



Prediction of Drug Lipophilicity Using Back Propagation Artificial Neural Network Modeling

SAADI SAAIDPOUR*

Department of Chemistry, Faculty of science, Islamic Azad University, Sanandaj Branch, Sanandaj, Iran.

*Corresponding author E-mail: sasaaidpour@iausdj.ac.ir, saadisaaidpour@gmail.com

<http://dx.doi.org/10.13005/ojc/300253>

(Received: February 28, 2014; Accepted: March 30, 2014)

ABSTRACT

A quantitative structure–property relationship (QSPR) study was performed to develop models those relate the structures of 150 drug organic compounds to their *n*-octanol–water partition coefficients ($\log P_{o/w}$). Molecular descriptors derived solely from 3D structures of the molecular drugs. A genetic algorithm was also applied as a variable selection tools in QSPR analysis. The models were constructed based on 110 training compounds, and predictive ability was tested on 40 compounds reserved for that purpose. Application of the developed models to a testing set of 40 drug organic compounds demonstrates that the new models are reliable with good predictive accuracy and simple formulation. Modeling of logarithm of $\log P_{o/w}$ of these compounds as a function of the theoretically derived descriptors was established by artificial neural network (ANN). The neural network employed here is a connected back-propagation model with a 4-4-1 architecture. Four descriptors for these compounds molecular volume (MV) (Geometrical), hydrophilic-lipophilic balance (HLB) (Constitutional), hydrogen bond forming ability (HB) (Electronic) and polar surface area (PSA) (Electrostatic) are taken as inputs for the models. The use of descriptors calculated only from molecular structure eliminates the need for experimental determination of properties for use in the correlation and allows for the estimation of $\log P_{o/w}$ for molecules not yet synthesized. The prediction results are in good agreement with the experimental value. The root mean square error of prediction (RMSEP) and square correlation coefficient (R^2) for ANN model were 0.1838, 0.9876 for the prediction set $\log P_{o/w}$, respectively.

Key words: *n*-Octanol–water partition coefficients; Quantitative structure–property relationship; Genetic algorithm; Artificial neural network; Prediction.

INTRODUCTION

The *n*-octanol/water partition coefficient is the ratio of the concentration of a chemical in *n*-octanol to that in water in a two-phase system at equilibrium. The logarithm of this coefficient, $\log P_{o/w}$, has been shown to be one of the key parameters in quantitative structure-activity / property

relationship (QSAR/QSPR) studies. The octanol-water partition coefficient is a measure of the hydrophobicity and hydrophilicity of a substance. Hydrophobic interactions are of critical importance in many areas of chemistry, including enzyme-ligand interactions, drug-receptor interactions, transport of drug to the active site, the assembly of lipids in bio-membranes, aggregation of

surfactants, coagulation, and detergency, etc^{1,2}. Hydrophobic “bonding” is actually not bond formation at all, but rather the tendency of hydrophobic molecules or hydrophobic parts of molecules to avoid water because they are not readily accommodated in the highly ordered hydrogen bonded structure of water³. Hydrophobic interaction is favored thermodynamically because of increased entropy of the water molecules that accompanies the association of non-polar molecules, which squeeze out water. The hydrophobic “bonding” resulting from an unwelcome reception of non-polar molecules in water involves vander Waals forces, hydrogen bonding of water molecules in 3D structure, and other interactions⁴.

Hydrophobicity affects drug absorption, bioavailability, hydrophobic drug-receptor interactions, metabolism of molecules, as well as their toxicity. Hydrophobicities of solutes are readily determined by measuring partition coefficients ($\log P_{o/w}$) using the shake-flask method (which involves distribution of a compound between an aqueous phase and an organic phase) and reversed phase high performance liquid chromatography^{5,6}. Since the experimental determination of the partition coefficient of a large set of compounds is a very tedious job, several methods of calculations of $\log P_{o/w}$ values have been proposed by different groups of authors, e.g., Rekker *et al.*⁷, Leo *et al.*^{8,9}, Crippen *et al.*¹⁰⁻¹¹, Bodor *et al.*¹², Klopman *et al.*¹³, Moriguchi *et al.*¹⁴, and Suzuki *et al.*¹⁵. Many studies on the modeling of $\log P_{o/w}$ values using topological, topographic, quantum chemical, and other descriptors have been reported where $\log P_{o/w}$ values have been the response variable to explore suitability of the descriptors/schemes in QSPR studies¹⁶⁻²².

