**Blackwell Publishing**

Full-length article

# Predicting MDCK cell permeation coefficients of organic molecules using membrane-interaction QSAR analysis[1]

Li-li CHEN, Jia YAO, Jian-bo YANG, Jie YANG[2]

*State Key Laboratory of Pharmaceutical Biotechnology, College of Life Sciences, Nanjing University, Nanjing 210093, China*

## Key words

## Abstract

**Aim:** To use membrane-interaction quantitative structure-activity relationship analysis (MI-QSAR) to develop predictive models of partitioning of organic compounds in gastrointestinal cells. **Methods:** A training set of 22 structurally diverse compounds, whose apparent permeability accross cellular membranes of Madin-Darby canine kidney (MDCK) cells were measured, were used to construct MI-QSAR models. Molecular dynamic simulations were used to determine the explicit interaction of each test compound (solute) with a dimyristoyl-phosphatidyl-choline monolayer membrane model. An additional set of intramolecular solute descriptors were computed and considered in the trial pool of descriptors for building MI-QSAR models. The QSAR models were optimized using multidimensional linear regression fitting and the stepwise method. A test set of 8 compounds were evaluated using the MI-QSAR models as part of a validation process. **Results:** MI-QSAR models of the gastrointestinal absorption process were constructed. The descriptors found in the best MI-QSAR models are as follows: 1) Clog$P$ (the logarithm of the 1-octanol/water partition coefficient); 2) $E_{HOMO}$ (the highest occupied molecular orbital energy); 3) $E_s$ (stretch energy); 4) $PM_Y$ (the principal moment of inertia $Y$, the inertia along the $y$ axis in the rectangular coordinates; 5) $C_t$ (total connectivity); and 6) $E_{nb}$ (the energy of interactions between all of the nonbonded atoms). The most important descriptor in the models is Clog$P$. **Conclusion:** Permeability is not only determined by the properties of drug molecules, but is also very much influenced by the molecule-membrane interaction process.

## Introduction

A majority of drugs used today are taken orally, so the drugs must traverse several semipermeable cell membranes before reaching the systemic circulation[1]. These membranes are biological barriers that selectively inhibit the passage of drug molecules and are composed primarily of a bimolecular lipid matrix, containing mostly cholesterol and phospholipids. The lipids provide stability to the membrane and determine its permeability characteristics. Drugs can cross a biological barrier by passive diffusion, facilitate passive diffusion, active transport, and pinocytosis, but passive diffusion is the most common method[1]. In the small intestine, in which the major part of oral drug absorption occurs, the barrier to drug absorption is a membrane comprising intestinal columnar

epithelial cells. For any novel drug that is developed, it is necessary to examine how it is absorbed in the small intestine. In the absorption process, transport across a cell membrane depends on the concentration gradient of the solute. Most drug molecules are transported across a membrane by simple diffusion from a region of high concentration (eg gastrointestinal fluids) to one of low concentration (eg blood)[2]. Because drug molecules are rapidly removed by the systemic circulation and distributed into a large volume of body fluids and tissues, drug concentration in the blood is initially low compared with that at the administration site, producing a large gradient. The diffusion rate is directly proportional to the gradient, but also depends on the molecule's lipid solubility, degree of ionization, and the size and the area of the absorptive surface[3]; that is, the diffusion rate depends

on the characteristics and area of the intestinal membrane and the drug's properties, as well as the drug concentration gradient on both sides of the membrane.

Given the limitations of biomembranes, artificial membranes are usually used instead of biomembranes to study the absorption of drugs in the intestine. *In vitro* studies show that artificial membranes have some properties that are similar to those of biomembranes *in vivo*; that is, artificial membranes can be good mimics of biomembranes[3–5]. The Caco-2 cell monolayer is commonly used as an artificial membrane. When cultured on semipermeable membranes, Caco-2 cells, which are derived from a human colon adenocarcinoma, differentiate into a highly functionalized epithelial barrier with remarkable morphological and biochemical similarity to small intestinal columnar epithelium. The $P_{app}$ values obtained from Caco-2 transport studies have been proven to correlate with those of human intestinal absorption, and the Caco-2 monolayer model has been proven to be extremely useful as a tool for mechanistic studies of drug absorption[3]. However, although the Caco-2 model is well characterized and has been proven to be useful, assays using the system are not perfect. The system requires a 3-week growth period and regular maintenance feeding, so it remains a relatively low throughput method. Irvine and co-workers tried another system using Madin-Darby canine kidney (MDCK) cells and found that it was possibly a useful tool for testing rapid membrane permeability[3]. MDCK cells are commonly used for studying cell growth regulation, metabolism, and transport mechanisms in distal renal epithelia. They can also differentiate into columnar epithelia and form tight junctions when cultured on semipermeable membranes. Irvine and co-workers studied a large number of compounds using both MDCK and Caco-2 assays to evaluate the suitability of MDCK cells as a possible tool for assessing membrane permeability and found that they were well suited. Others have also pointed out that MDCK cells are good candidates for modeling simple epithelia[6].

Based on the study carried out by Irvine and coworkers, we built membrane-interaction quantitative structure-activity relationship analysis (MI-QSAR) models to predict drug permeability through MDCK cells. The QSAR approach involves statistical analysis of various molecular descriptors for a series of biologically active molecules. The result of a QSAR study provides useful clues regarding the type of substituents that should be tested to improve the activity further. QSAR can play a vital role in lead exploitation. One successful example is the transformation of nalidixic acid with the help of QSAR into an important family of drugs: the quinolone carboxylates[7]. Since the method was established

in the 1960s, QSAR equations have been used to describe the biological activities of thousands of different drugs and drug candidates[8,9]. Some new approaches have been used in QSAR since its initial development, for example, principal component analysis (PCA), partial least squares (PLS), artificial neural network (ANN)[10]. MI-QSAR is an advanced form of QSAR, in which membrane-interaction properties are added to the descriptor pool. MI-QSAR models predicting drug permeability through Caco-2 cells have been built by Kulkarni and co-workers[11]. In the present study, we built MDCK cell models for MI-QSAR analysis.

