



# Prediction of Parallel Artificial Membrane Permeability Assay of Some Drugs from their Theoretically Calculated Molecular Descriptors

E. KONOZ\*, AMIR H. M. SARRAFI and S. ARDALANI

Department of Chemistry, Islamic Azad University  
Central-Tehran Branch, P.O.Box 13185/768-Tehran, Iran

*konozelaheh@gmail.com*

Received 9 December 2010; Accepted 23 January 2011

**Abstract:** Parallel artificial membrane permeation assays (PAMPA) have been extensively utilized to determine the drug permeation potentials. In the present work, the permeation of miscellaneous drugs measured as flux by PAMPA ( $\log F$ ) of 94 drugs, are predicted by quantitative structure property relationships modeling based on a variety of calculated theoretical descriptors, which screened and selected by genetic algorithm (GA) variable subset selection procedure. These descriptors were used as inputs for generated artificial neural networks. After generation, optimization and training of artificial neural network (5:3:1), it was used for the prediction of  $\log F$  for the training, test and validation sets. The standard error for the GA-ANN calculated  $\log F$  for training, test and validation sets are 0.17, 0.028 and 0.15 respectively, which are smaller than those obtained by GA-MLR model (0.26, 0.051 and 0.22, respectively). Results obtained reveal the reliability and good predictability of neural network model in the prediction of membrane permeability of drugs.

**Keywords:** Artificial neural network, Genetic algorithm, Quantitative structure-property relationship, Drugs permeation, Parallel artificial membrane permeability assays

## Introduction

Determination of the intestinal permeability is a key parameter for the selection of compounds for drug discovery. To evaluate the absorption of drugs with diverse structures across a membrane via the transcellular route, their permeability was measured using the parallel artificial membrane permeation assay (PAMPA). PAMPA was proposed by Kansy *et al.* in 1998<sup>1</sup>, which consists of hydrophobic filters coated with lecithin in an organic solvent solution. It is a rapid *in vitro* assay system that evaluates transcellular permeation and is applicable to high throughput screening.

PAMPA has been used for the prediction of oral absorption and blood brain barrier penetration<sup>2-7</sup>. A recent PAMPA (PAMPA-BBB) study using porcine brain lipids has been successfully utilized to improve the prediction of blood-brain barrier (BBB) penetration. The PAMPA-BBB assay is high throughput, accurate, low cost, reproducible and consumes a minimal amount of sample (<0.5 mg)<sup>8</sup>. Previously, the PAMPA permeability of peptide-related compound such as protected peptides, cyclic peptides and indole compounds was measured as models of peptidomimetics<sup>9-11</sup>. One approach to calculate physical properties from molecular structural descriptors is quantitative structure- property relationship (QSAR) methods<sup>13</sup>. QSAR methods yield explicit or implicit correlations which can be applied to compounds structurally related to (but not identical with) the compounds in the original training set. The quantitative structure- activity relationship procedure was applied to the quantitative analysis between chemical structures and PAMPA permeability coefficients. The classical QSAR, the so-called Hansch-Fujita approach is a representative of QSAR methods and has been widely used in medicinal chemistry<sup>12</sup>. The result of QSAR analysis showed that the hydrogen- accepting ability of molecules in addition of hydrophobicity at particular pH was significant in determining the variations in PAMPA permeability coefficients.

Kansy *et al.*<sup>13</sup> reported QSAR models for PAMPA flux of a large set of miscellaneous drugs with or without the addition of glycolic acid. Verma *et al.*<sup>14</sup> developed a QSAR model for the prediction of permeation of drugs by PAMPA at pH 7.4. They correlated the drugs permeations by PAMPA to partition coefficient in 1-octanol/water system (is measure of the overall hydrophobicity of a molecule) and the presence or absence respectively of certain structural features with unusual effects. The functional groups with hydrogen bonding capabilities include the -COOH, -SO<sub>2</sub>NH<sub>2</sub>, atomic -OH and the -N(CH<sub>3</sub>)<sub>2</sub> moieties. Their QSAR model for the estimation of drugs permeation of 94 drugs was shown R<sup>2</sup>=0.721 and standard error was 0.206.

The present work, a QSPR model based on genetic algorithm (GA) and artificial neural network (ANN) techniques was carried out to investigate the permeation of miscellaneous drugs measured as flux by PAMPA<sup>14,15</sup>.