There are some reports about the applications of ANN²³⁻²⁶ modeling to predict the *n*-octanol /water partition coefficient of organic compounds. In our previous papers, we reported on the application of QSPR techniques in the development of a new, simplified approach to prediction of compounds properties²⁷⁻³¹. Experimental determination of $\log P_{o/w}$ is often complex and time-consuming and can be done only for already synthesized compounds. For this

reason, a number of computational methods for the prediction of this parameter have been proposed. In this work a QSPR study is performed, to develop models that relate the structures of a heterogeneous group of 150 drug compounds to their *n*-octanol-water partition coefficients. The genetic algorithm was used to select the most informative descriptors from the calculated descriptors by Molecular Modeling Pro Plus software. The selected descriptors were used to develop an ANN model for predicting the $\log P_{o/w}$ for 40 drug compounds at 25 °C. The key aim of this work is to investigate molecular descriptors important in determining *n*-octanol-water partition coefficients.

MATERIALS AND METHODS

The QSPR model for the estimation of the $\log P_{o/w}$'s of various drug organic compounds is established in the following six steps: the molecular structure input and generation of the files containing the chemical structures is stored in a computer-readable format; quantum mechanics geometry is optimized with a semi-empirical (AM1) method; structural descriptors are computed; structural descriptors are selected; and the structure- $\log P_{o/w}$ model is generated by the ANN and statistical analysis.

Data set

All $\log P_{o/w}$ data for all 150 compounds was taken from the literature³². The data set was split into a training set (110 compounds) and a prediction set (40 compounds). The $\log P_{o/w}$'s of these compounds are deposited in Journal log as supporting material (Table 1 and 2).

Computer hardware and software

All calculations were run on a Dell Inspiron N5010 laptop computer with intel® Core™ i7 processor with Windows 7 operating system. The ChemDraw Ultra version 9.0 (ChemOffice 2005, CambridgeSoft Corporation) software was used for drawing the molecular structures³³. The MOPAC 7.0 (AM1 method) did the optimizations of molecular structures and descriptors were calculated by Molecular Modeling Pro Plus (MMPP) Version 6.3.3 (ChemSW, Inc.) software's^{34,35}. A genetic algorithm procedure was used for selection of descriptors using the STATISTICA Neural Networks software

package³⁶ and other calculations were performed in the MATLAB (version 7.0, MathWorks, Inc.) environment.

Molecular modeling and theoretical molecular descriptors

The derivation of theoretical molecular descriptors proceeds from the chemical structure of the compounds. In order to calculate the theoretical descriptors, molecular structures were constructed with the aid of ChemDraw Ultra version 9.0 and molecular structures were optimized using AM1 algorithm^{37,38}. The computational chemistry software Chem3D Ultra version 9.0 with MOPAC was used to build the molecules and perform the necessary geometry optimizations. A gradient cutoff of 0.01 was used for all geometry optimizations. We have chosen descriptors associated with the neutral molecules of drug in our calculations. As a result, 67 theoretical descriptors were calculated for each compound in the data sets (150 compounds).

Genetic algorithm for descriptor selection

The selection of relevant descriptors, which relate the $\log P_{o/w}$ to the molecular structure, is an important step to construct predictive models. The genetic algorithm was applied to the input set of 67 molecular descriptors for each chemical of the studied data sets and the related response, in order to extract the best set of molecular descriptors, which are, in combination, the most relevant variables in modeling the response of the training set chemicals. Genetic algorithm (GA), included in the STATISTICA Neural Networks, was used for variables selection (based on the training set) [39]. The population size was 100, maximum generation was set to 100, double crossover was set to 0.3 and a mutation rate of 1.0 was used. Finally, we obtained a 4-descriptor subset, which keeps most interpretive information for $\log P_{o/w}$. A total of 4 descriptors were calculated for each drug in the data set. The selected descriptors are molecular volume (MV), hydrophilic-lipophilic balance (HLB), Hansen's hydrogen bond forming ability (HB) and polar surface area (PSA).

Artificial neural network

An artificial neural network (ANN) or commonly just neural network (NN) is an

interconnected group of artificial neurons that uses a mathematical model or computational model for information processing based on a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network⁴⁰. We use a three-layer neural network with a number of neurons in the input layer equal to the number of parameters, a variable number of neurons in the hidden layer and one neuron in the output layer. The training of the neural network is done by the back-propagation algorithm⁴¹ using a STATISTICA Neural Networks software. Back propagation is the best known training algorithm for neural networks, and still one of the most useful. Back-propagation is a gradient descent on the error surface, the weights of the connections between neurons being adjusted in order to decrease the root mean squared error (rmse) between calculated and expected values for all molecules in the database. A back propagation ANN (multilayer perceptron or MLP) was selected as an empirical model.