## Materials and methods

**MDCK cell permeation coefficients** The dependent variable used in MI-QSAR analysis is the MDCK cell permeability coefficient, $P_{app}$. Irvine and co-workers performed permeability experiments on a data set of 55 structurally and chemically diverse drugs ranging in molecular weight from 130 to 470 and varying in net charge at pH 7.4[3]. MDCK cells were obtained from the American Type Culture Collection (ATCC; Rockville, MD, USA). MDCK cells were maintained in minimal essential medium containing 10% fetal bovine serum and fresh *L*-glutamine 2 mmol/L. Culture inserts were preincubated with culture medium at 37 °C for 1 h and then seeded with $6.64 \times 10^5$ cells per cm$^2$. MDCK monolayers were washed and fed with fresh medium 1 h post-seeding and again 24 h post-seeding. After 3 d of culturing, MDCK monolayers could be used to test the permeability of compounds. All compounds were tested in 6 replicate monolayers. Monolayers were incubated with donor and acceptor solutions for 60 min at 37 °C, 95% humidity, with 30 r/min reciprocal shaking[3,12]. The permeability coefficient was calculated according to the following equation[3]:

$$P_{app} = (\frac{dQ}{dt}) \times \frac{1}{C_0} \times \frac{1}{A} \tag{1}$$

In this equation $dQ/dt$ is the permeability rate, $C_0$ is the initial concentration in the donor compartment, and $A$ is the surface area of the filter.

Table 1 contains the $P_{app}$ values for 22 structurally diverse drugs used as a training set of compounds and 8 drug compounds as a test set. The 22 training set compounds and 8 test compounds were selected according to the criterion that members of the test set were to be representative of all members of the training set in terms of the range of $P_{app}$ values, molecular weights, and structural and chemical diversities, to achieve a composite representative subset. Table 1 also contains a composite summary of the human

**Table 1.** Molecular weight, MDCK permeability coefficient, and corresponding percentage of drug absorbed (as available) for the drugs of the training set and test set.

| ID | Drug | Molecular weight | Perme- ability ($\mu$m/s) | Percent- age absorbed |
|----|------|------------------|-----------|----------|
| **Training Set** | | | | |
| 1 | Acetaminophen | 151.16 | 0.35 | 80 |
| 2 | Acetylsalicylic acid | 180.16 | 0.074 | 100 |
| 3 | Acyclovir | 225.21 | 0.0021 | 16 |
| 4 | Alprenolol hydrochloride | 285.81 | 1.6 | 93 |
| 5 | Amoxicillin | 365.4 | 0.0024 | 94 |
| 6 | Atenolol | 266.34 | 0.18 | 50 |
| 7 | Bupropion hydrochloride | 276.20 | 1.3 | 87 |
| 8 | Cefatrizine propylene glycol | 462.5 | 0.025 | 76 |
| 9 | Corticosterone | 346.47 | 1.4 | 100 |
| 10 | Lamotrigine | 256.09 | 0.88 | 70 |
| 11 | Methylprednisolone | 374.48 | 0.16 | 82 |
| 12 | Nadolol | 309.4 | 0.014 | 34 |
| 13 | Ondansetron | 293.37 | 1.1 | 100 |
| 14 | Penicillin V | 350.39 | 0.0015 | 45 |
| 15 | Phenytoin | 252.27 | 1.2 | 90 |
| 16 | Pindolol | 248.32 | 0.59 | 90 |
| 17 | Practolol | 266.34 | 0.013 | 100 |
| 18 | Progesterone | 314.46 | 1.6 | 91 |
| 19 | Propranolol hydrochloride | 295.80 | 1.7 | 90 |
| 20 | Salicylic acid | 138.12 | 0.1 | 100 |
| 21 | Testosterone | 288.43 | 1.4 | 100 |
| 22 | Trimethoprim | 290.32 | 0.52 | 97 |
| **Test set** | | | | |
| 23 | Antipyrine | 188.22 | 1.5 | 100 |
| 24 | Dexamethasone | 392.46 | 0.2 | 100 |
| 25 | Guanabenz | 231.08 | 1.9 | 75 |
| 26 | Hydrocortisone | 362.46 | 0.31 | 91 |
| 27 | Propylthiouracil | 170.22 | 0.41 | 75 |
| 28 | AZT | 267.24 | 0.06 | 100 |
| 29 | Cephalexin | 347.39 | 0.0048 | 98 |
| 30 | Gabapentin | 171.24 | 0.0036 | 50 |

percentage absorption values of many of the drugs in the table. These data were obtained from published values. Compared with the $P_{app}$ value and the human percentage absorption value, it is obvious that the $P_{app}$ values are indicative of *in vivo* drug uptake. Figure 1 shows the structures of these thirty molecules.

**Building solute molecules and a dimyristoyl-phosphatidyl-choline (DMPC) monolayer** All the solute molecules of the training set and test set were built using HyperChem software[13]. A single DMPC molecule was built using HyperChem from the published crystal structure data[14]. The

DMPC molecule was selected as the model phospholipid in this study. The structure of the DMPC molecule is shown in Figure 2. An assembly of 25 DMPC molecules, 5×5×1 in *x*, *y*, and *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[15]. These workers built 2 decanoate bilayers, with 2×8×2 and 2×16×2 phospholipid molecules, respectively, and performed a molecular dynamic simulation for each of them. It was found that the estimated order parameters for these 2 model bilayers agreed with one another, which suggests 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 the simulation system that permits a minimum effective size (phospholipids) of the monolayer to be defined[16].

To prevent unfavorable van der Waals' interactions between solute molecules and the membrane DMPC molecules, the single DMPC molecule in the center of the monolayer, which was located at position *x, y*=3,3 of the 5×5 DMPC monolayer model, was taken out[11], creating a space for the test solute molecules to insert into. Each of the test solute molecules of the permeation data set was inserted at 3 different positions or depths in the DMPC monolayer, with the most polar group of the solute facing toward the head group region of the monolayer. Three corresponding MDS (molecular dynamic simulation) models were generated for each solute molecule with regard to the trial positions of the solute molecules in the monolayer. The 3 trial positions were as follows[4,5,17]: solute molecule in the head group region, solute molecule between the head group region and the aliphatic chains, and solute molecule in the tail region of the aliphatic chains.

The energetically favorable geometry of the solute molecule in the monolayer was sought after using each of these trial positions. The 3 different positions of AZT, one of the training set solute molecules, are shown in Figure 3A to illustrate this modeling procedure. The most energetically favorable geometry of this solute molecule in the model DMPC monolayer is shown in Figure 3B.