## Theory

### Data set

The data set of permeation drugs through PAMPA was taken from the values reported by Verma *et al.*<sup>14</sup> and shown in Table 1. This data set consist the logarithm of drugs permeation for 94 compounds. Permeation of miscellaneous drugs measured as flux by PAMPA with or without the addition of glycolic acid at pH 7.4. The data set was randomly divided into three groups including training, test and validation sets, which consists of 70, 12 and 12 drugs, respectively. The training set was used for model generation, test set was used for monitoring the extent of overtraining and validation set was used for evaluation of the prediction power for obtained model.

### Molecular descriptors

One important step in QSAR modeling is the numerical representation of molecular structures (often called molecular descriptor of the chemical structure). The built model performance and the accuracy of the results are strongly dependent on the way that the structural representation was performed. At first the structures of the compounds were drawn

with Hyperchem 7.0 program<sup>15</sup>. The geometry optimization was performed with the semi-empirical quantum method AM1<sup>16</sup>. The Hyperchem output files were used by the dragon program to calculate five classes of descriptors: topological, electrostatic, constitutional, geometrical and quantum chemical<sup>17</sup>. Some descriptors generated for each compound were encoded similar information about the molecule of interest. Therefore, it was desirable to test each descriptor and eliminate those that show high correlation ( $R > 0.90$ ) with each other. A total of 167 out of 542 descriptors showed high correlation and were removed from the next consideration. Subsequently GA-MLR variable subset selection method was used for selection of important descriptors. The names of descriptors and the statistical values of constructed GA-MLR model are shown in Table 2. These descriptors were used as inputs for the generated artificial neural network.

**Table 1.** Data set and corresponding observed, GA-MLR and GA-ANN predicted values of logF

No.	Name	logF <sub>exp</sub>	logF <sub>GA-MLR</sub>	logF <sub>GANN</sub>	$\Delta(\log F_{GANN} - \log F_{GA-MLR})$
1	Acetaminophen	1.01	2.03	1.05	0.04
2	Alprenolol	1.81	1.57	1.65	-0.16
3	Amitriptyline	1.72	2.04	1.66	-0.06
4	Aspirin	0.75	1.14	0.96	0.21
5	Atenolol	1.04	1.35	1.10	0.05
6	Atropine	1.57	1.51	1.64	0.07
7	Cefixime	0.70	0.55	0.68	-0.02
8	Cefoxitin	1.10	0.99	0.93	-0.17
9	Ceftriaxone	0.74	0.73	0.81	0.07
10	Chloroquine	1.67	1.84	1.65	-0.02
11	Chlorpropamide	1.65	1.07	0.94	-0.71
12	Chlorprothixene	1.26	1.48	1.46	0.20
13	Chorthalidone	0.79	0.84	0.97	0.18
14	Cimetidine	0.70	1.06	0.75	0.05
15	Codeine	1.67	1.68	1.67	0.00
16	Corticosterone	1.97	2.04	1.78	-0.19
17	Coumarin	1.95	1.43	1.59	-0.36
18	Dapsone	1.73	1.23	1.68	-0.05
19	Diazepam	1.68	1.90	1.68	0.00
20	Diffunisal	1.59	1.49	1.62	0.03
21	Dihydrocodeine	1.58	1.73	1.68	0.10
22	Diltiazem	1.88	1.56	1.47	-0.41
23	Doxorubicin	0.76	0.78	0.69	-0.07
24	Etilefrine	1.13	1.36	1.22	0.09
25	Etofilline	0.97	1.14	1.05	0.08
26	Famotidine	0.83	0.68	0.76	-0.07
27	Furosemide	0.85	1.05	0.87	0.02
28	Hydrochlorothiazide	0.70	0.87	0.65	-0.05
29	Hydroflumethiazide	0.72	0.86	0.75	0.03
30	Ketoprofen	1.58	1.69	1.69	0.11
31	Lorazepam	1.68	1.55	1.69	0.01
32	Metergoline	1.48	1.70	1.69	0.21

