecular solute-membrane interaction descriptors* can be extracted directly from the MDS trajectories and are listed in Part A of Table 3. These particular intermolecular descriptors were calculated using the most stable (lowest total potential energy) solute-membrane geometry realized from MDS sampling of the three initial positions, see Figure 2a, for each of the solutes. Figure 3 shows a plot the total potential energy vs MDS trajectory from which the most energetically favorable position of urea, one of the solutes of the test set, in DMPC is identified and shown in Figure 2b. It can also be seen in Figure 3 that the membrane-urea system attains energy equilibrium after about 1 ps of MDS. That is, the energy vs time plot is, on average, "flat" after 1 ps.

The intermolecular membrane-solute descriptors derived from MDS trajectories are given in Table 3A. $E_{XY}(Z)$ is a set of energy terms where Z denotes the electrostatic

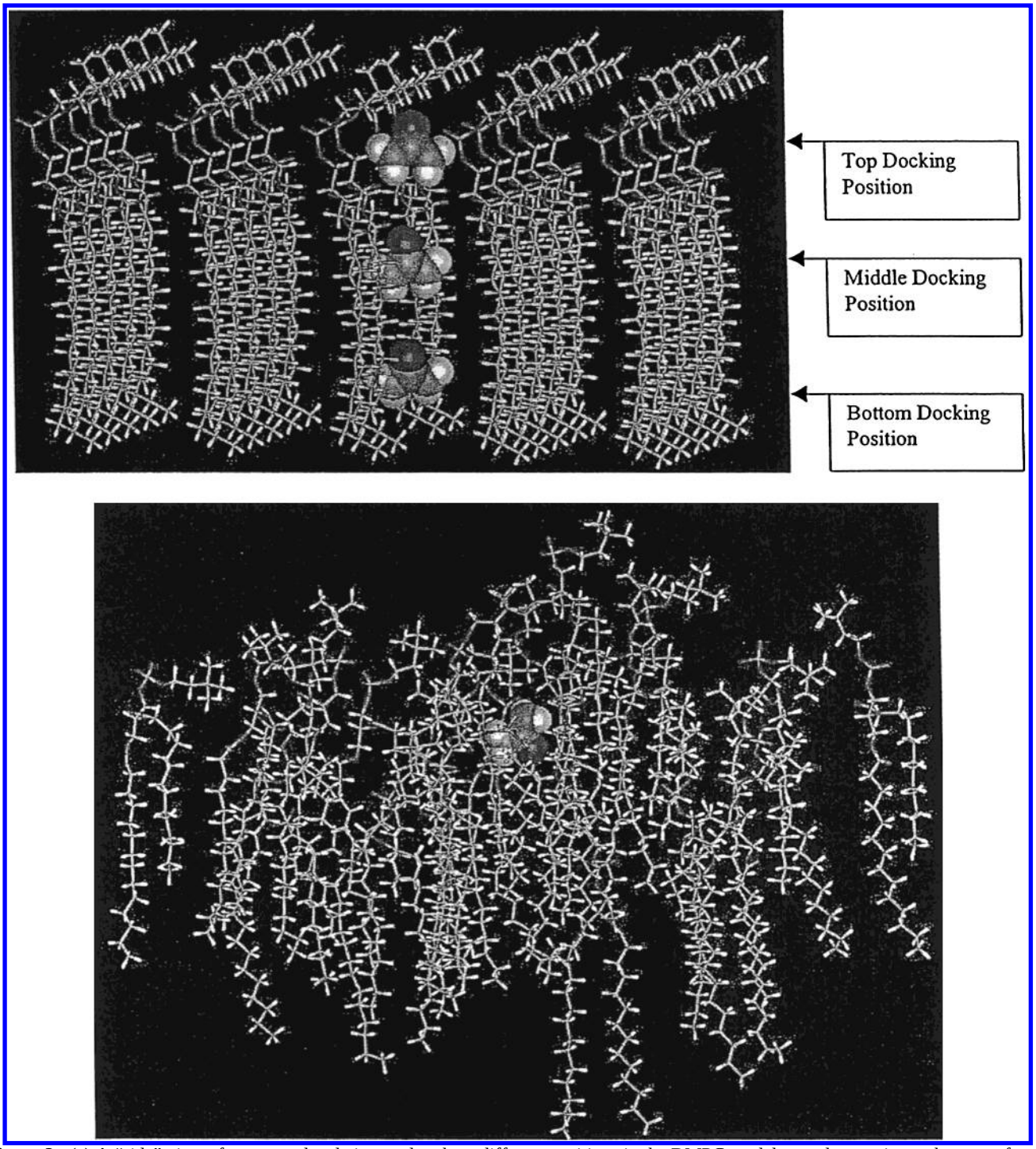

Figure 2. (a) A "side" view of a urea molecule inserted at three different positions in the DMPC model monolayer prior to the start of each of the three corresponding MDS used in the MI-QSAR modeling of urea. (b) The lowest energy geometry of a DMPC-urea complex in the MDS.

interaction energy, 1,4-nonbonded interaction energy, general van der Waals interaction energy, hydrogen bond energy, torsion energy, or combinations thereof between X and Y where X and Y can each be the solute, S, and/or the membrane, M, at the minimum total potential energy state of the system. $\Delta E_{XY}(Z)$ has the same meaning to its symbols as $E_{XY}(Z)$ but refers to the change in the particular energy term due to the uptake of the solute into the membrane at the minimum total potential energy position of the membrane -solute system. $E_{TT}(Z)$ and $\Delta E_{TT}(Z)$ are the same as $E_{XY}(Z)$ and $\Delta E_{XY}(Z)$, respectively, but the energy contributions are for the entire solute-model membrane complex.

Other *intermolecular solute-membrane descriptors* computed from the MDS and reported in Part A of Table 3 are the conformational entropy and change in conformational entropy due to solute uptake into the membrane, average depth of the solute in the monolayer, and change in membrane density upon solute uptake. Details about the methods and algorithms used to compute these descriptors can be found in refs 12 and 13.