**Molecular dynamic simulations** MDS were carried out using the Discover module in Material Studio[18] with compass force field and NVT ensemble(an ensemble in which the dynamics are modified to allow the system to exchange heat with the environment at a controlled temperature). The selection of the simulation temperature was based on the phase transition temperature for DMPC, which was 297 K[11]. A simulation temperature of 311 K was selected because it is
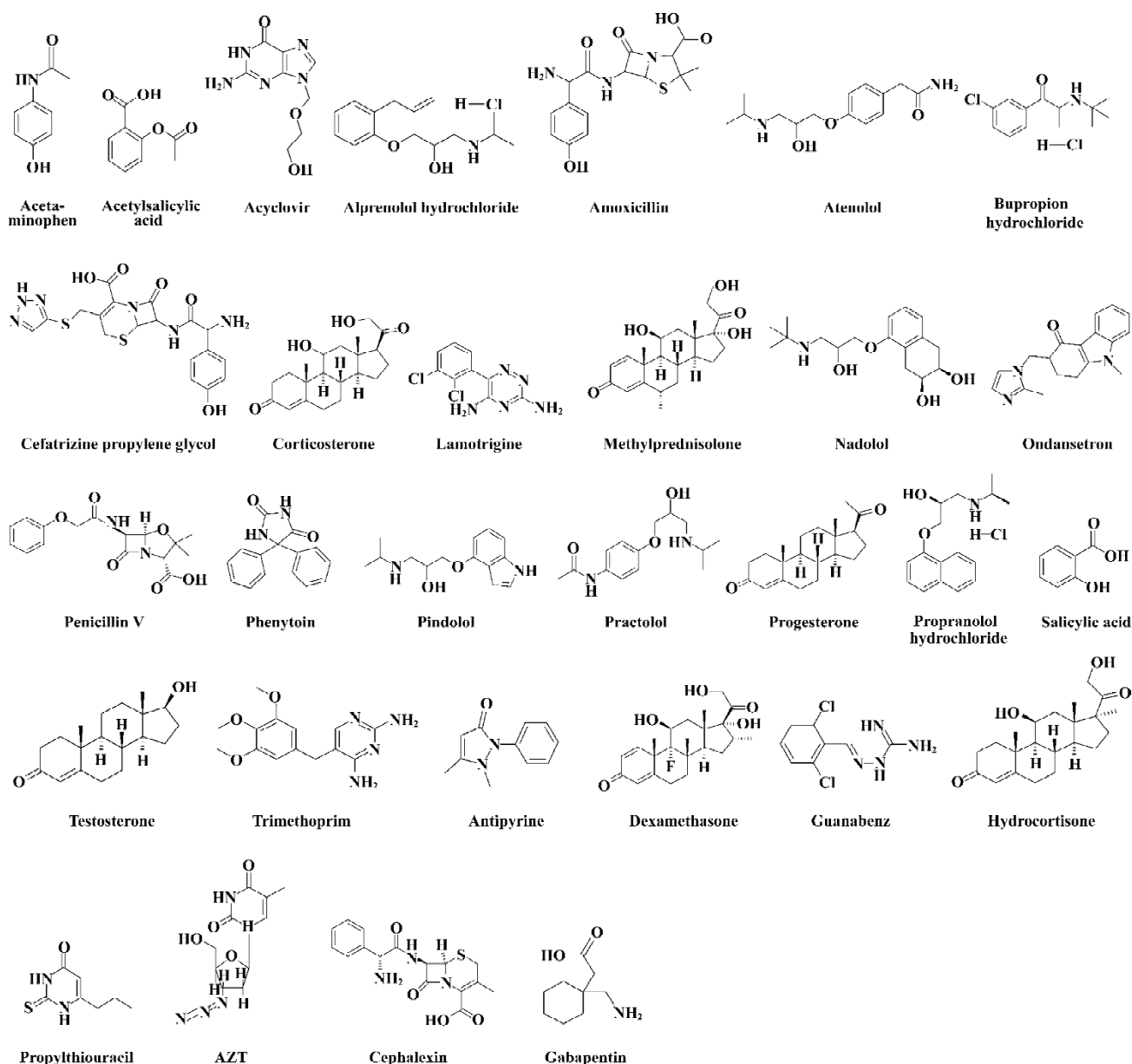
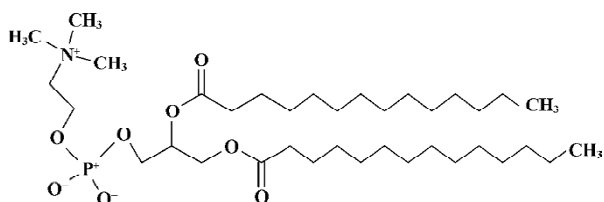**Figure 1.** Structures of training set compounds and test set compounds.