*Contd...*

33	Methylprednisolone	1.86	1.94	1.97	0.11
34	Methysergide	1.63	1.46	1.67	0.04
35	Metoclopramide	1.71	1.51	1.66	-0.05
36	Metoprolol	1.54	1.47	1.67	0.13
37	Metronidazole	1.16	0.92	1.00	-0.16
38	Midazolam	1.60	1.62	1.68	0.08
39	Molsidomine	1.15	1.28	1.01	-0.14
40	Morphine	1.39	1.50	1.83	0.44
41	Naloxone	1.74	1.52	1.69	-0.05
42	Naproxen	1.95	1.57	1.68	-0.27
43	Nitrendipine	1.01	1.23	1.02	0.01
44	Nordazepam	1.54	1.72	1.69	0.15
45	Norfloxacin	0.91	1.26	1.04	0.13
46	Omeprazole	1.66	1.40	1.60	-0.06
47	Oxprenolol-HCl	1.65	1.40	1.68	0.03
48	Pentamidine	0.70	0.88	0.78	0.08
49	Phenylbutazone	1.73	1.58	1.68	-0.05
50	Phenytoin	1.61	1.58	1.68	0.07
51	Pindolol	1.66	1.46	1.60	-0.06
52	Primidone	1.51	1.42	1.59	0.08
53	Probenecid	1.16	1.28	1.28	0.12
54	Procainamide	1.45	1.73	1.41	-0.04
55	Propoxyphene	1.90	2.17	1.68	-0.22
56	Proxiphylline	1.04	1.29	1.12	0.08
57	Pyrimethamine	1.79	1.60	1.97	0.18
58	Ranitidine	0.86	1.16	1.00	0.14
59	Saccharin	0.70	0.84	0.99	0.29
60	Salicylic acid	0.70	1.13	0.63	-0.07
61	Scopolamine	1.42	1.43	1.57	0.15
62	Sulfadiazine	0.75	0.71	0.94	0.19
63	Sulfisoxazole	0.70	0.84	0.93	0.23
64	Sulindac	1.15	1.40	1.59	0.44
65	Suprofen	1.12	1.15	0.91	-0.21
66	Tacrine	1.73	1.51	1.63	-0.10
67	Terbutaline	0.94	1.15	0.93	-0.01
68	Testosterone	1.90	2.14	1.66	-0.24
69	Tetracycline	1.26	1.17	1.25	-0.01
70	Theophyllin	1.09	1.10	0.93	-0.16
	Test set				
71	Fluocortolone	1.80	1.85	1.66	-0.14
72	Griseofulvin	1.79	1.67	1.66	-0.14
73	Ydrocortisone	1.96	2.01	1.97	0.01
74	Imipramine	1.82	1.94	1.66	-0.16
75	Papaverine	1.79	1.61	1.66	-0.13
76	Pentoxifylline	1.44	1.18	0.95	-0.49
77	Prednisone	1.88	1.90	1.68	-0.20

*Contd...*

78	Propranolol	1.82	1.57	1.67	-0.15
79	Proquazone	1.78	1.66	1.66	-0.12
80	Warfarin	1.85	1.64	1.65	-0.2
81	Verapamil	1.83	1.65	1.66	-0.17
82	Zolpidem	1.83	1.86	1.66	-0.17
Validation set					
83	Acyclovir	0.94	1.10	0.29	-0.65
84	Alprazolam	1.87	2.06	1.66	-0.21
85	Antipyrine	1.42	1.48	1.41	-0.01
86	Betamethasone	1.84	1.81	1.95	0.11
87	Carbamazepine	1.86	1.49	1.65	-0.21
88	Cocaine	1.93	1.67	1.66	-0.27
89	Isoproterenol	1.08	1.20	1.02	-0.06
90	Lidocaine	1.85	1.67	1.66	-0.19
91	Oxazepam	1.67	1.56	1.66	-0.01
92	Phenobarbital	1.88	1.58	1.66	-0.22
93	Prednisolone	1.90	1.91	1.97	0.07
94	Quinidine	1.94	1.66	1.66	-0.28

### Genetic algorithm

Genetic algorithm (GA) is a stochastic optimization method that has been inspired by evolutionary principles<sup>18-20</sup>. The distinctive aspect of a GA is that it investigates many possible solution simultaneously, each of which explores different regions in parameter space<sup>21</sup>. For the moment, one of the best available tutorial on variable selection using genetic algorithm published by Leardi *et al.*<sup>22</sup>. In the present paper, genetic optimization method was tried following the studies of Rogers *et al.*<sup>23</sup> and Luke<sup>24</sup> with a few minor modifications. In our GA program that was written by MATLAB 7.5 two modifications are made. The first is a stochastic reminder method, which allows an individual with above average fitness to be reproduced at least once<sup>21</sup>. The second is the inclusion of elitism, which protect the fittest individual in any given generation from crossover or mutation during reproduction. The genetic content of this individual simply moves on to the next generation intact. In original studies, the fitness function of the individual was determined by a function related to the residual error in the regression analysis of the training data. Here we try to use varieties of fitness functions, which are proportional to the residual error of the training set, prediction set and the number of selected variables according to the following equation:

$$\text{Fitness} = \frac{1}{\text{SEC} + \text{SEP} + (m)^w} \tag{1}$$

In this equation SEC and SEP are standard error of training and prediction set, respectively<sup>25</sup>, m is the number of variables in the represented model and w is a numerical value that implies the weights of m in the value of fitness. In fact the value of w determine the number of variables exist in the chromosome. Some experiments were done using different value for w. Here for the calculation of the fitness of each chromosome a linear model was constructed using variables consist in each chromosome separately by multiple linear regressions (MLR) and the values of SEC and SEP were calculated using this model. This procedure was applied for each chromosome separately. Descriptors selected by GA and the specifications of constructed GAMLR model are shown in Table 2. Since linear model cannot provide an appropriate model for prediction of logF, therefore, we decided to use ANN as nonlinear feature mapping for QSAR model construction.

**Table 2.** Specification of GA-MLR models

Descriptor	Notation	Coefficient	SE	Mean effect
Mean information index on atomic composition	AAC	-0.640	±0.149	1.550
R autocorrelation of lag 6/ weighted by atomic polarizabilites	R6p	1.481	±0.309	- 0.811
QXX COMMA2 value/ weighted by atomic masses	QXXm	-0.002	±0.001	0.275
centered fragments H attached to heteroatom	H-050	-0.045	±0.020	0.001
Moran autocorrelation - lag8/weighted by atomic Sanderson electro negativities	MATS8e	-0.317	±0.128	- 0.015
Constant		2.267	±0.221	

### *Artificial neural network*

A detailed description of the theory behind a neural network has been adequately described elsewhere<sup>18-24</sup>. In addition we have reported some relevant principles of ANNs in our previous papers<sup>25</sup>. The program for the feed-forward neural network that was trained by back-propagation strategy was written in MATLAB 7.5 in our laboratory. The descriptors appearing in the GA model were used as inputs for generation of the ANN. Therefore the number of inputs in the ANN was five and the number of nodes in the output layer was set to be one. The number of nodes in the hidden layer was optimized. The initial weights were randomly selected from a uniform distribution that ranged between -0.3 and +0.3. The initial biases values were set to be one. These values were optimized during the training of the network. The value of each input was divided to its mean value to bring the values of the input variables into the dynamic range of the sigmoid transfer function in the ANN. Before training, the network was optimized for the number of nodes in the hidden layer, learning rates and momentum. In order to evaluate the performance of the ANN, standard error of calibration (SEC) and standard error of prediction (SEP) were used<sup>26</sup>. Then the network was trained using the training set appear by back-propagation strategy for the optimization of the values of the weights and biases.

### **Results and Discussion**

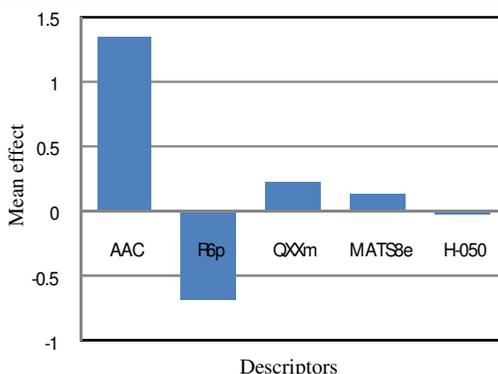
The data set and corresponding observed and GA-ANN predicted values of the permeation of drugs studied in this work are shown in Table 1. For the selection of the most important descriptors genetic algorithm were used. Then these descriptors were used as inputs for generated ANN. Applied GA contained a population of 100 individuals, which evolved for 300 generations. Then by comparison between the fitness values of individuals, the best model was chosen. It can be seen from Table 2 that five descriptors appeared in the best GA-MLR model. These descriptors are: Mean information content on atomic composition (AAC), R autocorrelation of lag 6/ weighted by atomic polarizabilites (R6p), QXX COMMA2 value/ weighted by atomic masses (QXXm), Moran autocorrelation – lag8/weighted by atomic Sanderson electronegativities (MATS8e) and number of hydrogen attached to heteroatom H050). (Table 3 represents the correlation matrix of these descriptors, which represent there aren't any significant correlation between these descriptors.

