ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
26 (2001) 241–254

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

www.elsevier.com/locate/jpba

ANN modeling of the penetration across a polydimethylsiloxane membrane from theoretically derived molecular descriptors

S. Agatonovic-Kustrin [a,*], R. Beresford [b], A. Pauzi M. Yusof [a]

[a] *School of Pharmaceutical Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia*
[b] *School of Pharmacy, University of Otago, PO Box 913, Dunedin, New Zealand*

Received 30 November 2000; accepted 6 February 2001

Abstract

A quantitative structure-permeability relationship was developed using Artificial Neural Network (ANN) modeling to study penetration across a polydimethylsiloxane membrane. A set of 254 compounds and their experimentally derived maximum steady state flux values used in this study was gathered from the literature. A total of 42 molecular descriptors were calculated for each compound. A genetic algorithm was used to select important molecular descriptors and supervised ANN was used to correlate selected descriptors with the experimentally derived maximum steady-state flux through the polydimethylsiloxane membrane (log *J*). Calculated molecular descriptors were used as the ANN's inputs and log *J* as the output. Developed model indicates that molecular shape and size, inter-molecular interactions, hydrogen-bonding capacity of drugs, and conformational stability could be used to predict drug absorption through skin. A 12-descriptor nonlinear computational neural network model has been developed for the estimation of log *J* values for a data set of 254 drugs. Described model does not require experimental parameters and could potentially provide useful prediction of membrane penetration of new drugs and reduce the need for actual compound synthesis and flux measurements.

Keywords: Polydimethylsiloxane membrane; Molecular descriptors; QSAR; ANNs; GNNs

1. Introduction

The skin offers a number of opportunities as a route of administration of drugs, both for topical application in local treatment of skin diseases and for transdermal application of drugs for systemic effects. The general principles governing the use of drugs applied to the skin are the same as those for other routes of drug administration.

The main function of skin is to provide a barrier that protects the body from foreign substances. Before any drug applied topically can act either locally or systematically, it must penetrate the skin's barrier. Skin consists of two layers, the epidermis and the dermis. The epidermis has no

* Corresponding author. Tel.: +60-4-6577888/2696; fax: +60-4-6570017.
E-mail address: nena@usm.my (S. Agatonovic-Kustrin).

0731-7085/01/$ - see front matter
PII: S0731-7085(01)00421-6

capillary blood flow but is made up of several layers of enzymatically active cells. The outermost layer, stratum corneum, behaves like a passive diffusion barrier [1]. It is responsible for limiting the passage of exogenous chemicals across the skin into the systemic circulation. It consists of layers of keratinized dead cells at the skin's surface and two alternating amorphous lipophilic and hydrophilic layers layer in the inner layers [2]. The dermis contains the capillary network that transports the drug to the systemic circulation. Thus, drugs with both lipid and water solubility characteristics have a better chance of diffusing through the epidermis [3]. Solute molecules may penetrate the skin through hair follicles or sweat ducts. However, these ways of penetration are of only minor importance because they represent just a small fraction of the total skin surface area.

Quantification of the flux of drug through skin barriers is the basis of Trans Dermal Drug Delivery (TD DD) and techniques for making such measurements are rapidly increasing in number and sensitivity. TD DD offers several important advantages over the usual periodic modes of drug administration. It is convenient, efficient and has fewer side effects. TD DD can provide a good substitute for the oral route as it bypasses gastrointestinal absorption (pH effects, enzymatic activity and drug interactions) and drug deactivation by digestive and liver enzymes (the 'first-pass effect'). The dosage may thus be lower than the amount given orally, and a constant drug–plasma concentration can be maintained enabling optimum blood concentration–time profile, predictable and prolonged blood levels and reduced frequency of dosage. TD DD can also provide fast route of medication in emergency and, if necessary, an easy and fast way to terminate drug effects since the drug–plasma concentration will quickly decrease when the formulation is removed from the skin.

However, there are also some disadvantages. The main disadvantage is that only relatively potent drugs are able to diffuse through the skin. Some may irritate the skin. Moreover, the environment of the skin, different skin types, and problems with adhesion can cause difficulties in the design of the transdermal delivery systems [4].

Thus, the development of TD DD systems is hindered by a number of difficulties arising from the inherently variable nature of the skin barrier. As a consequence, a therapeutically viable dose is dependent upon a number of factors, including skin conditions (hydration of the skin, temperature, PH, rubbing or injunction) thickness of the stratum corneum and location on the body [5].

Human ex vivo models and artificial membranes have been used for in vitro measurements of drug penetration where concern about the toxicity of novel compounds and costs of screening large number of candidates limit the use of in vivo techniques. Although such models cannot fully replicate in vivo conditions, they can provide useful indications of drug release in vivo. These methods are still expensive and time consuming, and they require large amounts of sample. Thus, a theoretical method that could predict drug penetration through skin with high precision would be of interest.

Knowledge of molecular structure is the key to understanding of the functioning of molecules. Intrinsic to chemistry is the concept that there is a relationship between bulk properties of compounds and their molecular structure that provides a connection between the macroscopic and the microscopic properties of matter. A change in a structure of a molecule usually produces an associated change in its properties. Finding one or more molecular descriptors that explain variations in physico-chemical properties or biological activity has resulted in the development of linear free energy relationships (LFER) [6] and quantitative structure activity/property relationships (QSAR/QSPR) as natural extensions of the LFER approach. QSARs are mathematical models that relate the biological activity of a compound to its physicochemical structure. One of the conditions for a successful QSAR model is that parameters, molecular structure descriptors and investigated activity, have values that are obtained in a consistent manner. Until recently, these models have used primarily empirically based descriptors. However, experimental determination is time consuming and is a subject to experimental variation and errors. A current trend in quantitative structure–property/activity relationship (QSPR/

QSAR) studies is the use of theoretical molecular descriptors that can be calculated directly from molecular structure. Using computational methods to derive them is faster and more convenient. QSAR once quantified, can be used to estimate the properties of other molecules even when their structure is only sketchy.

Potts and Guy [7,8], proposed a two variable QSAR models based on hydrophobicity, i.e. octanol/water partition coefficient (Koct) and molecular size (molecular volume or molecular mass) to predict the percutaneous flux of pharmacological and toxic compounds from their physiochemical properties. The model that was developed provided variable results for the complete data set and had only limited statistical accuracy, possibly because the analyzed data were obtained from several different sources and were associated with the risk of likely erroneous values [9]. Later, the same authors developed more successful model based on the effects of molecular size and hydrogen bond activity [10]. Lien and Gao [11] formulated a general mathematical model, involving partition coefficient, molecular mass and hydrogen bonding to correlate the molecular structures and skin permeability of a wide range of compounds through human skin. Recently, Pugh et al. [12] have developed a model for diffusion across human stratum corneum in terms of molecular weight, H-bonding and electronic charge.