Our inputs consist of variables describing the structure, and our target data are values of $\log P_{o/w}$; thus, a supervised learning method should be used. In this study, we will try to find a model that can predict the $\log P_{o/w}$ for each set of the four input variables, MV, HLB, HB and PSA for any drug compounds. Hence, our network requires four input units and one output neuron. As in most applications, one hidden layer turns out to be sufficient; after some trial and error, 4 neurons were placed into the hidden layer. The (4-4-1) neural network (Fig. 1), with one hidden and output layer was trained with 110 drugs by the back propagation algorithm.

RESULTS AND DISCUSSION

All descriptors were calculated for the neutral species. The $\log P_{o/w}$ is assumed to be highly dependent upon the MV, HLB, HB and PSA. The correlation coefficients between experimental $\log P_{o/w}$ and the MV, HLB, HB and PSA are 0.6777, -0.8893, -0.8058 and -0.7512, respectively. In the present study, the QSPR model was generated using a training set of 110 molecules. The test set of 40 molecules (Table 2) with regularly distributed

Table 1: Experimental values of $\log P_{ow}$ for drug organic compounds at 25 °C for training set

No.	Drug	$\log P_{ow}$	No.	Drug	$\log P_{ow}$
1	Amoxicillin	-1.71	56	Metoclopramide	2.34
2	Leucine	-1.55	57	Mebendazole	2.42
3	Gabapentin	-1.25	58	Lidocaine	2.44
4	Morphine-3-B-D-Glucuronide	-1.10	59	Carbamazepine	2.45
5	AMITROL	-0.97	60	Lorazepam	2.48
6	Hydrochlorothiazide	-0.66	61	Erythromycin	2.54
7	Maleic hydrazide acid	-0.56	62	Omeprazole	2.57
8	Chlorothiazide	-0.52	63	Alprazolam	2.61
9	Pyridoxine	-0.50	64	2,4-Dichlorophenoxyacetic acid	2.78
10	Acetic acid	-0.30	65	bupirone	2.78
11	Hydroflumethiazide	-0.10	66	Metipranolol	2.81
12	Cefixime	0.25	67	Diltiazem	2.89
13	Fluconazole	0.50	68	Deprenyl	2.90
14	Methotrexate	0.54	69	Prostaglandin E2	2.90
15	Imidacloprid	0.59	70	3-Bromoquinoline	2.91
16	Doxorubicin	0.65	71	Papaverine	2.95
17	Moxonidine	0.90	72	Alprenolol	2.99
18	Rufinamide	0.90	73	Primaquine	3.00
19	Clopyralid	1.07	74	Cocaine	3.01
20	Pilocarpine	1.20	75	Clonazepam	3.02
21	Ketorolac	1.26	76	p-F-Deprenyl	3.06
22	Ranitidine	1.28	77	Proquazone	3.13
23	Procainamide	1.42	78	Nordiazepam	3.15
24	Norfloxacin	1.49	79	Ketoprofen	3.16
25	Coumarin	1.53	80	Diphenhydramine	3.18
26	6-Acetylmorphine	1.55	81	Fluazifop	3.18
27	Prednisone	1.56	82	Bupropion	3.21
28	Clonidine	1.57	83	Mecoprop	3.21
29	Acetophenone	1.58	84	Pirimiphos-methyl	3.27
30	Flumazenil	1.64	85	Chlorpheniramine	3.39
31	N-Methylaniline	1.65	86	loxynil	3.43
32	Ergonovine	1.67	87	Phenylbutazone	3.47
33	Piroxicam	1.68	88	Progesterone	3.48
34	Oleandomycin	1.69	89	Propranolol	3.48
35	Pirimicarb	1.71	90	Indomethacin	3.51
36	Flumequine	1.72	91	Nitrendipine	3.59
37	Chlorsulfuron	1.79	92	Clofibrate	3.65
38	Pindolol	1.83	93	Haloperidol	3.67
39	Atropine	1.89	94	Chlorambucil	3.70
40	Benzocaine	1.89	95	Desipramine	3.79
41	Cromolyn	1.95	96	Azithromycin	3.87
42	Benzoic acid	1.96	97	Flurbiprofen	3.99
43	Amylobarbitone	2.01	98	Cozapine	4.10
44	Mefluidide	2.02	99	Ibuprofen	4.13
45	Betamethasone	2.06	100	Fluvastatin	4.17
46	Fluocortolone	2.06	101	Bromocriptine	4.20
47	Pentobarbital	2.08	102	Propoxyphene	4.37
48	Penicillin v	2.09	103	Imipramine	4.39
49	Procaine	2.14	104	Nortriptyline	4.39
50	Bisoprolol	2.15	105	Penbutolol	4.62
51	Quinoline	2.15	106	Chloroquine	4.69
52	Griseofulvin	2.18	107	Bifonazole	4.77
53	Hydrocortisone-21-acetate	2.19	108	Fenpropimorph	4.93
54	Ethirimol	2.22	109	Chlorpromazine	5.40
55	Phenytoin	2.24	110	Chlorprothixene	6.03