For inspection of the relative importance and contribution of each descriptor in the model, the value of mean effect (ME) was calculated for each descriptor and shown in the last column of the Table 2 and also were plotted in Figure 1. By interpreting the descriptors in this model, it is possible to gain some insight into factors that are likely related to the permeation of drugs. The first topological descriptor is the mean information index on atomic composition. This descriptor is the mean value of the total information content and is calculated as:

$$I_{AC} = -\sum_g \frac{A_g}{A_h} \cdot \log_2 \frac{A_g}{A_h} = -\sum_g P_g \cdot \log_2 P_g$$

**Table 3.** Correlation matrix for descriptors applying in this work

	AAC	R6p	QXXm	MATS8e	H-050
AAC	1				
R6p	-0.285	1			
QXXm	0.126	0.618	1		
MATS8e	-0.065	0.085	0.089	1	
H-050	0.423	-0.079	0.240	-0.099	1



**Figure 1.** Plot of the mean effects of descriptors

Where  $A^h$  is the total number of atoms (hydrogen included),  $A_g$  is the number of equal-type atoms in the  $g^{th}$  equivalence class, and  $p_g$  is the probability of randomly selecting a  $g^{th}$  type atom<sup>27</sup>. The positive value of mean effect for AAC (+1.550) in the GA-MLR model indicate that this descriptor contribute positively to value of logF of drugs. For example the values of  $I_{AC}$  for tacrine and alprenolol are 1.29 and 1.34, respectively and their logF values are 1.73 and 1.81, respectively.

The second descriptor is R autocorrelation of lag 6/ weighted by atomic polarizabilites. This descriptor is one GETAWAY (GEometry, topology and atom-weights assembly) type descriptors. Molecular descriptors based on the autocorrelation function  $AC_l$ , defined as:

$$AC_l = \int_a^b f(x)f(x+l)dx$$

Where  $f(x)$  is any function of the variable  $x$  and  $l$  is the lag representing an interval of  $x$ ;  $a$  and  $b$  define the total studied interval of the function. A property of the autocorrelation function is that it does not change when the origin of the  $x$  variable is shifted. To obtain spatial

autocorrelation molecular descriptors, function  $f(x_i)$  is any physico chemical property calculated for each atom of the molecule, there polarizability<sup>28</sup>. With increase value of polarizability and this index, log is decreased, that have compatibility with calculated mean effect of this descriptor (-0.811). For example the values of R6p for coumarin and suprofen are 0.09 and 0.27 and logF for these molecules are 1.95 and 1.12, respectively.

The next descriptor is QXXm (QXX COMMA2 value/ weighted by atomic masses) that is one molecular classified as comparative molecular moment analysis method based on the 3D geometrical representation of the molecule. This descriptor can encode calculates different molecular moments with respect to the center of mass, center of charge and center-of-dipole of the molecule<sup>29</sup>. The quadrupole components QXX is calculated with respect to a translated initial reference frame whose origin coincides with the center-of-dipole<sup>30</sup>. By increasing in dipole of molecular, value of this descriptor is increased. Corresponding to positive mean effect for this descriptor (+0.275), by increasing the magnitude of quadrupole moment, log increase. For example the values of QXXm for alprenolol and atenolol are 70.23 and 29.56 and logF for these molecules are 1.81 and 1.04, respectively.

The next descriptor is MATS8e; which is Moran autocorrelation - lag8/weighted by atomic Sanderson electronegativities, that is 2D autocorrelation descriptors. General index of spatial autocorrelation can be defined as:

$$I(d) = \frac{\frac{1}{\Delta} \cdot \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (W_i - \bar{W}) \cdot (W_j - \bar{W})}{\frac{1}{A} \cdot \sum_{i=1}^A (W_i - \bar{W})^2}$$

Where  $w_i$  is any atomic property,  $\bar{W}$  is its average value on the molecule,  $A$  is the atom number,  $d$  is the considered topological distance (*i.e.* the Zag in autocorrelation terms),  $\Delta$ , is a Kronecker delta ( $\delta_{ij} = 1$  if  $\delta_{ij} = d$ , zero otherwise).  $\Delta$  is the sum of the Kronecker deltas, *i.e.* the number of vertex pairs at distance equal to  $d$ <sup>36</sup>. The Moran coefficient usually takes a value in the interval [-1, +1]. Positive autocorrelation corresponds to positive values of the coefficient whereas negative autocorrelation produces negative values. Corresponding to negative mean effect for this descriptor (-0.015), by increasing moran coefficient, log decrease. For example the values of MATS8e for naproxen and chlorthalidone are -0.15 and 0.33 and logF for these molecules are 1.95 and 0.79, respectively.