Chen et a. [13,14] studied the penetration of a large number of heterogeneous compounds through a polydimethylsiloxane (PDMS) membrane. They derived different models for various subclasses of the investigated data set, based on measured solubility in isopropyl alcohol and the calculated charges on particular atoms within the molecule. Thus, these models require the experimental measurements of solubility in isopropyl alcohol, as there is no available method for its calculations. Cronin et al. have developed a three parameter QSPR model based on the number of hydrogen bond acceptor and donor groups and sixth-order path molecular connectivity, by use of stepwise regression analysis for the prediction of the maximum steady state flux through the PDMS membrane (log J) [15].

The success of regression analysis in QSAR depends upon an assumed linear relationship between the biological activity and one or more descriptors. However, as the number of descriptors increases, regression analysis becomes more complex. One problem likely to occur in large descriptor sets is repetition in information when descriptors are correlated. Latent variable techniques have become accepted methods of addressing this issue. These techniques include the use of principal components in regression analysis and the method of partial least squares [16]. A second problem encountered in using regression analysis is the a priori assumption of a model form (i.e. quadratic, cubic, use of cross terms, etc.). Thus, variable selection techniques such as stepwise forward and stepwise backward multiple linear regression analysis (MLR) [17] were introduced. However, these approaches can capture only linear relationships between molecular characteristics and predicted functional features. Over the last few years the Artificial Neural Networks (ANNs) modeling technique has attracted increasing interest as a most promising method for classification and multivariate calibration problems [18]. ANNs, in contrast to conventional methods, are capable of recognizing highly nonlinear relationships; hence, they provide an interesting new approach to QSAR and QSPR analysis [19–23].

A goal in this research was to develop a Genetic Neural Network (GNN) model to predict the penetration across PDMS membranes for different compounds using only calculated molecular descriptors. Genetic algorithm (GA) was used to select a subset of the descriptors that best describe the membrane penetration and Artificial Neural Network (ANN) to correlate selected descriptors steady state flux and develop a QSAR. The set of 256 compounds with experimentally derived maximum steady state flux values (J) through PDMS membrane was taken from previous studies of Chen and co-workers. They have determined maximum steady state flux values (J) for selected compounds by a constant accurate protocol under identical laboratory conditions, thus fulfilling the condition of successful QSAR modeling that parameters have values, which are obtained in a consistent manner.

1.1. Artifical Neural Networks

An ANN is a biologically inspired computer program designed to learn from data in a manner emulating the learning pattern in the brain. Most ANN systems are very complex high-dimensional nonlinear information processing systems. ANNs are composed of hundreds of single processing elements (PE), artificial neurons. PEs are connected with coefficients (weights), which constitute the neural structure, and are organised in sets of layers, the input layer, output layer, and hidden layers between. Neural networks gather their knowledge by detecting the patterns and relationships in data and learn (or are trained) through experience with appropriate learning exemplars, not from programming. The input layer neurons receive data from a data file. The output neurons provide the ANN's response to the input data. Hidden neurons communicate only with other neurons. They are part of the large internal pattern that determines solutions to problems and where the network learns interdependencies in the model. Each hidden or output unit has a number of incoming connections from units in the preceding layer. The weighted sum of the inputs simulates activation of the neuron. Thus, what is learned in a hidden neuron is based on all the inputs taken together. The activation signal is passed through an activation function (also known as a transfer function) to produce a single output of the neuron. Transfer function for the hidden units is needed to introduce nonlinearity into the network.

The behavior of a neural network is determined by the transfer functions of its neurons, by the learning rule, and by the architecture itself We have used a supervised network with a back-propagation learning rule. In this type of model, information from inputs (inputs = molecular descriptors) is fed forward through the ANN to optimize the connection weights among neurons. The output of the neuron (molecular descriptors) is related to the summed input by a sigmoid shaped transfer function. During training, optimization of the network weights is made by back-propagation of error (e.g. difference between predicted and measured drug flux), and the inter-unit connections are changed until the error in predictions is minimized across many data sets and until the network reaches a specified level of accuracy. These connection weights store the knowledge necessary to solve specific problems. Once the network is trained and tested it can be given new input information (descriptors) to predict the output (maximum steady-state flux).

2. Experimental

2.1. Software

Neural Networks TM (StatSoft®) was used for building the QSPR model and CAChe Project leader Version 3.11 (Oxford Molecular Ltd.) was used to calculate molecular descriptors from the molecular structure.

2.2. Descriptor generation and analysis

A total of 42 calculated structure features including constitutional, topological, chemical, geometrical and quantum chemical descriptors were calculated for each of the 254 compounds (Table 1). Once the molecular structures were encoded,

Table 1
Calculated structural descriptors

Class	Molecular descriptor
Constitutional descriptors	Functional group counts (amine, aldehyde, amide, carbonyl, carboxylate, cyano, ether, hydroxyl, methyl, methylene, nitro, nitroso, sulfide, sulfone, sulfoxide and thio)
Topological descriptors	Kier and Hall connectivity indices (χ^0–χ^2) and valence connectivity indices (χ^{0V}–χ^V), topological shape indices (κ^{0-2})
Chemical descriptors	Molar refractivity, molecular volume, log *P*, molecular mass, density
Geometrical descriptors	Solvent accessible surface, ring count, ring size
Quantum chemical descriptors	Dipole moment, HOMO and LUMO energies, dielectric energy, steric energy, heat of formation, total energy, minimum energy, electron affinity

Table 2
The influence of different topology on the ANNs performance[a]

ANNS[b]	42/7/2/1	42/7/5/1	30/7/8/1	24/10/8	16/9/1	13/8/1/	12/5/1	11/4/1	8/3/1
UP	0.0000	0.0000	0.0001	0.0005	0.001	0.0015	0.002	0.0025	0.0040
RMS_{tr}	0.31	0.38	0.48	0.44	0.50	0.44	0.36	0.46	0.57
RMS_{test}	0.71	0.67	0.72	0.72	0.66	0.58	0.59	0.75	0.77
RMS_{vai}	0.73	0.74	0.87	0.90	0.53	0.52	0.60	0.66	0.66
Σ	1.76	1.78	2.07	2.05	1.69	1.55	1.55	1.87	2.00

[a] *n*, number ofinputs; UP, unit penalty factor.
[b] Number of inputs — *hidden neurons* — outputs.

the next step in QSAR model development was to select relevant molecular descriptors that best encoded membrane penetration. Using a GA for selection and following a unit penalty factor of 0.000–0.004 the number of inputs was reduced from 42 to 12. Input selection has reduced the size and complexity of the network and focused the training on the most important data. This also reduced the training time and improved the network performance.