E. Construction and Testing of MI-QSAR Models. MI-QSAR models were constructed using the genetic function approximation, GFA, which is a multidimensional optimization method based on the genetic algorithm paradigm.[24,25]

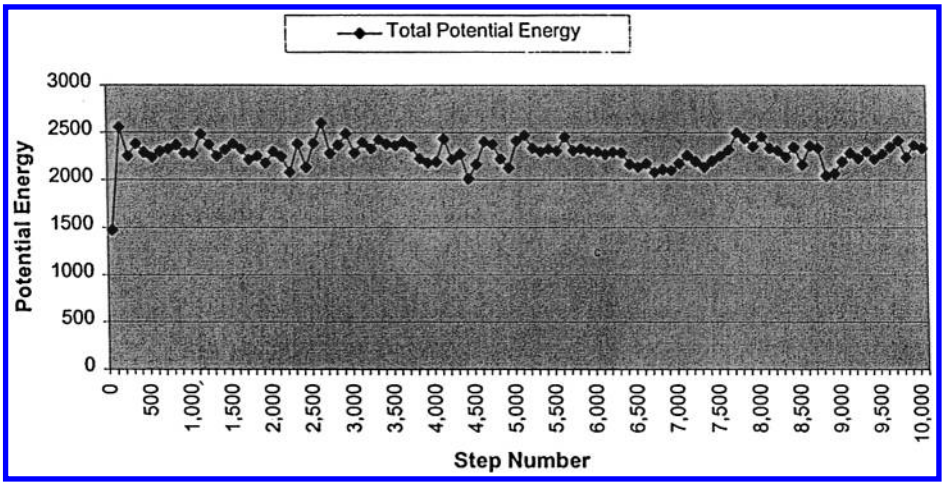

Figure 3. Total potential energy (kcal/mol) versus time (ps) trajectory plot for urea embedded in the model DMPC monolayer.

Table 2. General Intramolecular Solute Descriptors Used in the Trial MI-QSAR Descriptor Pool

HOMO	highest occupied molecular orbital energy
LUMO	lowest occupied molecular orbital energy
Dp	dipole moment
Vm	molecular volume
SA	molecular surface area
Ds	density
MW	molecular weight
MR	molecular refractivity
N(hba)	number of hydrogen bond acceptors
N(hbd)	number of hydrogen bond donors
N(B)	number of rotatable bonds
JSSA (X)	Jurs- Stanton surface area descriptors
Chi-N, Kappa-M	Kier and Hall topological descriptors
Rg	radius of gyration
PM	principle moment of inertia
Se	conformational entropy
Q(I)	partial atomic charge densities

Statistical significance in the optimization of a QSAR model using GFA is based on Friedman's lack of fit (LOF) measure.[26] The LOF measure is designed to resist overfitting which is a problem often encountered in constructing statistical models. Since the number of descriptors available in MI-QSAR analysis normally exceeds the number of observations (test compounds), the ability to prevent overfitting using GFA is critical to the successful construction of a statistically significant MI-QSAR model.

One of the parameters that can be adjusted in the GFA is the smoothing factor, SF. Adjusting the SF adjusts the model size, that is, the number of independent variables in the MI-QSAR model. The SF values were varied to generate a family of related MI-QSAR models representing the best 1−6-term models for the training set and for the training set and the test set combined. 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 multidimensional linear regression (MLR) can be used in GFA to establish functional data fits. MLR was used in this MI-QSAR Caco-2 cell permeability study.

To test and to validate the MI-QSAR models, the dependent variable, P_{caco-2}, was randomly "scrambled" with respect to the set of independent variables (descriptor set) of the compounds to see if meaningful correlations (QSARs) could be found among the scrambled data sets.[27] The absence of any significant correlation for each of the scrambled data sets is taken as evidence of the significance of the MI-QSAR models with respect to the original, nonscrambled data set. 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 optimization process of the GFA analysis. No significant cross-correlations were found for the descriptors of the best MI-QSAR models.

RESULTS

The best MI-QSAR models for Caco-2 cell permeability realized by considering the combination of general intramolecular solute, intermolecular dissolution/solvation-solute, and intermolecular membrane-solute descriptors are presented as a function of the number of terms, that is descriptors, included in a given MI-QSAR model:

1-term model

$$P_{caco-2} = 37.39 + 0.73F(H2O) \quad (2)$$

$$N = 30,\ R^2 = 0.75,\ Q^2 = 0.71$$

2term model

$$P_{caco-2} = 30.58 + 0.54F(H2O) + 0.07\Delta E_{TT}(hb) \quad (3)$$

$$N = 30,\ R^2 = 0.78,\ Q^2 = 0.72$$

3-term model

$$P_{caco-2} = 31.87 + 0.72F(H2O) + 0.07\Delta E_{TT}(hb) - 0.26E_{SS}(hb) \quad (4)$$

$$N = 30,\ R^2 = 0.80,\ Q^2 = 0.74$$

4-term model

$$P_{caco-2} = -14.62 + 0.71F(H2O) + 0.07\Delta E_{TT}(hb) - 0.26E_{SS}(hb) + 0.06E_{TT}(14) \quad (5)$$

$$N = 30,\ R^2 = 0.82,\ Q^2 = 0.75$$

Table 3. Intermolecular Interaction Descriptors in the Trial MI-QSAR Descriptor Pool: Part A Includes the Membrane-Solute Interaction Descriptors and Part B Lists the Intermolecular Dissolution and Solvation Descriptors of the Solute

Part A

membrane-solute descriptors − symbols	description of the membrane-solute descriptors
$\langle F(total)\rangle$	average total free energy of interaction of the solute and membrane
$\langle E(total)\rangle$	average total interaction energy of the solute and membrane
E_{INTER} (total)	interaction energy between the solute and the membrane at the total intermolecular system minimum potential energy
$E_{XY}(Z)$	Z = 1,4-nonbonded, general Van der Waal, electrostatic, hydrogen bonding, torsion and combinations thereof energies at the total intermolecular system minimum potential energy. X, Y can be the solute, S, and/or membrane, M
ΔE_{XY} (Z)	change in the Z = 1,4-nonbonded, general Van der Waal, electrostatic, hydrogen bonding, torsion and combinations thereof energies due to the uptake of the solute to the total intermolecular system minimum potential energy. X, Y can be the solute, S, and/or membrane, M
$E_{TT}(Z)$	Z = 1,4-nonbonded, general Van der Waal, electrostatic, hydrogen bonding, torsion and combinations thereof energies of the total [solute and membrane model] intermolecular minimum potential energy
ΔE_{TT} (Z)	change in the Z = 1,4-nonbonded, general Van der Waal, electrostatic, hydrogen bonding and combinations thereof of the total [solute and membrane model] intermolecular minimum potential energy
ΔS	change in entropy of the membrane due to the uptake of the solute
S	absolute entropy of the solute-membrane system
$\Delta\rho$	change in density of the model membrane due to the permeating solute
$\langle d\rangle$	average depth of the solute molecule from the membrane surface