Table 2: Molecular descriptors, experimental $\log P_{ow}$, predicted $\log P_{ow}$ and residuals values for external prediction set by ANN method

No.	Drug	MV	HLB	HB	PSA	LogPexp	ANN	Residual
1	Ascorbic acid	96.40	21.54	19.47	110.38	-1.85	-1.92	-0.07
2	N-Methyl-D-glucamine	109.22	20.45	18.15	113.18	-1.31	-1.49	-0.18
3	Morphine-6-B-D-Glucuronide	133.60	14.18	17.82	131.50	-0.76	-0.76	0.00
4	Etofylline	116.57	15.97	17.22	88.37	-0.49	-0.53	-0.04
5	Doxycycline	157.50	12.62	16.82	110.50	0.42	0.32	-0.10
6	Acetylsalicylic acid	120.00	13.18	11.68	69.92	0.90	0.84	-0.06
7	Captopril	116.64	13.48	10.81	63.93	1.02	0.97	-0.05
8	Ephedrine	110.90	9.09	12.37	50.40	1.13	1.19	0.06
9	Labetalol	167.80	10.53	14.93	98.74	1.33	1.17	-0.16
10	Phenobarbital	126.40	11.82	8.75	84.75	1.53	1.34	-0.19
11	Metsulfuron Me	171.20	12.95	12.86	95.10	1.58	1.41	-0.17
12	Bromazepam	138.67	12.41	8.42	57.51	1.65	2.02	0.37
13	Prednisolone	174.70	7.12	14.42	101.15	1.69	1.65	-0.04
14	Imazaquin	166.66	9.22	9.58	97.97	1.86	2.11	0.25
15	Celiprolol	179.50	10.20	10.96	97.22	1.92	2.10	0.18
16	Acebutolol	185.60	11.48	10.92	93.98	2.02	2.18	0.16
17	Pentamidine	180.80	10.81	10.14	89.90	2.08	2.32	0.24
18	Methylprednisolone	182.20	6.73	13.56	87.70	2.10	2.19	0.09
19	Prazosin	186.30	12.76	12.07	86.80	2.16	2.02	-0.14
20	Naloxone	174.94	7.67	13.41	73.16	2.23	2.20	-0.03
21	4-Phenylbutylamine	127.00	6.52	8.31	42.10	2.39	2.54	0.15
22	Oxprenolol	159.40	7.44	11.16	50.72	2.51	2.58	0.07
23	Lormetazepam	161.61	5.03	12.05	56.06	2.72	2.62	-0.10
24	5-Phenylvaleric acid	144.50	5.65	7.45	40.46	2.92	3.17	0.25
25	Fomesafen	188.53	9.50	9.78	75.00	3.00	2.89	-0.11
26	Flamprop	160.23	5.49	9.28	63.93	3.09	2.87	-0.22
27	Nifedipine	181.35	8.39	8.48	84.80	3.17	2.88	-0.29
28	Prostaglandin E1	206.20	5.76	11.63	82.90	3.20	3.18	-0.02
29	Naproxen	144.50	3.69	8.73	49.69	3.24	3.00	-0.24
30	Ethinylestradiol	164.46	3.58	11.70	40.46	3.42	3.10	-0.32
31	Hydroxyzine	209.98	8.09	10.68	35.94	3.55	3.93	0.38
32	Pericyazine	201.72	7.06	11.28	52.19	3.65	3.50	-0.15
33	Amlodipine	209.60	7.66	9.58	63.60	3.74	3.71	-0.03
34	Promethazine	161.44	2.31	6.75	22.80	4.05	4.20	0.15
35	Carvedilol	224.00	6.10	10.27	58.20	4.14	4.13	-0.01
36	Ketoconazole	207.50	4.84	9.30	55.20	4.34	4.08	-0.26
37	Amitriptyline	169.12	0.87	5.51	11.40	4.62	4.85	0.23
38	Metergoline	227.94	5.14	8.07	49.66	4.75	4.76	0.01
39	Buprenorphine	230.90	1.59	9.00	52.70	4.82	4.92	0.10
40	Clotrimazole	185.32	3.07	5.87	17.82	5.20	4.85	-0.35

Table 3: Statistical parameters obtained by applying the ANN model to the test set.