2.3. Optimal network architecture

A standard feed-forward network, with back-propagation rule and multilayer perception (MLP) model architecture [24] with maximum two hidden layers was chosen. Although it is possible to approximate any function with just one layer of hidden units, a huge number of hidden units may be required. An additional layer, which drastically reduces the number of hidden units and, consequently, the number of required weights was used to avoid this problem.

The set of 254 structurally different compounds and their experimentally derived values of a maximum steady state flux across a PDMS membrane (log *J*) used in this study was collected from the literature [14,15]. An initial neural network consisting of 42 inputs (calculated molecular descriptors), two hidden layers and one output neuron (log *J*) was used. The number of inputs and hidden neurons was optimized. (Table 2). A set of 200 compounds was selected for training and testing the ANN and 54 compounds were used as an external prediction set. Before each training run, both weights and biases were initialized with random values and 200 compounds in the working data set were split randomly into: the training set containing 160 data sets and testing set containing 40 data sets and the results of the five runs were averaged. During training, the performance of the ANN was evaluated with testing data. The training set was used to train the network and the testing set was used to determine the level of generalization produced by the training set and to monitor over-training the network, each with corresponding rooted mean squared (RMS) error. For an unbiased estimate of the generalization error, the ANN was presented with a validation data set that was not used at all during the training process. Thus, the validity of the QSAR model was evaluated with the validation data set, by predicting membrane penetration for compounds that have not been studied during model development.

2.4. Neural network analysis

The two forms of network analysis are model testing and sensitivity analysis. Both methods were performed concurrently with the training of the network. The testing and training set RMS errors were used to determine overall quality of a particular subset of descriptors (Table 2). Training was stopped, at each run, once the error performance of the network began to deteriorate, based on the training and testing set errors, when the training RMS error fails to improve over a given number of epochs and the testing RMS error starts to increase.

The second form of network analysis computes sensitivities of the network's outputs with respect

to each of its inputs. ANNs compute the output as a sum of nonlinear transformations of linear combinations of the inputs. Sensitivity reports show the sensitivity of the output variables, as a percentage, to the changes in the input variables. If the direction of the change in the output variable is always the same as the change in investigated descriptor, then the average sensitivity is positive. The set of percentages reveals the effect that a change in a particular input has on output.

3. Results and discussion

A first step in this study was to calculate a multitude of structural descriptors as mathematical representations of chemical structure. To describe physico-chemical structure, descriptors are utilized that account for three aspects of the compounds, namely the hydrophobic, electronic and steric effects of the compound. Hydrophobicity is often related to the ability of a compound to partition through a membrane. Steric effects relate to the ability to pass through a membrane or to bind to a receptor site. Electronic effects may relate to reactivity or metabolism. Connectivity and topological descriptors were calculated directly from the two-dimensional connection table representation of the structure, employing methods drawn from mathematical graph theory. Geometric descriptors were calculated from three-dimensional molecular models. Electronic descriptors were derived from empirical or molecular orbital calculations. One of the difficulties with the large number of descriptors is deciding which ones will provide the best regressions, considering both goodness of fit and the chemical meaning of the regression. Following a unit penalty factor of 0.00, 0.0001, 0.0005, 0.001, 0.0015, 0.002, 0.0025 and 0.004, the number of inputs was reduced from 42 to 30, 24, 16, 13, 12, 11, and finally to 8 inputs respectively, and the two hidden layer were replaced with one hidden layer (Table 2).

The best nonlinear GNN model for the estimation of log J values for a data set of 254 drugs was chosen by comparing the prediction obtained from several high scoring models. The model has 12 selected inputs with sensitivity greater that 1% and unit penalty factor greater than 0.0025 (Table 3). The QSPR that was developed indicates that inter-molecular interactions (dipole interaction, electron affinity), hydrogen bonding capacity of drugs (presence of amino and hydroxyl group), molecular shape and size (topological shape indices, molecular connectivity indices, ring count) are important for drug penetration through membranes. Furthermore, conformational stability (total energy, dielectric energy, LUMO) is also found to be an important parameter for membrane penetration. As expected, poor correlation was obtained with solvent-accessible surface areas [25]. However, calculated log P values, which is important in controlling the absorption, distribution, metabolism, elimination of drugs, failed to predict membrane penetration. This could be attributed to the incapacity of most lipophilicity parameters to account for intramolecular interactions, such as intramolecular hydrogen bonding.

Size, shape and symmetry of molecules play a key role in all processes where they interact and associate. The quantification of molecular shape and symmetry helps in understanding how the three-dimensional properties and bulk hindrance influence the interaction between molecules. Over the last 10 years, a variety of topological shape indices and molecular connectivity indices as theoretical descriptors, have emerged as alternative descriptors in quantitative structure–activity studies for characterization of molecular [26,27]. To-

Table 3
Sensitivity analysis of membrane permeability (log J)

Descriptor	Sensitivity
Dielectric energy	0.445
–OH	−0.180
Ring count	−0.091
LUMO	0.051
$-NH_2$	−0.049
EL. affinity	0.041
MW	−0.034
Total energy	−0.027
Dipole	0.012
κ^1	−0.008
χ^1	−0.019
χ^2	−0.014

pological shape indices and molecular connectivity indices are quantitative descriptors of molecular shape and molecular similarity [28]. The advantage of such descriptors is that they can be calculated for any chemical structure, real or hypothetical. The ANN model that was developed included topological shape descriptor of the first order (κ^1) and connectivity indices of the first and second order (χ^1 and χ^2) for effective quantification of molecular shape and bulk properties.

Topological shape indices [29] are the basis of a method of molecular structure quantification in which attributes of molecular shape and size are encoded into three indices (κ values 1–3), properties. Topological indices are suitable for describing similarity or dissimilarity of molecules. If two compounds have close values for a number of indices, they can be regarded as being similar. Topological shape indices or numerical graph invariants are derived from different classes of weighted graphs, representing various levels of chemical structural information. They are numerical quantifiers of molecular topology and encode information regarding the size, shape, branching pattern, cyclicity, and symmetry of molecular graphs. κ values are derived from counts of one-bond, two-bond and three-bond fragments, each count being made relative to fragment counts in reference structures, which possess a maximum and minimum value for that number of atoms. The first order shape index, κ^1, encodes molecular cycles, κ^2 encodes linearity and κ^3 encodes branching. The model shows that increase in κ^1 decreased membrane penetration due to increase in molecular size and lipid solubility.