The last descriptor is H-050 (number of hydrogen attached to heteroatom). The atom-centered fragment code is a short-range atom-centered code that describes each atom by its own atom type and the bond types and atom types of its first neighbors. Functionalities in a molecule can be represented by two to five atoms (corresponding to one bond to 4 bonds), which consist of a central atom and its neighboring bonded atoms<sup>31-34</sup>. Each fragment is represented by a single-value atom-centered fragment descriptor. All atom-centered fragment descriptors representing all fragments in the data set molecules are recorded in an arbitrary but fixed way in a uniform-length multidimensional vector. The use of these substructure descriptors greatly increases the specific chemical information regarding different functional groups, but cannot discriminate between different arrangements of functional groups within a molecule<sup>31</sup>. In general, substructure descriptors are the counts of predefined structural features in the molecules or binary variables specifying their presence/absence<sup>35</sup>. By increasing the number of hydrogen attached to heteroatom the value of H-050 increase and therefore the value of logF increase. For example the values of H-050 for chlorpropamide and dapson are 2 and 4 and log for these molecules are 1.65 and 1.73, respectively.

From the above discussion, it can be seen that all descriptors in the QSAR model has chemical meaning and these can account the structural features that affect on the permeation of the interested drugs.

The next step was the construction of artificial neural network. The selected descriptors were used as inputs for generated ANNs. Before training the network, the parameters of the number of nodes in the hidden layer, weights and biases learning rates and momentum values were optimized. Procedures for the optimization of these parameters were reported in our previous papers<sup>37</sup>. Table 4 shows the architecture and specification of the optimized network. After optimization of the ANN parameters, the network was trained for the adjustment of weights and biases values. Then the trained network was used to evaluate the logF values for molecules in validation set. The ANN and GA-MLR predicted values of logF all molecules in data set are shown in Table 1. Also, the statistical parameters obtained for the GA-MLR and ANN models are shown in Table 5.

**Table 4.** Architecture and specification of the generated ANNs

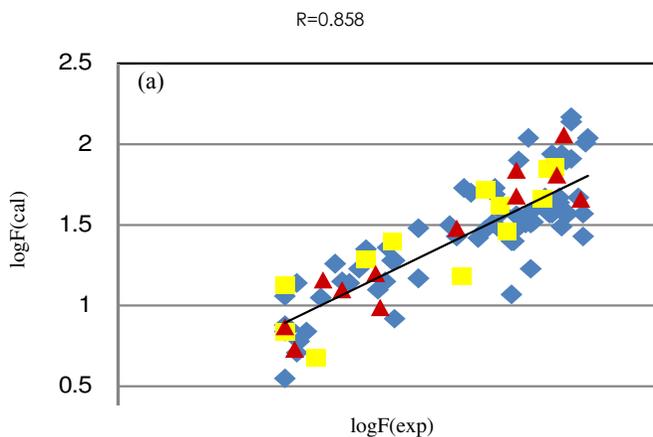
No. of nodes in the input layer	5
No. of nodes in the hidden layer	3
No. of nodes in the output layer	1
Weights learning rate	0.1
Bias learning rate	0.8
Momentum	0.5
Transfer function	Sigmoid

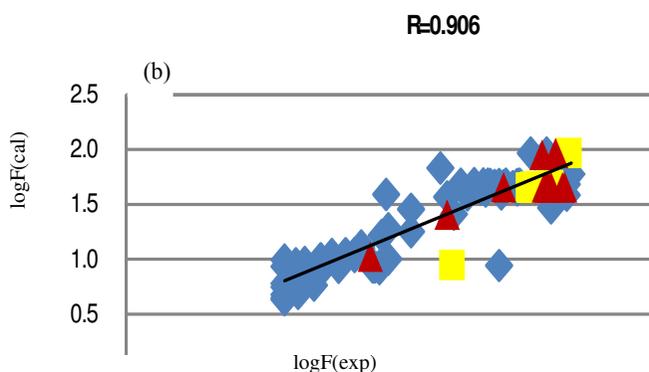
**Table 5.** Comparison of the statistical parameters obtained using the ANN and MLR models<sup>a</sup>

Model	SE <sub>c</sub>	SE <sub>t</sub>	SE <sub>v</sub>	R <sub>c</sub>	R <sub>t</sub>	R <sub>v</sub>	F <sub>c</sub>	F <sub>t</sub>	F <sub>v</sub>
ANN	0.174	0.028	0.151	0.910	0.976	0.911	322	200	48
MLR	0.263	0.051	0.218	0.817	0.953	0.885	22	12	4

<sup>a</sup>F is the statistical F value, R is the correlation coefficient and S.E. is the standard error of model c refers to the calibration(training) set; t refers to the test set; v refers to the validation set.