**Figure 2.** Chemical structure of a DMPC molecule.

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. The trajectory step size was 0.001 ps over a total simulation time of 10 ps for each solute of the training set and test set. After 10 ps of simulation, molecule-membrane interactions arrived at an energetically stable phase (Figure 4). Two-dimensional periodic boundary conditions were used (a=32 Å, b=32 Å, c=80 Å, and γ=96.0°) for the
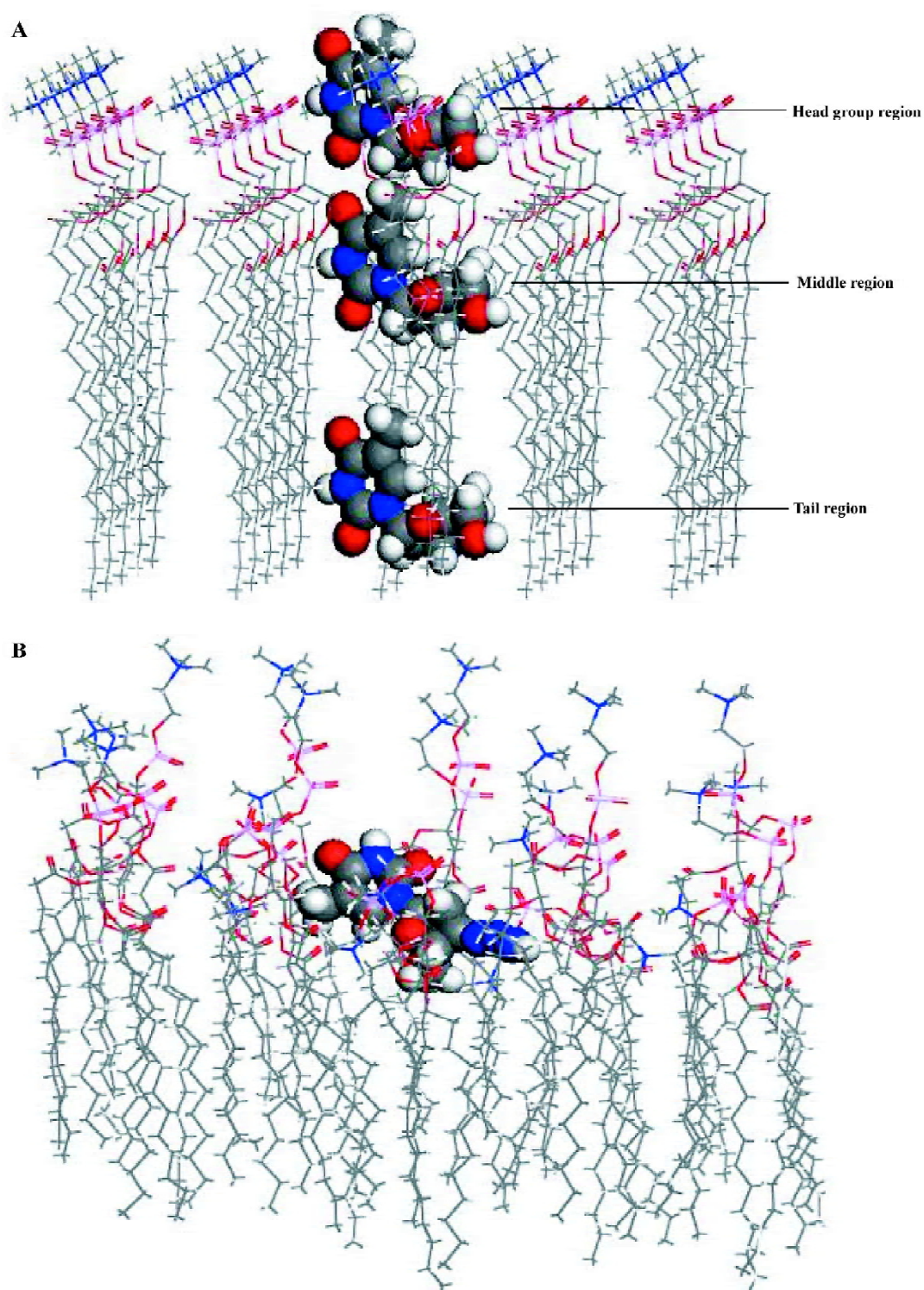
**Figure 3**. (A) "Side" view of an AZT molecule inserted at 3 different positions in the DMPC model monolayer, prior to the start of each simulation; (B) "Side" view of the most energetically favorable geometry of AZT in the DMPC model monolayer.
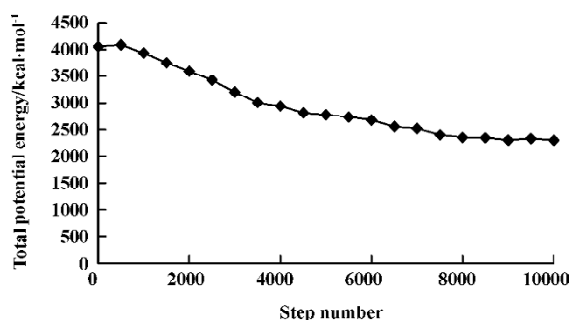
**Figure 4.** Total potential energy vs trajectory step number for acyclovir embedded in the model DMPC monolayer.

DMPC molecules of the monolayer model, but not the test solute molecule. Only a single solute molecule was explicitly considered in each MDS. The angle γ is the angle that an extended DMPC molecule makes with the "planar surface" of the monolayer. Each of the solute molecules was placed at each of the 3 different positions in the monolayer, as described earlier, with the most polar portion of the solute "facing" toward the head group region. A snapshot of molecule AZT and the membrane around it was taken after MDS, as shown in Figure 3b.

**Calculation of descriptors** Both intramolecular physi-cochemical properties and features of the solute molecules and intermolecular solute-membrane interaction properties were calculated. These properties and features will both be referred to as descriptors from this point forward because 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 the following 3 kinds: 1) solute aqueous dissolution and solvation descriptors; 2) solute-membrane interaction descriptors; and 3) general intramolecular solute descriptors. Table 2 and Table 3 reporting the trial pool of descriptors used in the MDCK cell permeation MI-QSAR modeling use 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 MDCK cell permeation process) of a solute. There are other intermolecular properties computed using intramolecular computational methods, which are not included in Table 2, for example Clog$P$ (the logarithm of 1-octanol/water partition coefficient), MP (melting point), Tc (critical tempeture), Sol (water solubility). All of these descriptors are intermolecular properties. They

**Table 2.** General intramolecular solute descriptors used in the MI-QSAR descriptor pool.

| Descriptors | Explanation of descriptors |
| --- | --- |
| HOMO | Highest occupied molecular orbital energy |
| LUMO | Lowest occupied molecular orbital energy |
| Dp | Dipole moment |
| $V_m$ | Molecular volume |
| SA | Molecular surface area |
| MW | Molecular weight |
| MR | Molecular refractivity |
| PM($x,y,z$) | Principal moment of inertia along the $x,y,z$ axis in the rectangular coordinate |
| HF | Heat of formation ($\Delta h_f$) for the structure at 298.15 K and 1 atm |
| H | The logarithm of Henry's law constant |
| $G$ | Standard Gibbs free energy |
| Ovality | The ratio of the molecular surface area to the minimum surface area, which is the surface area of a sphere having a volume equal to the solvent-excluded volume of the molecule |
| $E_b$ | Sum of the angle bending terms of the force-field equation |
| $E_c$ | Sum of the electrostatic energy representing the pairwise interaction of charged atoms |
| $E_v$ | Sum of pairwise van der Waals' interaction energy terms for atoms separated by more than 3 chemical bonds |
| $E_s$ | Stretch energy, the energy contribution associated with the deformation of a bond from its equilibrium bond length |
| $E_{sb}$ | Sum of the stretch-bend coupling terms of the force-field equation |
| $E_t$ | Sum of the dihedral bond rotational energy term of the force-field equation |
| E14 | Sum of pairwise van der Waals' interaction energy terms for atoms separated by exactly 3 chemical bonds |
| NRE | Repulsion energy |
| $C_t$ | A kind of index defined according to the number of atoms and bonds and how they connect together |

are classified as solvation and dissolution intermolecular descriptors and are reported in Table 3b. In Table 3a, there are solute-membrane interaction descriptors extracted directly from the MDS trajectories. These intermolecular descriptors were calculated using the most stable (lowest total potential energy) solute-membrane geometry from the 3 positions sampled for each of the solutes. For example, Figure 3b shows the lowest potential energy state of AZT in the membrane monolayer, which was used to estimate the solute-membrane interaction descriptors. Other solute-membrane interaction descriptors used in the QSAR descriptor trial set were determined using data from the MDS trajectories.

**Construction and testing of MI-QSAR models** Independent and useful descriptors can be extracted from all descriptors calculated earlier. The methods in common use are forward regression, backward regression and stepwise regression. Stepwise regression is the method used most frequently. The resulting regression equation from the stepwise regression method is not the best, rather it is an optimized result. SPSS was used to obtain the regression equation; it selected proper variables according to the partial sum of squares of regression in every step. The partial sum of squares for regression means that the sum of squares for regression increased or decreased after one variable was added to or deducted from the present regression equation. A parameter, $F$, is defined to check whether to introduce a variable or reject it when partial sum of squares for regression is of some value.