The molecular connectivity indices, χ values, describe the extent of skeletal branching. Connectivity indices are descriptor of molecular structure, a descriptor of size and shape based on a count of groupings of skeletal atoms, weighted by the degree of skeletal branching. Each carbon atom in a molecule skeleton is assigned a number according to its number of neighboring carbons. The molecular skeleton is then fragmented into all of its two carbon atom bonds. The sum of these values over the structure forms the χ index. Molecules could be further dissected into two bond fragments, three bond fragments and so on. Molecular structure is quantified so that weighted counts of substructure fragments are incorporated into numerical indices and an index is derived from a consideration of pairs of atoms forming bonds. χ^0, zero order (atomic) connectivity indices, provides information about the number of atoms in a molecule. Molecular connectivity index of the first order, χ^1, encodes single bond properties. It is a weighted count of bonds, related to the types and position of branching in the molecule. χ^2 (path) is derived from fragments of two-bond lengths. It also provides information about types and position of branching and may be indicative of the amount of structural flexibility. An increase in branching increases surface area and molecular volume [30] and results in an increased solubility and lowered partition coefficient. QSAR suggest that the increase in χ^1 and χ^2 decrease membrane penetration. Statistical analysis has shown that χ^1 and χ^2 are covariant to an extent. However, there is enough difference between the information in χ^1 and χ^2 to reflect structural features contributing in a different way from the numerical value. χ^2 can differentiate between structural isomers, while χ^1 values are identical. Low values of χ^1 and χ^2 are found for more elongated molecules or those with only one branching atom. An increase in the length of the carbon chain, the nonpolar portion of the molecule, results in an increase in lipid-solubility (log P) and molecular size. The partition coefficient correlates with numerous parameters that represent bulk properties, such as molecular mass, volume and surface area. Molecular size is limiting the absorption through membranes in general. Molecular mass is often used as molecular size descriptor. Medications with a low molecular mass that are unionized and liphophilic have higher flux value simply because diffusion is much easier. It is clear, from input sensitivity analysis, that an increase in molecular topology (κ^1, χ^1, χ^2) and an increase in ring count and molecular mass results in a flux decrease. Mechanistically, a more bulky molecule is less likely to pass through the membrane.

Functional groups exhibit a characteristic reactivity and characteristic chemical behavior when present in a compound. Particular functional groups are the most important in the specific

interaction between a drug and a transporter receptor. Undoubtedly, the functional group accounts for many of the dipole–dipole, dipole-induced dipole and hydrogen bond interactions. For each molecule, the orientation of the functional groups influences the dipole (electrostatic) moment value, and this descriptor can be used as a selector of active conformations. The QSAR models obtained have a large dependence on a number of hydroxyl and amino groups, substituents with polar bonds that are capable of hydrogen bonding. It is shown that properties associated with hydrogen bonding should be kept to a minimum to promote high membrane penetration. Since the presence of hydroxyl and amino groups in the molecule facilitates hydrogen bonding, an increase in the number of hydroxyl and amino groups decreases membrane penetration.

Dipole moment is also a measure of the polarity of a molecule. Electronic effects may relate to molecular reactivity or metabolism [31]. The greater the dipole moment, the greater the activity. Drugs with higher dipole moments have higher maximum steady-state flux through membranes.

Although solubility parameters, topological shape and connectivity indices are often successful in rationalizing solubilities and partition coefficients, they cannot account for conformational changes and they do not provide information about electronic influences through bonds or across space. These electronic influences may play a role in the magnitude of a biological activity, along with structural features encoded in indexes. Therefore, electronic effects are quantified explicitly by the use of molecular orbital calculations to estimate dielectric energy, Lowest Unoccupied Molecular Orbital (LUMO) energy, Highest Occupied Molecular Orbital (HOMO) energy and electron affinity.

The most important molecular descriptor in QSAR was found to be dielectric energy. Dielectric setting, i.e. the change in charge rearrangement of molecules, accompanies the change in hydrogen bond strength [32]. The QSPR model shows that the increase in dielectric energy promotes membrane penetration.

The LUMO energy is often used as a measure for electron affinity of a molecule or its reactivity as an electrophyle. Good electrophyles are those in which the electrons reside in low-lying orbitals. Electrophyles are often reducing agents. High values of LUMO indicate loosely bound electrons, which are reactive to nucleophilic attack. The energy difference between the HOMO and the LUMO energies is related to the minimum energy needed to excite an electron in the molecule and was used as an index of kinetic stability for a variety of polycyclic aromatic hydrocarbons [33]. An increase in electron acceptor properties increases binding affinity.

Electron affinity [34] also incorporates electron correlation and relaxation, whereas LUMO does not, and is also a measure of reduction capacity. Since living organisms function at an optimum redox potential range, it is assumed that redox potential of compounds of a certain type correlate with biological effect. An increase in molecular reactivity also increases metabolic processes. The access of a drug to the sites of oxidation–reduction reactions is hindered by the complex competing events during absorption. Therefore, correlation between redox potential and biological activity is important for compounds with similar structural and physical properties. As expected, compounds with lower total energies, higher LUMO energy and higher electron affinity have higher rates of membrane penetration.

Molecular interactions are determined fundamentally by molecular size, shape and charge distribution. If a molecule has a form similar to a geometrically defined solid, then a qualitative or semi-quantitative description is easily performed, assuming that the dimensions and other geometrical parameters of that solid represent the molecular shape. However, molecules rarely have a unique description and there may be several different forms in equilibrium. Thus, their dynamic structure should be recognized. Molecular modeling software separates the graphics and the simulation. The graphics are used for constructing molecules, editing topology, and visualizing properties, while the simulation carries out energy minimization or dynamics. These quantum chemical descriptors can give great insight into structure