Part B

dissolution and solvation − solute descriptors −symbols	description of the dissolution and solvation − solute descriptors
F(H2O)	the aqueous solvation free energy
F(OCT)	the 1-octanol solvation free energy
Log(P)	the 1-octanol/water partition coefficient
E(coh)	the cohesive packing energy of the solute molecules
T_M	the hypothetical crystal-melt transition temperature of the solute
T_G	the hypothetical glass transition temperature of the solute

5-term model

$$P_{caco-2} = -16.16 + 0.73F(H2O) + 0.06\Delta E_{TT}(hb) - 0.25E_{SS}(hb) + 0.07E_{TT}(14) - 0.12E_{TT}(tor) \quad (6)$$

$$N = 30,\ R^2 = 0.83,\ Q^2 = 0.74$$

6-term model

$$P_{caco-2} = -40.50 + 0.65F(H2O) + 0.06\Delta E_{TT}(hb) - 0.19E_{SS}(hb) + 0.10E_{TT}(14) - 0.03E_{TT}(tor) - 5.61\chi_3 \quad (7)$$

$$N = 30,\ R^2 = 0.86,\ Q^2 = 0.77$$

N is the number of compounds, R^2 is the coefficient of determination, and Q^2 is the cross-validated coefficient of determination.

The descriptors found in the best MI-QSAR models are as follows:

(1) F(H2O) is the aqueous solvation free energy.

(2) χ_3 is a topological index measuring the size and shape of a molecule.[28]

(3) E_{SS}(hb) is the intramolecular hydrogen bonding energy of the solute molecule *when it is in the lowest membrane-solute interaction state within the membrane.*

(4) ΔE_{TT}(hb) is the *change* in the hydrogen bonding energy of the entire membrane-solute system for the solute relocated from free-space to the position corresponding to the lowest solute−membrane interaction energy state of the model system.

(5) E_{TT}(14) is the 1,4-van der Waals plus electrostatic interaction energy of the entire membrane-solute system for the solute located at the position corresponding to the lowest solute membrane interaction energy state of the model system. The range in values of this descriptor over the training and test sets is 770−920 kcal/mol, a very large set of energies. However, there are over 700 torsion angles associated with E_{TT}(14). Thus, the *average* E_{TT}(1,4) per torsion angle is only about 1.1−1.3 kcal/mol.

(6) E_{TT}(tor) is the torsion energy of the entire membrane-solute system for the solute located at the position corresponding to the lowest solute-membrane interaction energy state of the model system. This descriptor is also large in energy having a range of values of 150−230 kcal/mol across the training and test sets of compounds. Again, for the more than 700 torsion angles associated with this descriptor, the *average* value of E_{TT}(tor) per torsion angle is only 0.20−0.33 kcal/mol.

The values of the six descriptors found in the 1−6-term MI-QSAR models for each compound in the training and test sets are given in Table 4.

The observed and predicted, using the 3−6-term MI-QSAR models, Caco-2 cell permeation coefficients of the test and training set compounds are listed in Table 5 and plotted in Figure 4. Clonidine, metoprolol, corticosterone, and aminopyrine are observed to permeate better than

Table 5. Observed and Predicted Caco-2 Permeability Coefficients for the 3−6-Term MI-QSAR Models

structure name	obsd $P_{caco-2} \times 10^6$	3-term	4-term	5-term	6-term
		Training Set			
diazepam	33.4	26.89	27.84	27.99	29.25
caffeine	30.8	27.91	25.73	25.99	25.43
phenytoin	26.7	21.97	21.95	23.50	23.82
alprenolol	25.3	19.99	20.26	21.40	22.03
testosterone	24.9	23.98	24.30	25.87	26.35
phencyclidine	24.7	29.22	27.95	26.86	27.15
desipramine	24.2	23.13	21.88	23.78	23.44
metoprolol	23.7	16.38	16.15	17.04	17.87
progesterone	23.7	31.82	31.29	31.88	31.75
salicylicacid	22	22.38	21.43	22.06	21.90
clonidine	21.8	17.27	15.87	14.58	15.48
corticosterone	21.2	16.53	15.64	14.82	15.44
indomethacin	20.4	18.40	20.04	20.51	22.63
chlorpromazine	19.9	24.63	22.65	23.94	23.41
nicotine	19.4	27.28	25.57	24.73	24.87
estradiol	16.9	14.34	13.94	15.39	16.14
pindolol	16.7	9.97	10.60	12.02	13.13
hydrocortisone	14	11.83	12.20	13.87	14.24
timolol	12.8	12.15	11.47	11.58	12.25
dexamethasone	12.2	10.68	14.01	12.95	15.89
scopolamine	11.8	16.91	18.83	19.41	13.54
dopamine	9.33	10.87	10.21	9.28	10.88
labetalol	9.31	8.90	7.56	9.03	8.33
bremazocine	8.02	12.77	13.63	12.70	6.39
nadolol	3.88	4.83	5.26	5.22	6.71
atenolol	0.53	3.78	2.17	3.56	3.29
terbutaline	0.47	7.18	4.57	4.76	4.50
ganciclovir	0.38	0.44	−0.90	−1.69	−2.23
sulfasalazine	0.3	4.66	1.77	1.83	2.36
acyclovir	0.25	1.95	1.72	2.56	2.88

structure name	obsd BA	3-term	4-term	5-term	6-term
		Test Set			
aminopyrine	36.5	25.56	27.20	26.01	28.25
propranolol	21.8	14.98	14.10	15.09	15.26
warfarine	21.1	19.23	21.08	20.84	23.16
meloxicam	19.5	21.53	26.84	26.60	28.61
zidovudine	6.93	12.84	10.80	10.36	10.67
urea	4.56	4.51	4.91	5.95	6.53
mannitol	0.38	−2.10	−0.27	0.07	0.70
sucrose	1.71	−1.87	2.22	1.50	−1.39