The stepwise regression method was used in this study, which was the combination of forward regression and backward regression methods. The single worst variable was picked out after each new variable was added. Two threshold constants, $F_{entry}$ and $F_{removal}$ ($F_{entry} < F_{removal}$), were given before the regression. If the $F$ value of one variable, which had the largest partial sum of squares for regression of all the variables not included in the regression equation, was larger than or equal to $F_{entry}$, the variable could be introduced into the equation. Conversely, if the $F$ value of one variable, which had the smallest partial sum of squares for regression among all the variables included in the regression equation, was less than or equal to $F_{removal}$, the variable should be removed from the regression equation. These steps were carried out alternately until there was no variable to be introduced into the regression equation and none to be removed from it. The adjustment of the threshold values $F_{entry}$ and $F_{removal}$ could affect the result of the selection of variables. If the prepared variables were few, it was appropriate to increase $F_{entry}$ in order to introduce variables to the regression

**Table 3.** Intermolecular interaction descriptors in the trial MI-QSAR descriptor pool.

| Part a | |
| --- | --- |
| Membrane-solute descriptors | Description of the membrane-solute descriptors |
| Potential | The energy of a particle or system of particles derived from position, or condition, rather than motion |
| Electrostatic | The energy created by the interaction of charges between the molecule and the membrane |
| $E_{nb}$ | The energy of interactions between non-bonded atoms |
| Angle | Valence angle bending |
| Bond | Bond stretching |
| Torsion | Dihedral angle torsion |
| Out-of-plane | Part of nearly all force fields for a covalent system |

| Part b | |
| --- | --- |
| Dissolution and solvation-solute descriptors | Description of the dissolution and solvation–solute descriptors |
| Clog$P$ | The logarithm of l-octanol/water partition coefficient |
| MP | The melting point for the structure at 1 atm |
| Tc | The temperature above which the gaseous form of the structure cannot be liquefied, no matter what pressure is applied |
| Sol | Solubility of the structure in water at 25 °C |

Note: Part a includes the membrane-solute interaction descriptors and Part b lists the intermolecular dissolution and solvation descriptors of the solute.

equation as much as possible. If the prepared variables were too many, $F_{removal}$ could be decreased to cut down the number of variables introduced into the regression equation.

SPSS was used to carry out the regressions and construct the models[19].

## Results and Discussion

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

$$P_{app}=115.657+319.687ClogP$$
$$n=22 \quad R^2=0.646 \quad S=0.396 \tag{2}$$

$$P_{app}=3113.84+374.691ClogP+338.881E_{HOMO}$$
$$n=22 \quad R^2=0.719 \quad S=0.362 \tag{3}$$

$$P_{app}=3115.743+399.894ClogP+374.586E_{HOMO}+35.07E_s$$
$$n=22 \quad R^2=0.764 \quad S=0.341 \tag{4}$$

$$P_{app}=3453.482+409.333ClogP+391.596E_{HOMO}+48.403E_s-0.0971PM_Y$$
$$n=22 \quad R^2=0.813 \quad S=0.312 \tag{5}$$

$$P_{app}=3609.933+387.817ClogP+356.922E_{HOMO}+32.256E_s-0.163PM_Y-16594C_t$$
$$n=22 \quad R^2=0.866 \quad S=0.272 \tag{6}$$

$$P_{app}=1029.094+398.972ClogP+459.781E_{HOMO}+24.048E_s-0.174PM_Y-18164.4C_t+2.594E_{nb}$$
$$n=22 \quad R^2=0.896 \quad S=0.248 \tag{7}$$

where $n$ is the number of compounds, $R^2$ is the coefficient of determination, and S is the standard error of the estimate (and its value could be different if the unit of the variables changed). The descriptors found in the best MI-QSAR models are as follows: 1) $Clog P$ is the logarithm of the l-octanol/water partition coefficient; 2) $E_{HOMO}$ is the highest occupied molecular orbital energy; 3) $E_s$ is stretch energy, the energy contribution associated with the deformation of a bond from its equilibrium bond length; 4) $PM_Y$ is principal moment of inertia $Y$, the inertia along the $y$ axis in the rectangular coordinates; 5) $C_t$ is total connectivity, which is a kind of index defined according to the number of atoms and bonds and their connecting sequence ($C_t$ is a structural parameter; molecules with different structures have different connectivity indices according to a given definition); and 6) $E_{nb}$ is the energy of interactions between all of the non-bonded atoms.

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

The observed and predicted (using the 3–6 term MI-QSAR models) MDCK cell permeation coefficients of the training and test set compounds are listed in Table 5 and plotted in Figure 5. Corticosterone, ondansetron, phenytoin, progesterone, propranolol hydrochloride, and testosterone are observed to permeate better than predicted by each of the MI-QSAR models, whereas acetylsalicylic acid, bupropion hydrochloride, methylprednisolone, and nadolol have a lower permeation coefficient than the one 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 6 contains plots of $R^2$ and S for the training set. $R^2$ increases with increasing numbers of descriptor terms, whereas the value of S decreases when the number of descriptor terms increases.
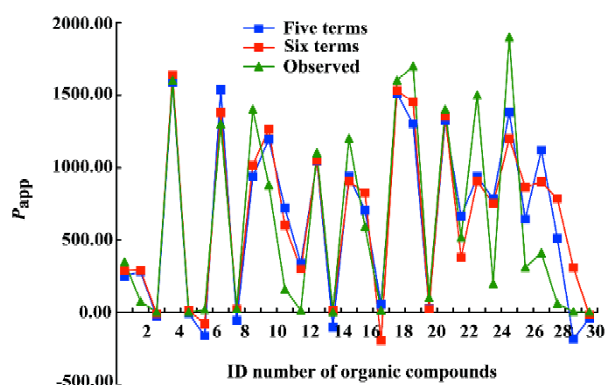


**Figure 5.** Observed versus predicted MDCK cell permeability coefficients for all the compounds of the training and test sets in the corresponding MI-QSAR model.



**Figure 6.** Diagnostic plot of the MI-QSAR. $R^2$ is the correlation coefficient and S is standard error of the estimate in the regression progress.