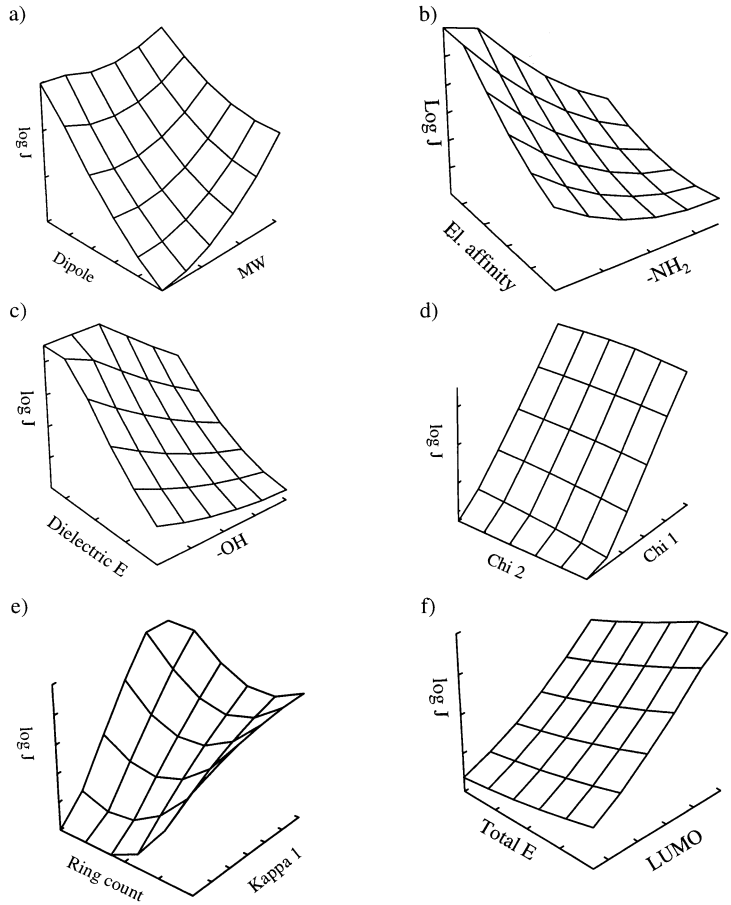

Fig. 1. Functional dependence surfaces for descriptors from the final model. The descriptors are (a) dipole and molecular mass; (b) electron affinity and -NH_2 group count; (c) dielectric energy and -OH group count; (d) χ^1 and χ^2; (e) ring count and κ^1; (f) total energy and LUMO.

and reactivity and can be used to establish and compare conformational stability, chemical reactivity and inter-molecular interactions. Energies were calculated for an optimized conformation with the most stable geometry, or minimum energy structure. Optimization was conducted to find a low energy structure for the steric energy, heat of formation and for total energy. Depending on the procedure used, the calculated energy was steric energy (from mechanics), heat of formation from Molecular Orbital Package (MOPAC) [35,36] and total energy from Zerner's Intermediate Neglect of Differential Overlap program (ZINDO) [37]. The total energy of a system was approximated by summing a series of empirically derived equations, which describe bond stretching, angle bending, torsional, and nonbonded (van der Waals and hydrogen bond) interactions. It was found that compounds with lower total energies had higher membrane penetration steady-state flux.

The interpretation of effects of individual descriptors is difficult as the model is multivariate and nonlinear. However, some insight into the degree of nonlinear behavior of descriptors was assessed with a functional dependence plot. The value of input variables was varied through its range, while all other inputs were held constant. The network output was plotted against two input descriptors to generate a functional depen-

Table 4
Performance of the ANN model (12-5-1) in the prediction of log J values

Systematic name	Experimental log J	Predicted 12-5-1
3-Nitrobenzaldehyde	−2.52	−2.013
2,5-Pyridinedicarboxylic acid	−5.205	−4.896
1-Fluoro-4-nitrobenzene	−1.6	−2.069
4-Aminoquinoline	−3.481	−3.828
2-Ethylimidazole	−2.975	−2.376
2-Thiophenmethanol	−2.179	−0.6193
3-Hydroxypyrimidine	−2.685	−2.701
6-Quinolinecarboxylic acid	−4.672	−4.097
Terephthalic acid	−5.145	−4.551
3,5-Dimethylpyrazole	−1.791	−1.756
1,2,5-Trimethylpyrrole	0.918	−0.7226
2-Methyl-5-nitroimidazole	−4.024	−3.317
Pyrrole	−0.891	−0.9402
4-Nitrobezoic acid	−3.358	−3.478
Diphenyl ether	−1.81	−1.902
Quinoline	−1.4	−1.656
2-Quinolinecarboxylic acid	−3.552	−3.551
7-Nitroindole	−2.659	−3.089
2-Methylimidazole	−2.797	−2.763
6-Hydroxynicotinic acid	−5.1	−4.081
1-Naphthoic acid	−2.985	−3.335
4-Carboxybenzaldehyde	−3.44	−3.115
1-Methylpyrrole	−0.657	−0.6429
2-Methyl-1-phenyl-2-propanol	−1.82	−1.754
2,4-Quinolinediol	−5.489	−4.397
2-Furaldehyde	−1.53	−1.697
Pyridazine	−1.865	−1.81
(2-Chloroethyl)benzene	−1.292	−1.148
Butyrophenone	−1.719	−1.314
8-Aminoquinoline	−2.278	−2.408
2,5-Dimethylfuran	−0.28	−0.5196
1-Methylimidazole	−1.813	−2.155
Benzofuran	−0.948	−1.272
Pyridine	−0.695	−0.7433
6-Chloronicotinic acid	−3.098	−3.124
Aniline	−1.75	−1.161
Pyrazole	−1.597	−1.782
6-Methoxyquinoline	−2.097	−2.304
Biphenyl	−2.05	−1.653
2-Thiophenacetic acid	−2.475	−2.757
2-Thiophenmethylamine	−1.400	−0.9129
Phenol	−1.570	−1.708
3,5-Dichloropyridine	−1.842	−1.209
2-Furoic acid	−2.476	−2.993
Butyl phenyl ether	−1.250	−1.338
Toluene	−0.388	−0.4872
4-Chlorobenzylalcohol	−2.504	−1.878
2,5-Dimethylpyrrole	−1.400	−0.8757
4-Aminophenol	−3.910	−3.307
2,5-Dimethylthiopheme	−0.468	−0.7362
2-Aminobenzylalcohol	−2.630	−3.111
5-Nitro-8-hydroxyquinoline	−4.220	−4.197
2-Hydroxyquinoline	−3.813	−3.581
7-Amino-2,4-dimethyl-1,8-naphridine	−3.663	−4.539
Chlorobenzene	−0.540	−0.6351
Furfuryl alcohol	−1.860	−2.205
2-Methyl-5-nitrobenzimidazole	−3.698	−3.704
4,7-Dichloroquinoline	−2.590	−2.449
Imidazole	−3.019	−2.921
5-Chloro-8-hydroxyquinoline	−3.166	−3.078
6-Methoxyquinaldine	−2.247	−2.369
Benzene	−0.256	−0.5152
2-Thiophencarboxaldehyde	−1.685	−1.684
Anisole	−1.030	−0.7275
Aminopyrazine	−2.587	−2.758
Picolinic acid	−3.282	−3.396
6-Aminoquinoline	−3.061	−3.024
2-Naphthol	−2.477	−2.205
2-Methyltiophen	−0.426	−0.6193
Ethyl-2-methylbenzoat	−1.480	−1.429
Isophthalic acid	−3.987	−4.228
Methyl benzoate	−1.460	−1.312
1-Buthylbenzene	−0.753	−0.8914
Methyl paraben	−2.740	−2.669
3-Hydroxybenzoic acid	−3.309	−3.379
Phenylbutylamine	−1.397	−1.363
Methylbenzylamine	−1.180	−0.9269
2-Chlorolepidine	−2.300	−2.046
Indole	−1.846	−1.566
8-Nitroquinoline	−3.395	−2.855
3-Quinolinecarboxylic acid	−4.41	−3.997
3-Chloroaniline	−2.015	−1.634
Benzimidazole	−2.944	−3.288
6-Nitroquinoline	−3.615	−3.063
2-Hydroxy-4-methylquinoline	−3.876	−3.599