Table 4. Values of the Six Descriptors Found To Be the Significant MI-QSAR Terms in Eqs 2−7[a]

structure name	E_{TT}(tor)	E_{TT}(14)	E_{SS}(hb)	ΔE_{TT}(hb)	χ_3	F(H2O)
			Training Set			
diazepam	196.9	847.4	0.0	0.0	0.0	−6.87
caffeine	180.3	792.0	0.0	0.0	0.0	−5.47
phenytoin	166.2	826.4	−1.8	−23.6	0.0	−11.89
alprenolol	167.7	830.9	−8.9	−6.0	0.0	−18.99
testosterone	168.2	833.9	0.0	−18.0	0.0	−9.04
phencyclidine	212.4	808.9	0.0	0.0	0.0	−3.67
desipramine	150.3	806.0	−0.9	−7.2	0.0	−11.66
metoprolol	169.4	820.2	−6.0	−13.3	0.0	−22.16
progesterone	185.3	823.1	0.0	0.0	0.0	−0.07
salicylicacid	173.8	809.9	−10.5	−7.6	0.0	−16.13
clonidine	215.3	798.9	0.0	−40.8	0.0	−15.97
corticosterone	208.3	806.4	−7.1	−48.6	0.0	−18.74
Indomethacin	188.1	855.6	−1.4	−6.8	0.0	−18.42
chlorpromazine	158.4	794.1	0.0	0.0	0.0	−10.00
nicotine	203.7	800.1	0.0	0.0	0.0	−6.34
estradiol	163.7	815.5	0.0	−39.4	0.0	−20.15
pindolol	169.6	829.9	−6.5	−61.7	0.0	−26.24
hydrocortisone	160.4	825.6	−15.6	−51.0	0.0	−28.04
timolol	178.7	808.7	−15.1	−21.7	0.0	−30.43
dexamethasone	230.8	877.4	−14.7	−64.4	0.0	−27.93
scopolamine	185.1	859.2	−6.4	−7.6	1.4	−22.16
dopamine	201.1	809.4	−5.7	−25.4	0.0	−28.43
labetalol	149.5	792.9	−25.9	−45.3	0.0	−36.37
bremazocine	216.6	836.4	−3.2	−48.3	1.5	−22.57
nadolol	187.2	823.4	−18.3	−50.4	0.0	−38.74
ntenolol	168.9	783.4	−7.5	−123.0	0.0	−28.82
terbutaline	172.0	770.3	−13.8	−54.9	0.0	−33.38
ganciclovir	204.1	783.3	−35.7	−126.0	0.0	−43.23
sulfasalazine	164.3	766.8	−7.5	−22.8	0.0	−37.92
acyclovir	183.8	805.9	−16.6	−127.4	0.0	−34.13
			Test Set			
aminopyrine	225.58	859.5	0	0	0	−8.72
propranolol	171.17	805.93	−6.55	−46.43	0	−20.89
warfarine	203.19	859.62	−3.49	−5.94	0	−18.10
meloxicam	217.59	917.53	−39.16	−20.45	0	−26.24
zidovudine	187.44	785.1	−8.4	−31.26	0	−26.08
urea	203.81	816.94	0	−186.09	0	−18.60
mannitol	186.59	838.82	−48.12	−102.16	0	−53.67
sucrose	205.11	866.78	−141.11	−132.76	0	−83.58

[a] See text and Tables 2 and 3 for definitions of the terms. All energies are in kcal/mol.

predicted by each of the MI-QSAR models, while nicotine and progesterone have a lower permeation coefficient than are predicted by any of the models. Nevertheless, none of the compounds in either the training or test sets are outliers for the 3−6-term MI-QSAR models. Figure 5 contains plots of R^2 and Q^2 for the training set and R^2 for the combined training and test sets, the *full* set, as a function of the number of descriptor terms for the 3−6-term MI-QSAR models. R^2, for both the training and full sets, increases with increasing number of descriptor terms. However, Q^2 dips in value for the 5-term model, perhaps suggesting overfitting is being approached with the 5- and 6-term models for the training set.

It appears from an analysis of eqs 2−7 that F(H2O) in the 1-term model accounts for much of the variance of P_{caco-2} across the training set. Nevertheless, the descriptors of the 2−6-term MI-QSAR models are all membrane-solute interaction properties and, therefore, are judged as being important in characterizing the *mechanism* of solute-membrane permeation. A composite analysis of all the MI-QSAR models, eqs 2−7, suggests that the 3-term MI-QSAR model captures the essential features of the postulated *mechanism* responsible for solute-membrane permeability as represented by P_{caco-2} values, see the **Discussion** section. The 3-term model does not represent a distinctly large statistical improvement over the 2-term model but rather includes descriptors indicative of each of the three components of the postulated *mechanism of permeation*.

The descriptors of 4−6-term MI-QSAR models successively refine the 3-term model fitting to the training set. The possible significance of the descriptors added in the 4−6-term MI-QSAR models to further revealing the essential mechanism of Caco-2 cell permeation can only be ascertained by consideration of an expanded training set. The interpretation that the 4−6-term MI-QSAR models are successive refinements of the "basic" 3-term MI-QSAR model is also supported by the mathematical forms of the MI-QSAR models. The [n + 1]-term MI-QSAR model can be viewed as essentially the [n]-term model with one new additional descriptor. The regression coefficients of corresponding descriptor terms across all of the MI-QSAR models are remarkably similar to one another, which indicates their respective roles in predicting P_{caco-2} are about the same in each MI-QSAR model irrespective of the number of descriptor terms in the model.