Parameter	R ²	SEP	%REP	RMSEP	$\overline{RE}\%$
Value	0.9876	0.1861	7.8427	0.1838	-0.0087

$\log P_{o/w}$ values was used to assess the predictive ability of the QSPR model produced in the ANN modeling.

ANN analysis

The $\log P_{o/w}$ data and molecular descriptors were divided into training (110) and test (40) data sets. A selection of 110 compounds of the remaining 150 compounds made up the training set. A small number of molecular descriptors (MV, HLB, HB and PSA) proposed were used to establish a QSPR model.

For evaluation of the predictive power of the generated ANN, the optimized model was applied for prediction of $\log P_{o/w}$ values of 40 compounds in the prediction set which were not used in the optimization procedure. For the

constructed model, the predictive ability of the ANN model was evaluated by calculation of statistical parameters.

The molecular descriptors, experimental $\log P_{o/w}$, predicted $\log P_{o/w}$ and residuals values for external prediction set by ANN method are presented in Table 2. The plots of predicted $\log P_{o/w}$ versus experimental $\log P_{o/w}$ and the residuals (experimental $\log P_{o/w}$ - predicted $\log P_{o/w}$) versus experimental $\log P_{o/w}$ value, obtained by the ANN modeling, and the random distribution of residuals about zero mean are shown in Figure 2. The stability and validity of model was tested by prediction of the response values for the prediction set. This model is applicable for prediction of $\log P_{o/w}$ from -1.85 to 5.20.

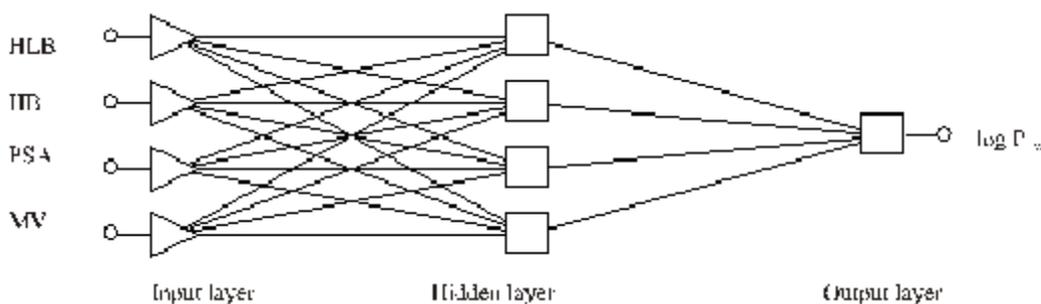


Fig. 1: Network architecture for studying the $\log P_{o/w}$ of drug

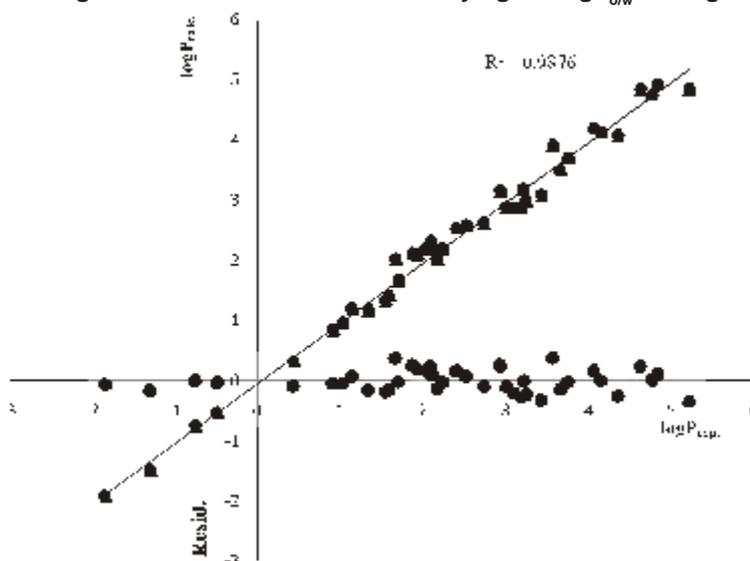


Fig. 2: Plots of predicted $\log P_{o/w}$ and residuals $\log P_{o/w}$ estimated by ANN modeling versus experimental $\log P_{o/w}$ for test molecules in prediction

The average relative errors ($\overline{RE\%}$) of prediction for ANN is -0.0087%. The optimal back-propagation neural network for $\log P_{o/w}$ had a 4-4-1 architecture (Figure 1) was chosen to correlate weighted inputs and outputs of the hidden layer. The optimal neural network architecture was then tested using test data subset as described above.