Table 4 (*Continued*)

Systematic name	Experimental log *J*	Predicted 12-5-1
Benzoic acid	−2.316	−2.349
1,5-Dimethyl-2-pyrrole carbonitrile	−1.791	−1.970
Furfuril amine	−1.116	−1.392
5-Nitroquinoline	−2.862	−3.233
4-*t*-Butyl toluene	−0.915	−1.108
1-Methyl-2-phenoxyethylamine	−1.63	−1.814
Phenethylamine	−1.257	−1.453
2-Amino-5-nitropyridine	−3.77	−3.851
4-Metoxy-2-quinolinic acid	−4.617	−4.364
1,3-Diethylbenzene	−0.774	−0.643
2,4-Dihydroxypyridine	−4.289	−4.491
1-Nitronaphtelene	−2.447	−2.818
8-Hydroxyquinaldine	−2.375	−2.794
4-Aminoacetophenone	−3.040	−2.795
Nitrobenzene	−1.556	−2.333
Benzaldehyde	−1.480	−1.294
Acetophenone	−1.640	−1.284
Ethylbenzene	−0.555	−0.5396
Fluorobenzene	−0.256	−0.5679
3-Chlorotoluene	−0.837	−0.8031
3-Xylene	−0.58	−0.5381
3-*t*-Butylphenol	−1.900	−1.767
4-Hydroxybenzoic acid	−3.530	−3.448
4-Chlorotoluene	−0.694	−0.8136
Butylbenzene	−0.895	−0.9544
Phenetole	−1.11	−0.9755
3-Anisaldehyde	−2.09	−1.716
Methyl-3-methylbenzoate	−1.43	−1.354
4-*t*-Butylbenzoic acid	−2.759	−2.776
Ethyl paraben	−2.69	−2.857
3-Pyridinecarboxaldehyde	−1.823	−2.02
3,5-Lutidine	−0.948	−0.9063
5-Chloro-3-pyridinol	−2.621	−2.377
4-*t*-Butylpyridine	−1.227	−1.225
Nicotinic acid	−3.76	−3.258
4-Picoline	−0.845	−0.820
3-Acetylpyridine	−1.992	−1.792
2-Aminopyridine	−2.682	−2.053
3-Aminopyridine	−1.895	−2.149
2-Chloro-6-methoxypyridine	−1.211	−1.188
2-Ethylpyridine	−0.718	−0.8758
2-Chloropyridine	−1.081	−1.060
2-Butoxypyridine	−1.155	−1.618
2-Fluoropyridine	−0.878	−0.9825
3-Methoxypyridine	−0.809	−1.182
2-Methoxy-5-nitropyridine	−2.653	−1.985

Table 4 (*Continued*)

Systematic name	Experimental log *J*	Predicted 12-5-1
2-Methoxy-5-aminopyridine	−2.23	−1.95
2-Hydroxy-5-nitropyridine	−3.747	−3.538
2-Hydroxypyridine	−2.499	−2.262
2-Amino-4-methylpyridine	−2.228	−2.083
2-Amino-S-chloropyridine	−2.625	−2.162
Ethyl nicotinate	−1.53	−1.82
Lepidine	−1.853	−1.834
6-Methylquinoline	−1.747	−1.813
8-Hydroxyquinoline	−2.358	−2.519
2-Methyl-8-nitroquinoline	−3.827	−3.923
Quinaldine	−1.622	−1.748
6-Isopropylquinoline	−1.897	−2.257
5-Aminoquinoline	−3.113	−3.152
3-Aminoquinoline	−2.934	−2.949
4-Hydroxyquinoline	−3.688	−2.904
8-Quinoline carboxylic acid	−4.213	−3.135
4-Quinoline carboxylic acid	−4.518	−4.270
1-Isoquinoline carboxylic acid	−4.132	−4.249
2-Methyl-5-butylpyridine	−1.113	−1.311
2.6-Dimetoxypyridine	−1.129	−1.045
6-Methoxy-8-nitroquinoline	−4.332	−4.249
2-Amino-4,6-dimethylpyridine	−2.253	−2.384
2-Methylindole	−1.983	−1.686
Naphthalene	−1.746	−1.360
1-Bromonaphthalene	−1.726	−1.761
1-Methylnaphthalene	−1.592	−1.483
2-Methoxynaphthalene	−1.918	−1.734
1,6-Dixydroxynaphthalene	−1.883	−2.320
2-Naphtalenacetic acid	−3.570	−4.163
1-Etoxynaphthalene	−2.790	−1.84
2-Methylbenzimidazole	−2.979	−3.246
2-Hydroxybenzimidazol	−3.922	−3.283
3-Phenyl-1-propylamine	−1.457	−1.581
1-Phenyl-2-propanol	−2.015	−1.811
3-Phenyl-1-propanol	−2.324	−1.825
3-Methyltiophene	−0.407	−0.6243
3-Thiopheneacetic acid	−2.411	−2.546
3-Thiophenecarboxaldehyde	−1.612	−1.473
3-Aminobenzoic acid	−3.727	−3.683
3-Toluic acid	−2.309	−2.273
3-Anisic acid	−2.579	−3.258

Table 4 (*Continued*)