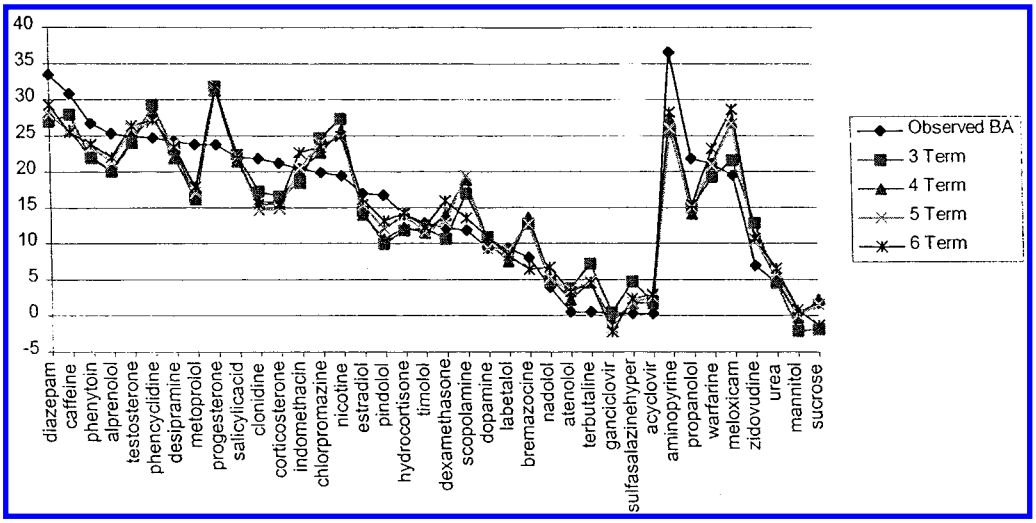

Figure 4. The observed vs predicted Caco-2 cell permeability coefficients for all the compounds of the training and test sets as a function of the size (number of descriptor terms, *X* axis) in the corresponding MI-QSAR model.

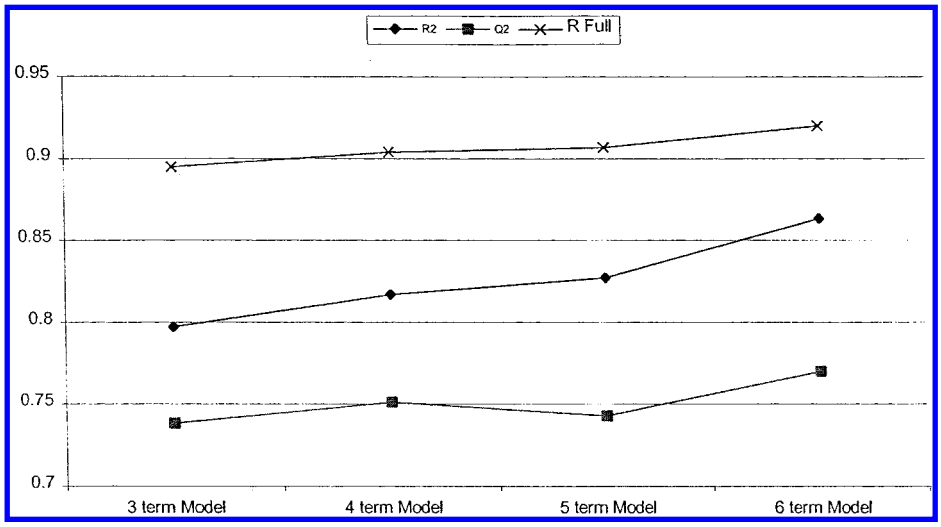

Figure 5. A diagnostic plot of the MI-QSAR models: *R2* is the correlation coefficient and Q2 is the cross-validation correlation coefficient of the best x-term model, where x is plotted on the *X*-axis for the 30 compounds of the training set. *R Full* is the correlation coefficient between the predicted and the observed value for all 38 compounds, that is the (training + test) compounds.

A test set of eight solute compounds was constructed from the parent Caco-2 cell permeation coefficient data set as one way to attempt to validate the MI-QSAR models. The drugs (solute molecules) of the test set were selected so as to span the entire range in Caco-2 cell permeability for the composite training set. The observed and predicted P_{caco-2} values for this test set are given at the bottom of Table 5 and plotted as the last eight compounds in Figure 4. There are no outliers, but aminopyrine and propranolol, compounds 1 and 2 of the test set, are predicted to have a lower permeability coefficients than observed. Conversely, meloxican has a higher observed P_{caco-2} value than is computed from any of the MI-QSAR models.

DISCUSSION

The aqueous solvation free energy, F(H2O) has been shown to correlate to aqueous solubility[12,13] as would be expected. Increasingly negative F(H2O) values corresponds to increasing aqueous solubility of a solute. In the P_{caco-2} MI-QSAR models it is seen that F(H2O) is positively correlated to P_{caco-2}. This relationship indicates that water soluble compounds will have lower permeability coefficients than hydrophobic compounds. This observation is similar to those found in the literature where Log P has been shown to have a relationship to Caco-2 cell permeability. An increase in Log P, reflecting an increase in lipophilicity, often corresponds to an increase in Caco-2 cell permeability. However, the relationship between Log P and Caco-2 cell permeability is not well defined. Some researchers report a sigmoidal relationship, while others report a poor linear relationship.[2,4] A significant linear relationship between F(H2O) and P_{caco-2} is seen in the MI-QSAR models reported here starting with the 1-term MI-QSAR model. Our interpretation of this relationship is that the other descriptors of the MI-QSAR models, *which focus on explicit membrane-solute interactions*, are not considered in the models/relationships of other workers. Hence, the models developed by other workers necessarily contain "noise" in the Log P−Caco-2 cell permeation comparisons and relationships.

ΔE_{TT}(hb) is the difference in the total hydrogen bond energy of the solute in the membrane minus the solute being in free space and the membrane by itself. No hydrogen bonding can occur within, or between, DMPC molecules.

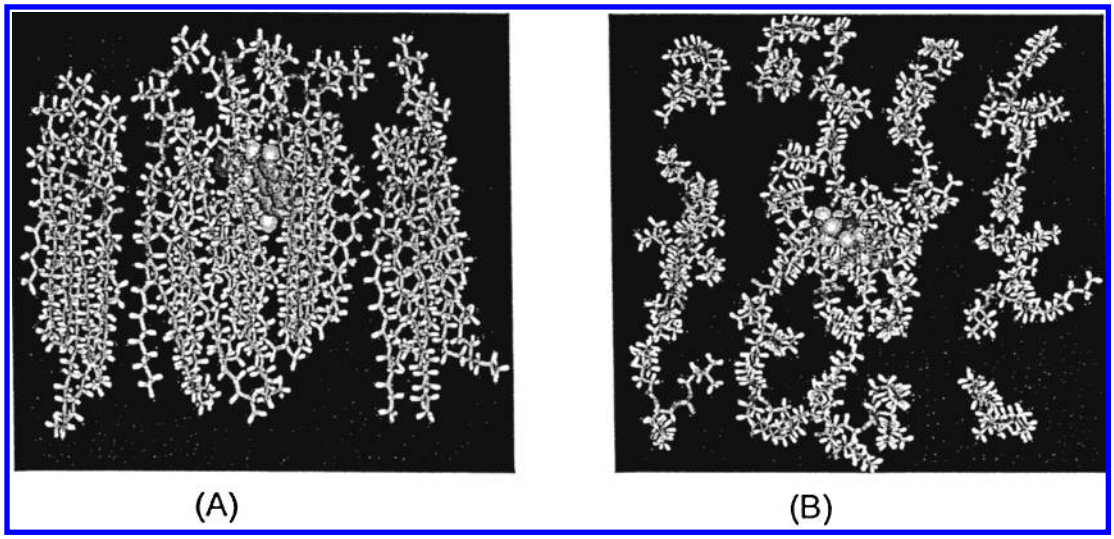

Figure 6. (A) Side view and (B) top view of the most stable structure for the model DMPC membrane−acyclovir complex. Acyclovir is one of the least permeable compounds of the training set.