Interpretation of Descriptors

The QSPR developed indicated that molecular volume (MV), hydrophilic-lipophilic balance (HLB), hydrogen bond forming ability (HB) and polar surface area (PSA) significantly influence drug n-octanol/water partition coefficients.

The molecular volume and the molecular surface area are used mostly as bulk/cavity terms. There is no unique way to define the molecular volume or surface area, but most approaches try to define a surface contour similar to the vander Waals volume. Molecular volume determines transport characteristics of molecules, such as intestinal absorption or blood-brain barrier penetration. Volume is therefore often used in QSPR studies to model molecular properties and biological activity. The steric effects characterize bulk properties of a molecule and can be described with molecular volume. The molecular volume is clearly the most important descriptor for $\log P_{o/w}$. In order for a solute to enter into aqueous solution, a cavity must be formed in the solvent for the solute molecule to occupy. Water as a solvent would much prefer to interact with itself or other hydrogen bonding or ionic species than with a non-polar solute, so there is an increasing penalty (and thus higher $\log P_{o/w}$) for larger solutes. By increasing molecular volume leads to increasing cavity formation energy in water, the larger the solute, the greater the energy demand to make cavity and the lower the solubility in water. According to, increasing the MV increases extent of $\log P_{o/w}$ of the drug organic compounds.

A parameter of utmost importance in the development of pharmaceutical emulsions is the evaluation of their critical hydrophilic-lipophilic balance (HLB). A number on the scale of one to 40 according to the HLB system, introduced by W.C. Griffin^{42,43}. The HLB system is based on the concept that some molecules have hydrophilic groups, other molecules have lipophilic groups, and some have both.

Hydrophilic compounds have a high HLB value (generally over 10), whereas lipophilic compounds have values ranging from 1 to 10. Compounds with self-balance between their lipophilic and hydrophilic portions are extremely efficient as emulsifying agents because they tend to concentrate at the oil/water interface. The Hydrophilic-lipophilic balance of a compounds is a measure of the degree to which it is hydrophilic or lipophilic, determined by calculating values for the different regions of the molecule, as described by Griffin in 1949 and 1954. Griffin's method for non-ionic compounds as described in 1954 works as follows:

$$HLB = 20 \times \frac{Mh}{M} \quad \dots(4)$$

where Mh is the molecular mass of the hydrophilic portion of the Molecule, and M is the molecular mass of the whole molecule, giving a result on an arbitrary scale of 0 to 20. An HLB value of 0 corresponds to a completely hydrophobic molecule, and a value of 20 would correspond to a molecule made up completely of hydrophilic components. The lower the HLB number the more oil soluble the product; and in turn the higher the HLB number the more water-soluble the product. The results indicate that the HLB increases as $\log P_{o/w}$ decreases.

Hydrogen-bonding may be divided into an electrostatic term and a polarization/charge transfer term. A particularly strong type of polar interaction occurs in molecules where a hydrogen atom is attached to an extremely electron-hungry atom such as oxygen, nitrogen, or fluorine. In such cases, the hydrogen's sole electron is drawn toward the electronegative atom, leaving the strongly charged hydrogen nucleus exposed. In this state the exposed positive nucleus can exert a considerable attraction on electrons in other molecules, forming a protonic bridge that is substantially stronger than most other types of dipole interactions. This type of polarity is so strong compared to other van der Waals interactions, that it is given its own name: hydrogen bonding. Understandably, hydrogen bonding plays a significant role in solubility behavior. Hydrogen bonding not a true bond, but a very strong form of dipole-dipole attraction. The O-H and N-H bonds in

molecular structures are strongly polarized and the positive charge is located on H⁺. In this study, we have a dipolar protic solvent (water) containing hydrogen bond donor (O-H bonds) and hydrogen bond acceptor (lone pairs of oxygen atom). Hydrogen bond donor solutes are simply those containing a hydrogen atom bound to an electronegative atom. Hydrogen bond acceptors solutes are that have a lone pair available for donation, and include N and O atoms in their structures. The hydrogen bonding (Hansen) a measure of the tendency of a molecule to form hydrogen bonds. Intermolecular hydrogen bonds can decrease the hydrophobicity ($\log P_{o/w}$) of a molecule, but identifying the existence of hydrogen bonds between a compound and water. This is easy to understand since a large hydrogen bond forming value means that the corresponding solute can provide more hydrogen bond acceptor sites and/or hydrogen bond donor sites

As the hydrogen bond formation increases, water solubility increases (or *n*-octanol solubility decreases), this is agreed to the fact that water has large dipolarity/polarizability. The intermolecular hydrogen bonding can dramatically influence solubility properties.