Systematic name	Experimental log *J*	Predicted 12-5-1
4-Anisic acid	−3.226	−3.002
3-Chlorobenzoic acid	−2.371	−2.367
3-Nitrobenzoic acid	−2.753	−3.295
4-Aminobenzoic acid	−3.488	−3.821
4-Chlorobenzoic acid	−3.088	−2.388
4-Acetoxybenzoic acid	−3.107	−3.653
Benzylamine	−1.387	−1.213
Benzyl alcohol	−2.222	−1.852
4-Xylene	−0.457	−0.5351
1,3-Diisopropylbenzene	−1.060	−0.9969
Mesitylene	−0.701	−0.6647
1,3,5-Triethylbenzene	−1.083	−1.002
3-Fluoronitrobenzene	−1.620	−1.951
3-Methoxyaceto-phenone	−1.990	−1.586
4-Anisaldehyde	−2.070	−1.737
4-Isopropylbanzalde-hyde	−1.640	−1.497
Methyl-4-*t*-butylbenz-oate	−1.710	−1.86
Dibenzyl	−1.980	−1.918
3-Phenoxytoluene	−2.010	−2.056
2-Aminoacetophenone	−2.160	−2.171
2-Anisaldehyde	−2.030	−1.529
2-Chloroacetophenone	−1.830	−1.316
2-Chlorobenzaldehyde	−1.580	−1.24
2-Chloronitrobenzene	−1.540	−2.117
2-Chlorotoluene	−0.771	−0.7855
Ethyl salicylate	−1.610	−2.281
2-Fluoroaniline	−1.310	−1.437
2-Fluorobenzaldehyde	−1.300	−1.526
2-Fluoronitrobenzene	−1.84	−1.837
2-Fluoropropiophenone	−1.44	−1.546
2-Fluorotoluene	−0.349	−0.6874
2-Hydroxyacetophenone	−1.78	−1.983
2-Isopropylaniline	−1.69	−1.516
2-Metoxyacetophenone	−2.02	−1.786
Methyl 2-nitrobenzoate	−2.68	−2.56
Methyl 2-metoxybenzoate	−2.19	−1.709
Methyl salicylate	−1.67	−2.133
2-Nitrotoluene	−1.72	−2.038
2-Xylene	−0.644	−0.5438
2-Nitrobenzoic acid	−2.86	−3.08
Salicylic acid	−2.57	−3.06
4-Hydroxybenzamide	−3.83	−4.228
3-Hydroxy-4-metoxy-benzoic acid	−4.37	−4.042
4-Chloro-3-nitroaceto-phenone	−3.33	−3.047

Table 4 (*Continued*)

Systematic name	Experimental log *J*	Predicted 12-5-1
1,2,4-Trimethylbenzene	−0.74	−0.5734
Phenylurea	−3.31	−3.061
Benzohydroxamic acid	−3.27	−3.445
Benzamide	−3.07	−3.412
Ethyl cinnamate	−1.95	−1.905
Phenyl acetate	−1.65	−1.647
Benzonitrile	−1.550	−1.252
Thioanisole	−1.390	−1.142
Iodobenzene	−1.300	−1.548
Styrene	−0.711	−0.5906
2-Chlorophenoxyacetic acid	−2.930	−3.209
2-(3-Hydroxyp henyloxy) ethanol	−3.540	−5.153
3-Methoxybenzyl acetate	−2.130	−2.787
Phenoxyacetic acid	−2.458	−3.135
3-Phenylbutyraldehyde	−1.959	−1.673
2-Phenylpropion-aldehyde	−1.686	−1.468
Propyl paraben	−2.720	−3.01
3-Chloro-4-methylan-iline	−1.960	−1.667
3-Amino-1,2,4-triazole	−3.270	−4.158
2-Pyrazine carboxylic acid	−4.067	−4.17
3-Amino-5,6-dimethyl-1,2,4-triazine	−3.865	−4.697
Anthracene	−3.839	−3.578
Acridine	−2.683	−3.696
2-Quinoxalinol	−4.164	−3.69
2,4-Dimethyl-6-hydrox-yp yrimidine	−3.3	−3.399
4-Methylpyrimidine	−1.022	−1.278
Isoquinoline	−1.677	−1.69
Methoxymethylphenyl sulphide	−1.684	−1.646
3-Iodoanisole	−1.805	−1.319
2-Chloroanisole	−1.761	−1.161
4-Bromoveratrole	−2.34	−1.372
4-Bromotoluene	−1.421	−1.328
2-Anisidine	−2.023	−1.558
3-Fluorobenzyl chloride	−1.12	−1.071
2-Chloro-4-fluoroaceto-ph enone	−1.937	−2.141
4-Chloro-4-fluorobutir-ophenone	−2.21	−2.845
2-Fluorobenzoic acid	−2.29	−2.656
5-Methylbenzimidazole	−3.076	−3.579

dence surface. This gives an indication of how the network output alters in response to the two selected input variables. Fig. 1 displays functional dependence surfaces of the six most important descriptors. Nonlinearity of inputs is clearly evident suggesting the complex relationship between the input descriptors and flux value.

As expected, the model shows a strong correlation (up to $R^2 = 0.919$) between predicted and experimentally measured flux values (Table 4). Since the slope ($b = 1.03$, $t_b = 1.64$) was not significantly different from unity, the method did not show proportional error. In other words, the sensitivity was the same for measured and predicted values. A proportional error leads to a change in b so that the difference between b and unity gives an estimate of the proportional error. However, the intercept was significantly different from zero at the 0.05 significance level ($a = 0.13$, $t_a = 2.8$) indicating that there is a constant systematic error that leads to the method bias. Indeed, the predicted values are slightly lower that the experimentally measured flux values.

4. Conclusion

A 12-descriptor nonlinear computational neural network model has been developed for the estimation of log J values for a data set of 254 drugs. The training set RMS error was 0.36 and the testing set RMS error was 0.59. Based on the RMS errors of the training and testing sets and high correlation of predicted vs. experimentally derived M/P values (R^2 greater than 0.91, it is clear that a link exists between structure and membrane penetration. The strength of the link was measured by the quality of the external prediction set. With an RMS error of 0.60 and a good visual plot, the external prediction set ensures the quality of the model. Unlike previously reported models, the QSPR model described here does not require experimental parameters and could potentially provide useful prediction of membrane penetration of new drugs and reduce the need for actual compound synthesis and flux measurements.