Thus, the hydrogen bond energy of the membrane by itself is zero and

$$\Delta E_{TT}(hb) = E_{SS}(hb) - E'_{SS}(hb) + E_{MS}(hb) \qquad (8)$$

where $E'_{SS}(hb)$ is the intramolecular solute hydrogen bonding energy for the solute in free-space. In the MI-QSAR models containing $\Delta E_{TT}(hb)$, eqs 3−7, the regression coefficients of this descriptor term are positive and about equal. Thus, if intramolecular hydrogen bonding of the solute *decreases* upon uptake into the membrane, $E_{SS}(hb)$, and/or *increases* for the solute in free space, $E'_{SS}(hb)$, the permeation coefficient of the solute will increase. A decrease in intramolecular solute hydrogen bonding should correspond to an increase in the conformational flexibility of the solute. Solute conformational flexibility within the membrane is very important for high permeability as other MI-QSAR model descriptors, see below, also indicate. However, while $\Delta E_{SS}(hb)$ is the preferred descriptor with FH2O in a 2-term MI-QSAR model, see eq 3, $E_{SS}(hb)$ is the next preferred descriptor and is found in the best 3-term MI-QSAR model, eq 4. Thus, in eqs 4−7 the terms

$$\{a\Delta E_{TT}(hb) - bE_{SS}(hb)\} = \{[a - b]E_{SS}(hb) - aE'_{SS}(hb) - aE_{MS}(hb)\} \qquad (9)$$

are always present indicating the most important contribution of the 3-term MI-QSAR model to refining the 2-term MI-QSAR model is to correct the statistical weighting of $E_{SS}(hb)$ in the 2-term model since it is inherent to the $\Delta E_{TT}(hb)$ descriptor.

If intramolecular hydrogen bonding of the solute decreases upon uptake into the membrane, solute-membrane hydrogen bonding will likely increase. According to eq 8, and the MI-QSAR models, an increase in solute-membrane hydrogen bonding will diminish solute permeability. Thus, the joint interpretation of $\Delta E_{TT}(hb)$ and $E_{SS}(hb)$ in the MI-QSAR models is that they capture the balance of hydrogen bonding of the solute with itself in and out of the membrane, and with the DMPC molecules of the membrane, that is at play in the solute-membrane permeation process.

Solute and DMPC conformational flexibility is represented by $E_{TT}(14)$ in eqs 5−7 and $E_{TT}(tor)$ of eqs 6 and 7. $E_{TT}(14)$ is the van der Waals and electrostatic energies associated with each set of atoms separated exactly, and only, by *one* torsion angle in the solute molecule and all the DMPC molecules of the model membrane. This contribution to the total conformational energy measures the composite rigidity of an average torsion rotation *of the entire solute-membrane system*. As $E_{TT}(14)$ increases the molecules of the membrane-solute system, on average, are moving away from minimum energy conformer states and exploring more conformational states That is, the molecules are expressing greater flexibility. This greater flexibility results in a higher permeation coefficient of the solute molecule based on the positive regression coefficients for $E_{TT}(14)$ in eqs 5−7. Presumably, an increase in conformational flexibility of the membrane-solute system makes it easier for the solute to navigate through the membrane.

$E_{TT}(tor)$ is always positive in energy value and measures the force field torsional potential energy for the bonds about which rotations occur in the membrane-solute system. The greater the value of $E_{TT}(tor)$, the greater the average flexibility of the membrane-solute system with regard to torsion angle flexibility for the same reasons as expressed for $E_{TT}(14)$. However, the regression coefficient for this descriptor is negative in eqs 6 and 7, and, consequently P_{caco-2} is predicted to decrease as $E_{TT}(tor)$ increases. Thus, it would seem that $E_{TT}(tor)$ is acting as a refinement term to $E_{TT}(14)$ in eqs 6 and 7 in the same way that $E_{SS}(hb)$ "refines" $\Delta E_{TT}(hb)$ in eqs 5−7.

The joint roles of $E_{SS}(hb)$ and $\Delta E_{TT}(hb)$, as expressed by eq 9, and their influence on solute permeability, may be reflected in the preferred MDS "docking" locations of the solutes within the model-membrane. Solutes having low permeation coefficients tend to dock near the polar heads of the model membrane monolayer, see acyclovir in Figure 6. These solutes generally have strong intermolecular hydrogen bond and/or electrostatic interactions with headgroups and/or the C═O groups of the phospholipids. Solutes with high permeability coefficients either have no preferred docking sites in the monolayer or preferentially locate in the tail