Analyzing Equations 2–7, it appears that $ClogP$ in the 1-term model accounts for much of the variance of $P_{app}$ across the training set. Both principal moment of inertia $Y$ and total

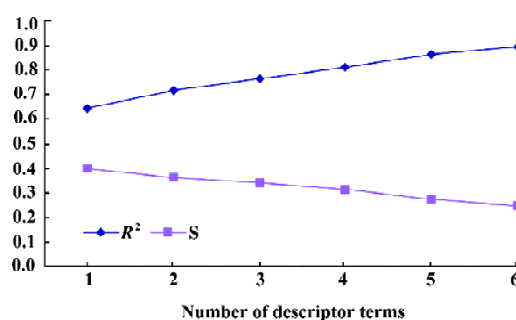**Table 4.** Values of the 6 descriptors found to be the significant MI-QSAR terms in Equations 2–7.

| ID | Structure name | ClogP | $E_{HOMO}$/eV | $E_s$/kcal·mol⁻¹ | $PM_Y$[1] | $C_t$ | $E_{nb}$/kcal·mol⁻¹ |
|---|---|---|---|---|---|---|---|
| Training set | | | | | | | |
| 1 | Acetaminophen | 0.61 | −8.62 | 6.10 | 942.88 | 3.40E−02 | 1393.32 |
| 2 | Acetylsalicylic acid | 1.24 | −9.72 | 3.38 | 765.36 | 1.96E−02 | 1404.13 |
| 3 | Acyclovir | −1.45 | −8.74 | 9.80 | 1663.84 | 4.01E−05 | 1392.69 |
| 4 | Alprenolol hydrochloride | 3.81 | −8.85 | 6.50 | 3020.37 | 3.47E−03 | 1384.42 |
| 5 | Amoxicillin | −0.15 | −9.27 | 7.61 | 3030.93 | 2.57E−04 | 1408.59 |
| 6 | Atenolol | 0.56 | −9.11 | 7.08 | 5650.81 | 2.84E−03 | 1432.93 |
| 7 | Bupropion hydrochloride | 4.05 | −8.95 | 5.19 | 2792.27 | 9.82E−03 | 1305.57 |
| 8 | Cefatrizine propylene glycol | −0.89 | −8.75 | 18.43 | 4873.42 | 1.52E−05 | 1456.15 |
| 9 | Corticosterone | 3.08 | −10.15 | 12.51 | 3983.53 | 1.36E−04 | 1471.39 |
| 10 | Lamotrigine | 2.39 | −8.84 | 6.15 | 1897.16 | 4.63E−03 | 1391.16 |
| 11 | Methylprednisolone | 2.78 | −10.03 | 13.56 | 5053.62 | 9.65E−05 | 1398.77 |
| 12 | Nadolol | 1.27 | −9.07 | 5.81 | 4280.39 | 8.18E−04 | 1370.14 |
| 13 | Ondansetron | 0.97 | −8.66 | 16.00 | 2238.49 | 2.23E−04 | 1396.91 |
| 14 | Penicillin V | 0.57 | −9.13 | 6.36 | 5382.44 | 2.73E−04 | 1442.01 |
| 15 | Phenytoin | 2.26 | −9.59 | 3.77 | 1432.63 | 8.68E−04 | 1371.56 |
| 16 | Pindolol | 1.01 | −8.24 | 7.22 | 3405.34 | 2.00E−03 | 1401.22 |
| 17 | Practolol | 0.64 | −8.53 | 5.08 | 5380.10 | 2.84E−03 | 1274.31 |
| 18 | Progesterone | 4.63 | −9.99 | 8.68 | 3716.02 | 2.36E−04 | 1423.14 |
| 19 | Propranolol hydrochloride | 2.80 | −8.23 | 5.60 | 3766.75 | 1.42E−03 | 1401.56 |
| 20 | Salicylic acid | 1.46 | −9.45 | 2.64 | 412.02 | 4.81E−02 | 1401.56 |
| 21 | Testosterone | 3.84 | −9.95 | 7.98 | 2910.24 | 4.09E−04 | 1423.14 |
| 22 | Trimethoprim | 0.51 | −8.63 | 11.61 | 2604.18 | 9.45E−04 | 1274.31 |
| Test set | | | | | | | |
| 23 | Antipyrine | 0.79 | −8.27 | 8.62 | 1063.6 | 8.02E−03 | 1344.44 |
| 24 | Dexamethasone | 2.65 | −10.25 | 16.84 | 4504.26 | 8.35E−05 | 1449.02 |
| 25 | Guanabenz | 2.95 | −8.71 | 6.82 | 1547.79 | 1.39E−02 | 1293.76 |
| 26 | Hydrocortisone | 2.37 | −10.02 | 12.88 | 4435.55 | 1.18E−04 | 1527.47 |
| 27 | Propylthiouracil | 3.25 | −9.10 | 6.19 | 860.4 | 3.40E−02 | 1301.05 |
| 28 | AZT | −1.54 | −6.22 | 3.52 | 2232 | 1.89E−03 | 1376.28 |
| 29 | Cephalexin | −0.54 | −8.86 | 5.33 | 3630.21 | 2.57E−04 | 1571.03 |
| 30 | Gabapentin | 0.96 | −10.08 | 3.96 | 807.524 | 2.55E−02 | 1431.10 |

connectivity have a negative effect on the permeation coefficient. In 1-term to 5-term models, the descriptors are all intramolecular descriptors, but in the 6-term model, one membrane-solute interaction descriptor, $E_{nb}$, is added in.

Eight compounds from the parent MDCK cell permeation coefficient data were selected to construct a test set to validate the MI-QSAR models. The molecules of the test set were selected so as to span the entire range of MDCK cell permeability for the composite training set. At the bottom of Table 5 are the observed and predicted $P_{app}$ values for this test set. Figure 5 plots the test set as the last 8 compounds. There are no outliers, but antipyrine and guanabenz, compounds 1 and 3 of the test set, are predicted to have lower permeability coefficients than observed. Conversely,

dexamethasone, hydrocortisone, propylthiouracil and AZT have higher predicted $P_{app}$ values in all the models than the observed value.