The Polar Surface Area (PSA) is defined as the surface sum over of polar atoms, e.g. oxygen, nitrogen, etc, as well as hydrogen atoms attached to them. PSA is a commonly used medicinal chemistry metric for the optimization of cell permeability. Molecules with a greater polar surface area are usually believed to be poor at permeating cell membranes. For molecules to penetrate the blood-brain barrier, PSA should be small. This quantity has been used successfully to correlate and predict several transport properties of drugs. The polar surface area term reflect electrostatic and polarization interactions between the solute and the solvent. These interactions are dipole and induced-dipole interactions, together with higher multipole analogues. Therefore, it seems natural to design descriptors based on the electrostatic potential, which gives a complete description of molecular charge distribution, to investigate these properties. Molecular polar surface area is calculated based on the methodology published by Ertl *et al.*⁴⁴ as a sum of fragment contributions. O- and N- centered

polar fragments are considered. PSA has been shown to be a very good descriptor characterizing drug absorption, including intestinal absorption, bioavailability and blood-brain barrier penetration. The results indicate that the PSA increases as $\log P_{o/w}$ decreases.

Statistical parameters

For evaluation of the predictive power of the generated ANN, the optimized models were applied for prediction of $\log P_{o/w}$ values of test compounds in the prediction set, which were not used in the optimization procedure. For the constructed models, four general statistical parameters were selected to evaluate the prediction ability of the model for $\log P_{o/w}$. For this case, the predicted $\log P_{o/w}$ of each sample in prediction step was compared with the experimental $\log P_{o/w}$.

R-Squared ranges from zero to one. The correlation coefficient values closer to 1 represents the better fit of the model. The square of the correlation coefficient (R^2), which is, indicated the quality of fit of all the data to a straight line is calculated for the checking of test set, and is calculated as:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \dots(5)$$

The predictive applicability of a regression model is described in various ways. The most general expression is the standard error of prediction (SEP) which is given in the following formula:

$$SEP = \left[\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-1} \right]^{0.5} \quad \dots(6)$$

The third statistical parameter was relative error of prediction (REP) that shows the predictive

ability of each component, and is calculated as:

$$REP(\%) = \frac{100}{\bar{y}} \left[\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \right]^{0.5} \quad \dots(7)$$

Root mean square error of prediction (RMSEP) is a measurement of the average difference between predicted and experimental values, at the prediction stage. RMSEP can be interpreted as the average prediction error, expressed in the same units as the original response values. The RMSEP was obtained by the following formula:

$$RMSEP = \left[\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right]^{0.5} \quad \dots(8)$$

where y_i is the experimental $\log P_{o/w}$ of the drug in the sample i , \hat{y}_i represented the predicted $\log P_{o/w}$ of the drug in the sample i , \bar{y} , is the mean of experimental $\log P_{o/w}$ in the prediction set and n is the total number of samples used in the prediction set.

The statistical parameters values of R^2 , SEP, REP (%) and RMSEP of prediction set for the ANN model are shown in **Table 3**.

CONCLUSIONS

Results of this study reveal that ANN can be used successfully in development of a QSPR model to predict the n -octanol/water partition coefficients of heterogeneous set of drug organic compounds. Descriptors appear in these QSPR model provide some information related to different molecular properties, which can participate in the intermolecular interactions that affected on the n -octanol/water partition coefficients. The good agreement between experimental results and predicted values confirms the validity of obtained model. The calculated statistical parameters of these models reveal the superiority of ANN model. The result shows that ANN model can describe accurately the relationship between the structural parameters and n -octanol/water partition coefficients of drug compounds. This procedure allowed us to achieve a precise and relatively fast method for determination of $\log P_{o/w}$ of different series of drug compounds and to predict with sufficient accuracy the $\log P_{o/w}$ of new drug derivatives.