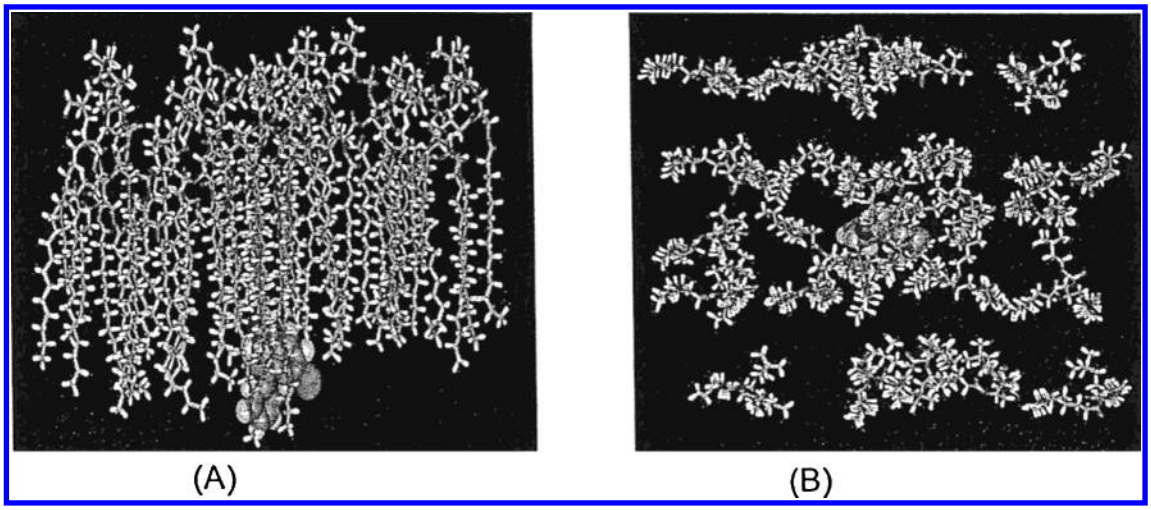

Figure 7. (A) Side view and (B) top view of the most stable structure for the model DMPC membrane−diazepam complex. Diazepam is one of the most permeable compounds in the training set.

regions of the DMPC phospholipids, see, for example, diazepam in Figure 7. These solutes are flexible and/or have limited hydrogen bond and/or electrostatic interactions with the membrane.

It has been shown in past studies that Caco-2 cell permeability correlates with the number of hydrogen bond donor, or acceptor, groups in the solute molecule. The fewer the number of donors and/or acceptors, then the better the permeability of the solute.[6,7] Still, there are compounds that have several hydrogen bonding sites but, at the same time, have high permeation coefficients. One explanation for this apparent conflict, which is consistent with the presence of F(H2O), ΔE_{TT}(hb), and E_{SS}(hb) in the MI-QSAR models, comes from the hypothesis of Stein.[30,31] This hypothesis asserts that the rate-limiting step in the transport of a polar solute across a cell membrane is aqueous desolvation. For a polar solute to transverse a cell membrane, the hydrogen bonds formed with water molecules must be broken. The energy required to break these intermolecular solute−solvent hydrogen bonds can be significant and lead to a major transport barrier. However, if such a polar solute molecule is capable of forming strong intramolecular hydrogen bonds, in place of the solute-water hydrogen bonds, then the energy barrier for the transport of the solute across a lipophilic cell membrane will be reduced. In addition, strong intramolecular solute hydrogen bonding will minimize the hydrogen bonding/electrostatic binding of the solute to the polar headgroups of the phospholipids that can also inhibit solute permeation.

Chi-3 is one of the topological indices developed to encode both molecular size and shape information within a common measure.[27] Caco-2 cell permeability is negatively correlated to Chi-3 in eq 7. Thus, the form of Chi-3 in eq 7 suggests that the more bulky/large is a solute molecule, the less will be its permeability through a Caco-2 cell membrane which makes intuitive sense. Still, it should be kept in mind that Chi-3 contributes little to the prediction of the Caco-2 permeation coefficient in eq 7, since only three compounds have nonzero Chi-3 values. Chi-3 may be a marginal descriptor in terms of significance for *this particular* the training set.

The previous MI-QSAR studies of eye irritation[12,13] led to QSAR models which can be mechanistically interpreted as consisting of two contributing factors:

1. Aqueous Solubility − A parabolic relationship is found between eye irritation potency, MES, and aqueous solubility of the solute irritant. In practice, most eye irritants have aqueous solvation free energies, F(H2O), in a range which display a direct linear relationship (half of the parabola) to eye irritation potency measures.

2. Membrane-Solute Interaction/Binding − A linear relationship is found between increasing (favorable) binding energy of the solute to the phospholipid-rich regions of a membrane and the magnitude of its corresponding MES measures.

These same two factors also appear to partially govern Caco-2 cell permeation, but both contributions exhibit *opposite* relationships to P_{caco-2} measures as compared to MES measures. An increase in aqueous solubility, as measured by an increasingly negative value of F(H2O), decreases P_{caco-2}. The less favorably the solute interacts with the membrane, and/or water, as measured by ΔE_{TT}(hbd), E_{SS}(hb), and Chi-3, the larger is the P_{caco-2} measure. But, overall, the same two factors that govern the eye irritation potency of a solute may, in fact, also play significant roles in its cellular permeation behavior.

There is an additional factor that appears to be important in governing solute permeability that is not found in the eye irritation MI-QSAR models. The greater the *conformational flexibility of the solute* within the membrane, the greater the permeability of the solute. In the case of the P_{caco-2} values of the training and test set compounds, conformational flexibility is expressed in the MI-QSAR models mainly by E_{TT}(14) and E_{TT}(tor) as well as by ΔE_{TT}(hb) and E_{SS}(hb).

If the six terms in eq 7 are grouped together in the following manner

$$P_{caco-2} = -40.50 + [0.65F(H2O)] + [0.06\Delta E_{TT}(hb) - 5.61\chi_3] + [-0.19E_{SS}(hb) + 0.10E_{TT}(14) - 0.03E_{TT}(tor)] \quad (10)$$

then each of the terms within the three sets of bold brackets "define" a contribution to the inferred general mechanism of Caco-2 cell permeation. Hence, eq 10 can be generalized to the form

$$P_{caco-2} = (\text{a constant value}) - [\text{aqueous solubility}] - [\text{membrane-solute binding}] + [\text{conformational flexibility of the solute in the membrane}] \quad (11)$$

An important strength of the MI-QSAR approach is to be able to construct simple and statistically significant relationships such as eqs 2−9 and a corresponding general mechanistic equation like eq 11. That is, MI-QSAR analysis is able to generate meaningful ADME property models employing a limited number of descriptors that can be directly interpreted in terms of physically reasonable mechanisms of action. There is no need to resort to generating very large numbers of [only] *intramolecular solute* descriptors and then producing a model that meets the statistical constraints of acceptance by performing some type of data reduction.

ACKNOWLEDGMENT

The financial support of the Procter & Gamble Company is greatly appreciated. We also acknowledge very helpful discussions with Edward D. Thompson of the Procter & Gamble Company, Martin Dowty of P&G Pharmaceuticals, and Mario Cardozo and M. Yazdanian of Boehringer Ingelheim Pharmaceuticals. Resources of the Laboratory of Molecular Modeling and Design at UIC, Accelrys, Inc. and of The Chem21 Group, Inc. were used in performing this work.