The most important descriptor in the models is ClogP. ClogP is the computed logarithm of the l-octanol/water partition coefficient, and it has been ubiquitously used as a quantitative measure of hydrophobicity/lipophilicity since the 1960s[20,21]. A high ClogP value of a molecule implies that the molecule dissolves easily in hydrophobic materials and dissolves poorly in water. In the MI-QSAR models we built, ClogP had a positive effect on $P_{app}$, which implies that chemicals with high lipophilicity have better permeability through membranes than hydrophilic structures. It is easy to see why this is the case, because biomembranes are mainly composed

**Table 5.** Observed and predicted MDCK permeability coefficients for the 3–6 term MI-QSAR models.

| | Structure name | Observed $P_{app}$ | 3-term | 4-term | 5-term | 6-term |
|---|---|---|---|---|---|---|
| | Training set | | | | | |
| 1 | Acetaminophen | 350.0 | 343.16 | 529.78 | 246.92 | 286.17 |
| 2 | Acetylsalicylic acid | 74.0 | 89.80 | 244.65 | 280.48 | 289.21 |
| 3 | Acyclovir | 2.1 | –393.51 | –248.96 | –26.92 | –8.91 |
| 4 | Alprenolol hydrochloride | 1600.0 | 1550.53 | 1566.97 | 1586.89 | 1636.99 |
| 5 | Amoxicillin | 2.4 | –150.74 | –165.04 | –10.67 | 10.81 |
| 6 | Atenolol | 18.0 | 177.07 | –89.12 | –162.70 | –81.56 |
| 7 | Bupropion hydrochloride | 1300.0 | 1566.62 | 1588.50 | 1537.19 | 1379.35 |
| 8 | Cefatrizine propylene glycol | 25.0 | 127.43 | 80.46 | –59.50 | 21.67 |
| 9 | Corticosterone | 1400.0 | 985.72 | 959.89 | 935.15 | 1015.25 |
| 10 | Lamotrigine | 880.0 | 977.29 | 1085.06 | 1195.33 | 1262.35 |
| 11 | Methylprednisolone | 160.0 | 946.65 | 830.13 | 720.90 | 600.99 |
| 12 | Nadolol | 14.0 | 428.39 | 285.60 | 339.90 | 297.99 |
| 13 | Ondansetron | 1100.0 | 821.84 | 1017.46 | 1043.64 | 1050.40 |
| 14 | Penicillin V | 1.5 | 147.21 | –102.81 | –104.00 | 11.30 |
| 15 | Phenytoin | 1200.0 | 559.43 | 666.59 | 937.19 | 904.82 |
| 16 | Pindolol | 590.0 | 686.72 | 659.43 | 705.59 | 823.54 |
| 17 | Practolol | 13.0 | 356.30 | 100.34 | 55.06 | –195.28 |
| 18 | Progesterone | 1600.0 | 1528.22 | 1494.60 | 1508.96 | 1530.93 |
| 19 | Propranolol hydrochloride | 1700.0 | 1348.61 | 1281.68 | 1301.12 | 1450.92 |
| 20 | Salicylic acid | 100.0 | 253.68 | 439.76 | 24.13 | 21.74 |
| 21 | Testosterone | 1400.0 | 1204.89 | 1233.46 | 1324.80 | 1357.07 |
| 22 | Trimethoprim | 520.0 | 494.62 | 592.32 | 662.24 | 379.69 |
| | Test set | | | | | |
| 23 | Antipyrine | 1500.0 | 635.25 | 851.38 | 935.33 | 904.85 |
| 24 | Dexamethasone | 200.0 | 924.92 | 900.40 | 785.27 | 750.10 |
| 25 | Guanabenz | 1900.0 | 1271.01 | 1428.98 | 1381.52 | 1198.71 |
| 26 | Hydrocortisone | 310.0 | 760.79 | 691.49 | 642.19 | 864.34 |
| 27 | Propylthiouracil | 410.0 | 1225.72 | 1438.47 | 1119.10 | 900.19 |
| 28 | AZT | 60.0 | 292.54 | 340.14 | 510.15 | 785.71 |
| 29 | Cephalexin | 4.8 | –230.65 | –330.13 | –184.49 | 308.96 |
| 30 | Gabapentin | 3.6 | –137.18 | 12.50 | –42.70 | –18.71 |

of double phospholipid layers, so small hydrophobic molecules can pass through the layer with little obstruction. Commonly, lipophilic drugs permeate the small intestinal columnar epithelium quicker and easier than hydrophilic drugs.

In Equation 3, another descriptor, $E_{HOMO}$ appears, which represents HOMO energy. HOMO energy is the energy of the highest occupied molecular orbit, which is the opposite of LUMO energy, the energy of the lowest unoccupied molecular orbit[22]. The greater $E_{HOMO}$ is, the greater the electron-donating capability; conversely, the smaller $E_{LUMO}$ is, the smaller the resistance to accept electrons. Compounds that present larger values of $E_{HOMO}$ are more electron donor and the compounds that present smaller values of $E_{LUMO}$ are

more electron acceptor. These variables are interpreted as measures of molecular reactivity and stability. As $E_{HOMO}$ increases (relative to other molecules), the molecule is less stable and more reactive. For $E_{LUMO}$, the situation is the opposite. HOMO energy is relevant because Equation 3 measures the electron-donating character of the drug molecules. $P_{app}$ has a positive relationship with HOMO energy, which means that drug molecules with higher electron-donating capacities can permeate membranes better.

As seen in the literature, energy is often discussed in many research programs. There is another energy descriptor, stretch energy, which influences $P_{app}$ as $E_{HOMO}$ does in our models. Stretch energy is the energy contribution associated with the deformation of a bond from its equilibrium bond

length. It is a kind of vibration energy, which is composed of bend energy and stretch energy. From Models 3–6, we can see that stretch energy is positively correlated with the value of $P_{app}$. Increasingly positive stretch energy values correspond to increasing $P_{app}$. Drug molecules with higher stretch energies metamorphose better and can permeate membranes more easily.

PM$_Y$ is the principal moment of inertia along the *y* axis. It gives information about how the product of force and distance influence the value of $P_{app}$. $P_{app}$ decreases as the value of PM$_Y$ increases. The coefficients of PM$_Y$ in Models 4–6 are very small when contrasted with the coefficients of other variables. This indicates that a change in PM$_Y$ value has comparatively little effect on the value of $P_{app}$ and the permeability of the drug molecules.

Compared with the descriptors discussed earlier, connectivity is an element that rarely appeared in QSAR models constructed by other researchers. Connectivity is a parameter defined according to the conformation of a molecule, and presumes that there is some functional relationship between molecule properties and connectivity. One standard definition of connectivity is as follows[23,24]:

Figure 7 is the structure of isopentanol without hydrogens. The figures are the bond numbers of each carbon connected with other carbons.
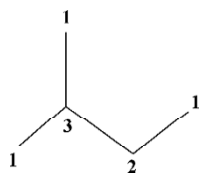


**Figure 7.** Structure of isopentanol without hydrogens.

$$Connectivity = \frac{1}{\sqrt{1 \times 3 \times 2 \times 1 \times 1}} \; [24]$$

From the foregoing equation, we can see that connectivity reflects the connection and ramification conditions of the molecules. Connectivity is the quantitative description of the molecular structure. Molecules with different structures have different connectivity values. The connectivity method has been widely used in structure-activity analyses. In the models we constructed, connectivity is included in Model 5 and Model 6 and it has a negative effect on the $P_{app}$ value. The larger the connectivity value, the more complicated the structure will be, and there will be more branch chains. Therefore an increase in connectivity will decrease the permeability of a molecule.