REFERENCES

- Selassie, C. D.; Abraham, D. J.; In Burger's Medicinal Chemistry and Drug Discovery, Wiley, New Jersey, **2003**, 15-18.
- Franke, R.; Theoretical Drug Design Methods, Elsevier, Amsterdam, **1984**, p.130.
- Taylor, P.J.; Hansch, C.; Sammes, P.G.; Taylor, J. B.; In Comprehensive Medicinal Chemistry, Pergamon Press, Oxford, **1990**, 211-212.
- Mannhold, R.; Molecular Drug Properties, Measurement and Prediction, Wiley-VCH, Weinheim, **2008**, 315-327.
- Debnath, A.K.; Ghose, A. K.; Viswanadhan, V. N.; In Combinatorial Library Design and Evaluation, Marcel Dekker, New York, **2001**, 73-129.
- Jurs, P.C.; S.L.; Dixon, E.G.; L.M.; In Chemometric Methods in Molecular Design, Waterbeemd, VCH, Weinheim, **1995**, 25-35.
- Rekker, R.F.; The Hydrophobic Fragment Constants. Its Derivation and Applications. A Means of Characterizing Membrane Systems, Elsevier, Amsterdam, **1977**, 155-198.
- Tayar, N.; Tsai, R. S.; Testa, B.; Carrupt, P.A.; Leo, A.; *J. Pharm. Sci.* **1991**, *80*, 590-598.
- Hansch, C.; Leo, A.; Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington DC, **1995**, 125-161.
- Ghose, A. K.; Crippen, G. M.; *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 21-35.
- Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J.; *J. Phys. Chem.* **1998**, *102*, 3762-3772.
- Bodor, N.; Gabanyi, Z.; Wong, C.; K.; *J. Am. Chem. Soc.* **1989**, *111*, 3783-3786.
- Klopman, G.; Wang, S.; *J. Comput. Chem.*

- 1991, 12, 1025-1032.
14. Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y.; *Chem. Pharm. Bull.* **1992**, *40*, 127-130.
15. Suzuki, T.; Kudo, Y.; *J. Comput. Aided Mol. Des.* **1990**, *4*, 155-198.
16. Kier, L. B.; Hall, L. H.; *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, **1976**.
17. Saxena, A. K.; *Quant. Struct.-Act. Relat.* **1995**, *14*, 142-150.
18. Andrey, A. T.; Kunal, R.; *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 179-186.
19. Wouter, H. J. V.; Eñaut, U. R.; Henk, J. M. V.; Christopher, J. C.; Joop, L. M. H.; *Chem. Res. Toxicol.* **1998**, *11*, 847-854.
20. Estrada, E.; Molina, E.; *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 791-797.
21. Du, Q.; Arteca, G. A.; *J. Comput. Aided Mol. Des.* **1996**, *10*, 133-144.
22. Roy, K.; Saha, A.; *Internet Electron. J. Mol. Des.* **2003**, *2*, 288-305.
23. Huuskonen, J. J.; Livingstone, D. J.; Tetko, I. V.; *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 947-955.
24. Tetko, I. V.; Tanchuk, V. Y.; Villa, A. E. P.; *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1407-1421.
25. Molnar, L.; Keseru, Papp, G. M.; Gulyas, A.; Z.; F.; Darvas, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 851-853.
26. Duprat, A. F.; Huynh, T.; Dreyfus, G.; *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 586-594.
27. Ghasemi, J.; Saaidpour, S.; Brown, S. D.; *J. Mol. Struct. (Theochem)* **2007**, *805*, 27-32.
28. Ghasemi, J.; Saaidpour, S.; *Chem. Pharm. Bull.* **2007**, *55*, 669-674.
29. Ghasemi, J.; Saaidpour, S.; *J. Incl. Phenom. Macrocycl. Chem.* **2008**, *60*, 339-351.
30. Ghasemi, J.; Saaidpour, S.; *QSAR Comb. Sci.* **2009**, *28*, 1245-1254.
31. Ghasemi, J.; Saaidpour, S.; *J. Chromatogr. Sci.* **2009**, *47*, 156-163.
32. Avdeef, A.; *Absorption and drug development: solubility, permeability, and charge state*, John Wiley & Sons, Inc., **2003**, 59-66.
33. ChemOffice 2005, CambridgeSoft Corporation, Web: <http://www.cambridgesoft.com>.
34. Web: <http://www.psu.ru/science/soft/winmopac/>.
35. Web: <http://www.chemsw.com/>.
36. Web: <http://www.statsoft.com/>.
37. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P.; *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
38. Young, D. C.; *Computational Chemistry*, John Wiley & Sons, Inc., **2001**, 36-37.
39. Devillers, J.; *Genetic Algorithms in Molecular Modeling*, Academic Press, London, **1996**, 67-76.
40. Zupan, J.; Gasteiger, J.; *Neural Networks in Chemistry and drug design*, WILEY-VCH, Weinheim, **1999**, 125-139.
41. Rumelhart, D.; Hinton, G. E.; Williams, R. J.; *Nature* **1986**, *323*, 533-536.
42. Griffin, W. C.; *J. Soc. Cosmetic Chem.* **1949**, *1*, 311-326.
43. Griffin, W. C.; *J. Soc. Cosmetic Chem.* **1954**, *5*, 249-256.
44. Ertl, P.; Rohde, B.; Selzer, P.; *J. Med. Chem.* **2000**, *43*, 3714-3717.