References

[1] R.C. Wester, H.J. Maibach, Cutaneous pharmacokinetics: ten steps to percutaneous absorption, Drug Metab. Rev. 14 (1983) 169–172.
[2] R.H. Guy, Transdermal drug delivery and cutaneous metabolism, Xenobiotica 17 (1987) 325–343.
[3] L. Brown, R. Langer, Transdermal delivery of drugs, Ann. Rev. Med. 39 (1988) 221–229.
[4] C.A. Howard, Pharmaceutical Dosage Forms and Drug Delivery Systems, Williams & Wilkins, Malvern, PA, 1995.
[5] P.M Elias, E.R. Cooper, A. Korc, B.E. Brown, Percutaneous transport in relation to stratum corneum structure and lipid composition, J. Invest. Dermatol. 76 (1981) 297–301.
[6] P Politzer, J.S. Murray, Quantitative Treatments of Solute/Solvent Interaction, Elsevier, New York, 1994.
[7] R.O. Potts, R.H. Guy, Predicting skin permeability, Pharm. Res. 9 (1992) 663–669.
[8] R.H. Guy, R.O. Potts, Structure-permeability relationships in percutaneous penetration, J. Pharm. Sci. 81 (1992) 603–604.
[9] V.P. Shah, G.L. Flynn, R.H. Guy, H.I. Maibach, H. Schaefer, J.P. Skelly, et al., Workshop report on in vivo percutaneous penetration/absorption, Washington D.C, Rev. Skin. Pharmacol. 4 (1991) 220–228.
[10] R.O. Potts, R.H. Guy, A predictive algorithm for skin permeability: the effects of molecular size and hydrogen bond activity, Pharm. Res. 12 (1995) 1628–1633.
[11] E.J. Lien, H. Gao, QSAR analysis of skin permeability of various drugs in man as compared to in vivo and in vitro studies in rodents, Pharm. Res. 12 (1995) 583–587.
[12] W.J. Pugh, I.T. Degim, J. Hadgraft, Epidermal permeability-penetrant structure relationships: 4, QSAR of permeant diffusion across human stratum corneum in terms of molecular weight, H-bonding and electronic charge, Int. J. Pharm. 197 (2000) 203–211.
[13] Y. Chen, W.L. Yang, L.E. Matheson, Prediction of flux through polydimethylsiloxane membranes using atomic charge calculations, Int. J. Pharm. 94 (1993) 81–88.
[14] Y. Chen, P. Vayuhasuwan, L.E. Matheson, Prediction of flux through polydimethylsiloxane membranes using atomic charge calculations: application to an extended data set, Int. J. Pharm. 137 (1996) 149–158.
[15] M.T.D. Cronin, J.C. Dearden, R. Gupta, G.P. Moss, An investigation of flux across polydimethylsiloxane membranes by use of quantitative structure-permeability relationship, J. Pharm. Pharmacol. 50 (1998) 143–152.
[16] J Chen, W.J. Peijnenburg, X. Quan, S. Chen, Y. Zhao, F. Yang, The use of PLS algorithms and quantum chemical parameters derived from PM3 hamiltonian in QSPR studies on direct photolysis quantum yields of substituted aromatic halides, Chemosphere 40 (2000) 1319–1326.
[17] P Constans, J.D. Hirst, Nonparametric regression applied to quantitative structure-activity relationships, J. Chem. Inf. Comput. Sci. 40 (2000) 452–459.

[18] S.W. Sarle, Neural Networks and Statistical Models. Proceedings of the Nineteenth Annual SAS Users Group International Conference, 1994.
[19] T. Aoyama, Y. Suzuki, H. Ichikawa, Neural networks applied to structure-activity relationships, J. Med. Chem. 33 (1990) 905–908.
[20] S. So, W.G. Richards, Application of neural networks: quantitative structure-activity relationships of the derivatives of 2,4-diamino-5- (sub stituted-benzyl)pyrimidines as DHFR inhibitors, J. Med. Chem. 35 (1992) 3201–3207.
[21] D.T. Manallack, D.D. Ellis, D.J. Livingstone, Analysis of linear and nonlinear QSAR data using neural networks, J. Med. Chem. 34 (1994) 3758–3767.
[22] J. Zupan, J. Gasteiger, Neural networks: a new method for solving chemical problems or just a passing phase?, Analyt. Chim. Acta. 248 (1992) 1–30.
[23] H. Lohninger, Evaluation of neural networks based on radial basis functions and their application to the prediction of boiling points from structural parameters, J. Chem. Inf. Comput. Sci. 3 (1993) 736–744.
[24] H. White, Artificial Neural Networks: Approximation and Learning Theory, Blackwell, Oxford, 1992.
[25] N. el Tayar, R.S. Tsai, B. Testa, P.A. Carrupt, C. Hansch, A. Leo, Percutaneous penetration of drugs: a quantitative structure-permeability relationship study, J. Pharm. Sci. 80 (1991) 744–749.
[26] G. Grassy, B. Calas, A. Yasri, R. Lahana, J. Woo, S. Iyer, et al., Computer-assisted rational design of immunosuppressive compounds, Nat. Biotechnol. 16 (1998) 748–752.
[27] D. Gorse, A. Rees, M. Kaczorek, R. Lahana, Molecular diversity and its analysis, Drug Discov. Today 4 (1999) 257–264.
[28] L.H. Hall, L.B. Kier, The molecular connectivity χ indices and κ shape indices in structure-property modeling, in: D.B. Boyd, K. Lipkowitz (Eds.), Reviews of Computational Chemistry, vol. 2, Indiana University-Purdue University at Indianapolis (IUPU), 1991, p. 367–422.
[29] V.K. Gombar, D.V.S. Jam, Quantification of molecular shape and its correlation with physico-chemical properties, Indian J. Chem. 26A (1987) 554–555.
[30] A. Verloop, W. Hoogenstraaten, J. Tysker, in: E.J. Ariens (Ed.), Drug Design, vol. 7, Academic Press, New York, 1976.
[31] M.A. Farnum, C.F. Zukoski, The effect of the dipole moment on protein-protein interactions, Biophys. J. 76 (1999) A127–A127.
[32] S.O. Shan, D. Herschlag, The change in hydrogen bond strength accompanying charge rearrangement: implications for enzymatic catalysis, Proc. Natl. Acad. Sci. USA 93 (1996) 14474–14479.
[33] J Aihara, Reduced HOMO-LUMO gap as an index of kinetic stability for polycyclic aromatic hydrocarbons, J. Phys. Chem. 103 (1999) 7487–7495.
[34] S. Trohalaki, R. Pachter, The utility of electron affinity as a QSAR descriptor in predictive toxicology, Abstr. Pap. Am. Chem. S. 217 (1999) U674–U674.
[35] M.J.S. Dewar, E.G. Zobisch, E.F. Healy, J.J.P. Steart, AM1 — a new general-purpose quantum-mechanical molecular mode, J. Am. Chem. Soc. 115 (1993) 5348–5348.
[36] P.O. Lowdin, M.C. Zerner, The role of quantum-chemistry as a foundation for the principles of chemistry — Introduction, Int. J. Quantum Chem. 49 (1994) 129–130.
[37] N.Y. Ohrn, J.R. Sabin, M.C. Zerner, Proceedings of the international symposium on the application of fundamental theory to problems of biology and pharmacology — Introduction, Int. J Quantum Chem. 20(Suppl.) (1993) R5–R6.