$E_{nb}$ is the energy of interactions between non-bonded atoms. It includes van der Waals' energy, electrostatic energy and hydrogen bond terms in some older forcefields, as the following equation describes[25],

$$E_{non\text{-}bond} = E_{vdW} + E_{coulomb} + E_{hbond} \tag{8}$$

The non-bond energy terms in Equations 6 and 7 suggest that permeability increases with increasing binding of the solute to the phospholipid regions of the membranes.

Kulkarni and coworkers carried out MI-QSAR analysis using the permeability coefficients of some drug molecules tested by Caco-2 cells. Thirty molecules were included in their training set, and significant MI-QSAR models were built. Twenty-three intramolecular descriptors, and 11 membrane-solute interaction descriptors were calculated. In our study, we calculated more than 70 descriptors for every drug molecule in the training set, including electronic, steric, and thermodynamic properties. MI-QSAR models were built based on the analysis of these descriptors. In addition, the present study confirmed the conclusion of Irvine and coworkers, that MDCK cells are suitable for studying drug permeability through biomembranes.

## Acknowledgements

## References

1   Moo JC, David PT, Clay TC, Thomas JV, Jeffrey FS. The Madin Darby Canine Kidney (MDCK) epithelial cell monolayer as a model cellular transport barrier. Pharm Res 1989; 6: 71–7.

2   Magali T, Peter R, Daniel HS. Pharmacology. Churehill Livingstone: Harcourt Asia; 1998.

3   Irvine JD, Takahashi L, Lockhart K, Cheong J, Tolan JW, Selick HE, *et al*. MDCK (Madin-Darby Canine Kidney) cells: a tool for membrane permeability screening. J Pharm Sci 1999; 88: 28–33.

4   Kulkarni A, Hopfinger AJ, Osbrone R, Bruner LH, Thompson ED. Prediction of eye irritation from organic chemicals using membrane-interaction QSAR analysis. Toxicol Sci 2001; 59: 335–45.

5   Kulkarni A, Hopfinger AJ. Membrane-interaction QSAR analysis: application to the estimation of eye irritation by organic compounds. Pharm Res 1999; 16: 1244–52.

6   Rothen-Rutishauser B, Kraemer SD, Braun A, Guenthert M, Wunderli-Allenspach H. MDCK cell cultures as an epithelial *in vitro* model: cytoskeleton and tight junctions as indicators for the definition of age-related stages by confocal microscopy. Pharm Res 1998; 15: 964–71.

7   Alka K. C-QSAR: a database of 18000 QSARs and associated biological and physical data. J Comput Aided Mol Des 2003; 17: 187–96.

8   Hugo K. QSAR and 3D-QSAR in drug design. Part 1: methodology.

Drug Discov Today 1997; 2: 457–67.

9   Hugo K.  QSAR and 3D-QSAR in drug design.  Part 2: applications and problems.  Drug Discov Today 1997; 2: 538–46.

10  Lennart E, Erik J.  Multivariate design and modeling in QSAR. Chemometrics Intell Lab Syst 1996; 34: 1–19.

11  Kulkarni A, Han Y, Hopfinger AJ.  Predicting Caco-2 cell permeation coefficients of organic molecules using membrane-interaction QSAR analysis.  J Chem Inf Comput Sci 2002; 42: 331–42.

12  Braun A, Hammerle S, Suda K, Rothen-Rutishauser B, Gunthert M, Kramer SD, *et al*.  Cell cultures as tools in biopharmacy.  Eur J Pharm Sci 2000; 11 Suppl 2: S51–60.

13  HyperChem.  HyperChem, Release 6.0 for MS Windows.  Waterloo (Ontario, Canada): Hypercube; 2001.

14  Hauser H, Pascher I, Pearson RH, Sundell S.  Preferred conformation and molecular packing of phosphatidylethanolamine and phosphatidylcholine.  Biochim Biophys Acta 1981; 650: 21–51.

15   van der Ploeg P, Berendsen HJC.  Molecular dynamic simulation of a bilayer membrane.  J Chem Phys 1982; 76: 3271–6.

16  Stouch TR.  Lipid membrane structure and dynamics studied by all atom molecular dynamics simulations of hydrated phosphatidylcholine vesicles.  Mol Simulation 1993; 1: 335–62.

17  Gurtovenko AA, Patra M, Karttunen M, Vattulainen I.  Cationic DMPC/DMTAP lipid bilayers: molecular dynamics study. Biophys J 2004; 86: 3461–72.

18  Accelrys: Materials Studios.  San Diego (USA): Accelrys; 2001.

19  SPSS.  SPSS for Windows. Chicago (USA): SPSS Inc; 2001.

20  Butina D, Segall MD, Frankcombe K.  Predicting ADME properties in silico: methods and models.  Drug Discov Today 2002; 7 Suppl 11: S83–8.

21  Albert PL.  Screening for human ADME/Tox drug properties in drug discovery.  Drug Discov Today 2001; 6: 357–66.

22  Honorio KM, Da Siliva ABF.  An AM1 study on the electron-donating and electron-accepting character of biomolecules.  Int J Quant Chem 2003; 95: 126–32.

23  Kier LB, Hall LH.  Molecular connectivity in chemistry and drug research.  New York: Academic Press; 1976.

24  Kier LB, Hall LH.  Molecular connectivity in structure-activity analysis.  London: John Wiley; 1986.

25  Iyer M, Mishra R, Han Y, Hopfinger AJ.  Predicting blood-brain barrier partitioning of organic molecules using membrane-interaction QSAR analysis.  Pharm Res 2002; 19: 1611–21.

26  Deretey E, Feher M, Schmidt JM.  Rapid prediction of human intestinal absorption.  Quant Struct-Act Relat 2002; 21: 493–506.

27  Levet-Trafit B, Gruyer MS, Marjanovic M, Chou RC.  Estimation of oral drug absorption in man based on intestine permeability in rats.  Life Sci 1996; 58: 359–63.

28  Zhao YH, Abraham MH, Hersey A, Luscombe CN.  Quantitative relationship between rat intestinal absorption and Abraham descriptors.  Eur J Med Chem 2003; 38: 939–47.

29  Peter B.  Modeling liquid properties, solvation, and hydrophobicity: A molecular size-based perspective.  Perspect Drug Discov Des 2000; 19: 19–45.