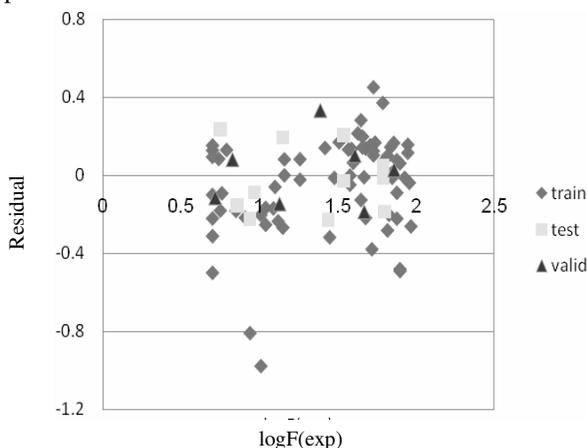
The standard errors of ANN model for training, test and validation sets are 0.17, 0.028 and 0.15, respectively, which would be compared with the values of 0.26, 0.051 and 0.22, respectively, for the GA-MLR model. Figure 2a and 2b shows a plot of the GA-MLR and ANN calculated *versus* the experimental values of logF for the data set molecules.





**Figure 2.** Plot of the (a) GA-MLR and (b) GA-ANN calculated logF against experimental value for all drugs in data set.

Correlation coefficient ( $R$ ) of 0.858 and 0.906 for these plots confirm the suitability of the GA-MLR and ANN models to predict of logF, respectively. Comparison between these values and other statistical parameters in table 5 reveals that nonlinear ANN model produced better results with good predictive ability than linear model. The residual of the GA-ANN calculated values of the logF are plotted against their experimental values in Figure 3. The propagation of the residuals on both sides of zero line indicates that no systematic error exists in the development of the neural network. Acetaminophen, acyclovir and chlorpropamide have high residual value. Figure 3. shows that predicted values for these molecules are not very good. By study structures of these molecules, it was find that acetaminophen has a very simple structure (compared with the training set), acyclovir has a  $N$ -penta and hexagonal structure and chlorpropamide is a amid that are bonded chloride and sulfide groups, so that they have different structures with other molecule. The overall root-mean square error for the GA-ANN model was 0.17, which was less than those obtained by grammatical *et al.*<sup>38</sup> as well as GA-MLR model (0.26). Results obtained reveals that there are some nonlinear relation between the permeation of drugs and the selected structural molecular descriptors.



**Figure 3.** Plot of the residuals *versus* experimental values of logF for all molecules in data set

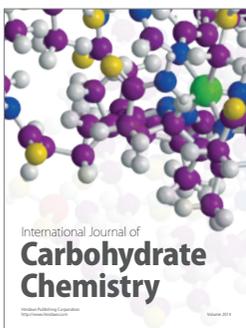
## Conclusion

In the present study, genetic algorithm techniques with artificial neural network approaches were used to develop the QSAR model for prediction of permeation of drugs by PAMPA. The effectiveness of the evolutionary programming algorithm is demonstrated by the selection of the best set descriptors. The key strength of neural networks is their ability to allow for flexible mapping of the selected features by manipulating their functional dependence implicitly, unlike regression analysis. Neural network handles both linear and nonlinear relationships without adding complexity to model. The result suggests that a small number of chemically meaningful descriptors will provide the most predictive QSPR. The statistical results showed that the best model was GA-ANN that combines genetic algorithm as variable selection technique and artificial neural network as feature mapping method. From the analysis, we can conclude that the topological, geometrical, GETAWAY, 2D autocorrelation and atom-centered fragments descriptors have an overall good modeling capability, providing their usefulness in QSPR studies. These descriptors contain local or distributed information molecular structure, so in most case more than one type of these descriptors are needed to reach an acceptable modeling power. Finally descriptors appearing in these QSPR models related to different molecular properties, which can participate in the permeation of drugs by PAMPA.