REFERENCES AND NOTES

(1) Pinto, M.; Robine-Leon, S.; Appay, M.; Kedinger, M.; Triadou, N. et al. Caco-2 cell monolayer a surrogate marker for in vivo intestinal permeability in humans. *Biol. Cell* **1983**, *47*, 323−328.
(2) Artursson, P. Cell cultures as models for drug absorption across the intestinal mucosa. *Control. Rev. Ther. Drug. Sys*. **1991**, *8*, 305−330.
(3) Levin, V. A. Relationship of octanol/water partition coefficient and molecular weight to rat brain capillary permeability. *J. Med. Chem*. **1980**, *23*, 682−684.
(4) Hilgers, A. R.; Conradi, R. A.; Burton, P. S. Caco-2 cell monolayers as a model for drug transport across the intestinal mucosa. *Pharm. Res.* **1990**, *7*, 902−910.
(5) Yamashita, S.; Tanala, Y.; Endoh, Y.; Taki, Y.; Sezaki, H. Analysis of drug permeation across caco-2 monolayer: implication for predicting in vivo drug absorption. *Pharm. Res.* **1997**, *14*, 486-491.
(6) Burton, P. S.; Conradi, R. A.; Hilgers, A. R.; Ho, N. F. H.; Maggiora, L. L. The relationship between peptide structure and transport across epithelial cell monolayer. *J. Control. Relat.* **1992**, *19*, 87−98.
(7) Kim, D.; Vurotn, P. S.; Borchardt, R. T. A correlation between the permeability characteristics of a series of peptides using an in vitro cell culture model (caco-2) and those using an in situ perfused rat ileum model of the intestinal mucosa. *Pharm. Res*. **1993**, *10*, 1710−1714.
(8) Palm, K.; Luthman, K.; Ungell, A.; Strandlund, G.; Artursson, P. Correlation of drug absorption with molecular surface properties. *J. Pharm. Sci.* **1996**, *85*, 32−39.
(9) Amidon, L.; Sinlo, P. J.; Fleisher, D. Oral absorption potential of a drugs as predicted by in situ perfusion in the rat. *Pharm. Res.* **1988**, *5*, 651−654.
(10) Pade, V.; Stavchansky, S. Link between drug absorption solubility and permeability measurements in Caco- cells. *J. Pharm. Sci*. **1998**, *87*, 1604−1607.
(11) Stenberg, P.; Norinder, U.; Luthman, K.; Artursson, P. Experimental and computational screening models for the prediction of intestinal drug absorption. *J. Med. Chem*. **2001**, *44*, 1927−1937.
(12) Kulkarni, A. S.; Hopfinger, A. J.; Osborne, R.; Bruner, L. H.; Thompson, E. D. Prediction of Eye Irritation from Orgranic Chemicals Using Membrane-Interaction QSAR Analysis. *Toxicological Sciences* **2001**, *59*, 335−345.
(13) Kulkarni, A. S.; Hopfinger, A. J. Membrane-Interaction QSAR Analysis: Application to the estimation of eye irritation by organic compounds. *Pharm. Res.* **1999**, *16*, 1244−1252.
(14) Yazdanian, M.; Glynn, S. L.; Wright, J. L.; Hawi, A. Correlating Partitioning and Caco-2 Cell Permeability of Structurally Diverse Small Molecular Weight Compounds. *Pharm. Res.* **1998**, *15*, 1490−1494.
(15) Pearlstein, R. A. CHEMLAB-II Users Guide; CHEMLAB Inc.: Chicago, 1988.
(16) Hyperchem HyperChem, Release 4.5 for MS Windows; Hypercube Inc.: Waterloo, Ontario, 1998.
(17) Hauser, H.; Pascher, I.; Pearson, R. H.; Sundell, S. Preffered conformation and moelcular packing of phosphatidylethanolamine and phosphatidylcholine. *Biochim. Biophys. Acta* **1981**, *650*, 21−51.
(18) Mopac Mopac 6.0 release notes; Frank J. Seiler Research Laboratory, United States Air Force Academy: 1990.
(19) Ploeg, P. v. d.; Berendsen, H. J. C. Molecular dynamic simulation of a bilayer membrane. *J. Chem. Phys*. **1982**, *76*, 3271−3276.
(20) Stouch, T. R. Lipid membrane structure and dynamics studied by all atom molecular dynamics simulations of hydrated phospholipid bilayer. *Mol. Simulation* **1993**, *1*, 335−362.
(21) Doherty, D. C. Molsim Version 3.0 User's Guide; The Chem21 Group, Inc.: 2000.
(22) Bloom, M.; Evans, E.; Mouritsen, O. Physical properties of the fluid lipid-bilayer component of cell membranes. A perceptive. *Quarterly Rev. Biophys* **1991**, *24*, 293−397.
(23) Berendsen, H. J. C.; Postman, J. P. M.; Gunsteren, W. F. v.; Nola, A. D.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys*. **1984**, *81*, 3684−3690.
(24) Rogers, D.; Hopfinger, A. J. Applications of Genetic Function Approximation to Quantitative Structure−Activity Relationships and Quantitative Structure−Property Relationships. *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 854−866.
(25) Rogers, D. WOLF 6.2 GFA Program, 1994.
(26) Friedman, J. Multivariate Adaptive Regression Splines; Department of Statistics, Stanford University: Standford, 1988.
(27) Waterbeemd, H. V. D. *Chemometric Methods in Molecular Design*; VCH Publishers: New York, 1995.
(28) Kier, L. B. *Molecular connectivity as a descriptor of structure for SAR analysis*; Marcel Dekker: New York, 1980.
(29) Hopfinger, A. J. *Conformational Properties of Macromolecules*; New York and London, 1973.
(30) Stanton, D. T.; Jurs, P. C. Development and Use of Charge Partial Surface Area Structural Descriptors in Computed Assisted Quantitative Strucutre-Property Relationship Studies. *Anal. Chem*. **1990**, *62*, 2323−2329.
(31) Stein, W. D. *Transport and diffusion accross cell membranes*; Academic Press: New York, 1986.
(32) Stein, W. D. *The molecular basis of diffusion across cell membranes. In the movement of molecules across cell membranes*; Academic Press: New York, 1967.

CI010108D