## References

1. Kansy M, Senner F and Gubernator K, *J Med Chem.*, 1998, **41**, 1007-1010.
2. Sugano K, Hamada H, Machida M and Ushio H, *J Biomol Screen.*, 2001, **6**, 189-196.
3. Sugano K, Takata N, Machida M, Saitoh K and Terada K, *Int J Pharm.*, 2002, **241**, 241- 251.
4. Sugano K, Nabuchi Y, Machida M and Aso Y, *Int J Pharm.*, 2003, **257**, 245-251.
5. Zhu C, Jiang L, Chen T M and Hwang K K A, *Eur J Med Chem.*, 2002, **37(5)**, 399-407.
6. Veber D F, Johnson S.R, Cheng H Y, Smith B.R, Ward K.W and Kopple K D, *J Med Chem.*, 2002, **45**, 2615-2623.
7. Kerns E H, *J Pharm Sci.*, 2001, **90(11)**, 1838-1858.
8. Kerns E H, Fan K, McConnell O J and Carter G T, *Eur J Med Chem.*, 2003, **38**, 223-232.
9. Obata K, Sugano K, Machida M and Aso Y, *Drug Dev Ind Pharm.*, 2004, **30**, 181-185.
10. Ano R, Kimura Y, Urakami M, Shima M, Matsuno R, Ueno T and Akamatsu M, *J Bioorg Med Chem.*, 2004, **12(1)**, 249-255.
11. Ano R, Kimura Y, Shima M, Matsuno R, Ueno T and Akamatsu M, *Bioorg J Med Chem.*, 2004, **12(1)**, 257-264.
12. Hansch C and Am Fujita T, *J Chem Soc.*, 1964, **86**, 1616-1626.
13. Kansy M, Fischer H, Kratzat K, Senner F, Wagner B and Parrilla I, *Verlag Helvetica Chemica Acta*, 2001, 447-464.
14. Verma R P, Hansch C and Selassie C D, *J Comput Aided Mol Des.*, 2007, **21(1-3)**, 3-22.
15. Hyperchem 7.0, Molecular Modelling System, Hypercube Inc., 2002.
16. Dewar M J S, Zoebisch E G, Healy E F and Stewart J J P, *J Am Chem Soc.*, 1985, **107**, 3902-3909.
17. Todeschini R, Consonni V, Mauri A and Pavan M, DRAGON Software Version 3.0, 2003.
18. Shaffer R E and Small G W, *J Anal Chem.*, 1997, **69**, 236A-242A.
19. Leardi R and Gonzalez A L, *Chemom Intell Lab Sys.*, 1998, **41**, 195-201.

20. Holland J, *Adaptation in Natural and Artificial System*, The University of Michigan Press, Ann Arbor, MI, 1975.
21. Cartwright H M, *Applications of Artificial Intelligence in Chemistry*, Oxford University Press, Oxford, 1993.
22. Leardi R and Gonzalez A L, *Chemom Intell Lab Sys.*, 1998, **41**, 195-207.
23. Hagan M T, Demuth H B and Beal M, *Neural Network Design*, 1996, **4**, 2136-2141.
24. Bose N K and Liang P, *Neural Network Fundamentals*, McGraw-Hill, New York, 1996.
25. Jalali-Heravi M and Fatemi M H, *Anal Chim Acta.*, 2000, **415**, 95-103.
26. Blank T B and Brown S T, *Anal Chim.*, 1993, **65**, 3084-3091.
27. Dancoff S M and Quastler H, *Essays on the Use of Information Theory in Biology*, University of Illinois, Urbana (ILL)., 1953, 263-272.
28. Todeschini R and Consonni V, *Handbook of Molecular Descriptors*, Wiley- VCH, Weinheim, Volume 11, 2000.
29. Silverman B D and Platt D E, *J Med Chem.*, 1996, **39**, 2129-2140.
30. Silverman B D, Pitman M C, Platt D E and Rigoutsos I, *J Biomolecular Structure and Dynamics*, 1999, **16**, 1169-1175.
31. Adamson G W, Lynch M F and Town W G, *J Chern Soc (C)*, 1970, 3702-3706.
32. Hodes L, Hazard G F, Geran R I and Richman S, *J Med Chem.*, 1977, **20**, 469-475.
33. Hodes L, *J Chem Inf Comput Sci.*, 1981, **21(3)**, 128-132.
34. Baumann K and Clerc J T, *J Anal Chim Acta.*, 1997, **348**, 327-343.
35. Hagadone T R, *J Chem Inf Comput Sci.*, 1992, **32**, 515-521.
36. Moran P A P, *Biometrika*, 1950, **37**, 17-23.
37. Fatemi M H, Jalali-Heravi M and Konuze E, *Anal Chim Acta*, 2003, **486**, 101-108.
38. Gramatica P, Pilutti P and Papa E, *QSAR Comb Sci.*, 2003, **22**, 364-373.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

