ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
25 (2001) 227–237

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

www.elsevier.com/locate/jpba

Theoretically-derived molecular descriptors important in human intestinal absorption

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

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

Received 20 June 2000; received in revised form 24 August 2000; accepted 1 October 2000

Abstract

A quantitative structure–human intestinal absorption relationship was developed using artificial neural network (ANN) modeling. A set of 86 drug compounds and their experimentally-derived intestinal absorption values used in this study was gathered from the literature and a total of 57 global molecular descriptors, including constitutional, topological, chemical, geometrical and quantum chemical descriptors, calculated for each compound. A supervised network with radial basis transfer function was used to correlate calculated molecular descriptors with experimentally-derived measures of human intestinal absorption. A genetic algorithm was then used to select important molecular descriptors. Intestinal absorption values (IA%) were used as the ANN's output and calculated molecular descriptors as the inputs. The best genetic neural network (GNN) model with 15 input descriptors was chosen, and the significance of the selected descriptors for intestinal absorption examined. Results obtained with the model that was developed indicate that lipophilicity, conformational stability and inter-molecular interactions (polarity, and hydrogen bonding) have the largest impact on intestinal absorption.

Keywords: Human intestinal absorption; Molecular descriptors; ANN; GNN

1. Introduction

The oral route is the one most preferred for systemic administration of drugs so it is important that any new drug has good bioavailability.

Absorption from the gastrointestinal tract, as well as penetration of other membrane barriers may be passive or active. Passive transport is governed by physico-chemical properties whereas active transport involves specific binding of a molecule to a binding site on a transport protein. Attempts have been made to explain and predict drug absorption from a variety of physico-chemical parameters [1–5]. Drugs must dissolve in the gastrointestinal tract to be available to cross the intestinal membrane. Dissolution is effected by the aqueous solubility, ionizability (pK_a) and lipophilicity (octanol/water log P). Although there are specialized membrane transport mechanisms for natural cell substances and small polar

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

0731-7085/01/$ - see front matter
PII: S0731-7085(00)00492-1

molecules may simply diffuse across, most organic compounds penetrate tissue cells as though the boundaries were lipid in nature. Single physico-chemical properties of drug molecules such as lipophilicity can be roughly correlated to passive drug transport across cell membranes, but only for homologous series of compounds. Log P is thus a crucial factor governing passive membrane partitioning; an increase in log P enhances permeability while reducing solubility.

There are 32 known human transporter families, including the mitochondrial transporter family (14 members), GABA transporter family (13 members), amino acid permease family (eight members), and the organic anion/cation transporter family (eight members) [6]. The intestinal oligopeptide transporter (cloned as Pept-1) plays a major role in protein synthesis and drug therapy [7]. Some transporter proteins (such as Pep T1) pump solute into the cell, whereas others (such as P-gp) work in the other direction. Studies of the relationships between the chemical structure of drugs (steric and electrostatic fields) and their affinity for the small intestinal oligopeptide carriers have shown that carrier permeability is sensitive to composition, size and hydrophobicity of the ligands [8].

In vivo animal studies and human ex vivo intestinal absorption models have been used to predict drug absorption [9]. These methods are expensive, time consuming, and require a large amount of sample. Moreover, these models produce variable results for structurally different compounds. Drug permeability in cell cultures or intestinal tissues has been shown to be a useful predictor of drug absorption in vivo even when drugs of different structural classes are studied [10]. The most useful are CaCO-2 cells derived from a human colon carcinoma cell line. Permeability measurements are based on the rate of appearance of the test compound in the receiver compartment. The apical surface of the monolayer retains many characteristics of the intestinal brush border and expresses functional transport proteins [11] and metabolic enzymes [12,13]. Although tissue and particular cell culture models are better predictors of drug absorption than single physico-chemical parameters, they are also labour intensive and require actual compound synthesis and absorption measurements. Thus, a theoretical method that could predict absorption with high precision would be of interest.

A goal in the design of this study was the development of a Genetic Neural Network (GNN) model to predict the degree of drug absorption from the gastrointestinal tract, depending on the calculated molecular descriptors. The molecular structure of any molecule determines its function. 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 biological activity has resulted in the development of quantitative structure property relationships (QSPR). While some molecular descriptors can be determined experimentally, using computational methods to derive them is much faster and more convenient. Such relationships, once quantified, can be used to estimate the properties of new drugs or hypothetical drugs only from the sketch of their structure. The GNN model uses artificial neural network (ANN) to correlate activity with descriptors that are preselected by a genetic algorithm (GA).

Recent investigations involving computational neural networks and genetic algorithms serve as examples of the application of the QSPR methods. Three-layer, feed-forward neural networks have provided excellent results in several QSPR studies. The genetic algorithm has been shown to be very effective in performing descriptor selection. Wessel et al. [14] have developed a neural network model consisting of six input neurons (descriptors), four hidden neurons and one output neuron (% of intestinal absorption). The six descriptors were the NSB-number of single bonds, SHDW-6-normalized 2D projection of the molecule on YZ plane, CHDH-1-charge on donatable hydrogen atoms, SAAA-2-surface area of hydrogen bond acceptor, SCAA-2-surface area $\times$ charge of hydrogen bond acceptor and GRAV-3-cube root of gravitation index. This study is also focused on the prediction of human intestinal drug absorption from molecular structure, but it uses radial basis function (RBF) networks and different molecular descriptors to build a model.

1.1. Artificial neural network (ANN)

An artificial neural network is a biologically inspired computer program designed to simulate the way in which the human brain processes information. ANNs are constituted from hundreds of single processing elements (PE), artificial neurons. Each PE has weighted inputs, transfer function and one output. PEs are connected with coefficients (weights) which constitute the neural structure and are organised in layers, the input layer, the output layer, and the hidden layers between them. 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. Units in the input layer do not process. They simply pass an output value onto units in the second layer. Each hidden or output unit has a number of incoming connections from units in the preceding layer. The weighted sum of the inputs comprises the activation of the neuron. The activation signal is passed through an activation function (also known as a transfer function) to produce a single output of the neuron. Thus what is learned in a hidden neurone is based on all the inputs taken together. This hidden layer is the place where the network learns interdependencies in the model. Hidden neurons communicate only with other neurons. They are part of the large internal pattern that determines a solution to the problem. PE is essentially an equation which balance inputs and outputs. Transfer functions for the hidden units are needed to introduce nonlinearity into the network.

Multilayer perceptrons (MLPs) and radial basis function (RBF) networks are the two most commonly used types of feed-forward network. The only difference is the way in which hidden units combine values coming from preceding layers in the network. An MLP models the response function using the composition of sigmoid functions. A radial basis function network (RBF) has a hidden layer of radial units, each modeling a Gaussian response surface. An MLP unit responds (non-linearly) to the distance of points from the line of the sigmoid projection; in a radial basis function network units respond (non-linearly) to the distance of points from the centre represented by the radial unit. The response surface of a single radial unit is a Gaussian (bell-shaped) function, peaked at the centre, and descending outwards.

MLP units are defined by their weights and threshold, which together give the equation of the defining line. In contrast, a radial unit is defined by its centre point and radius. Since these functions are non-linear, it is not necessary to have more than one hidden layer to model any shape of function.

2. Experimental

Neural Networks TM (StatSoft®) was used for building the IA model. For calculating drug properties from molecular structure Pallas 2.1 (Compu Drug Int.), ChemSketch 3.5 (ACD Inc.) and CAChe Project leader Version 3.11 (Oxford Molecular Ltd.) were used.

The set of 86 structurally different compounds and their experimentally-derived intestinal absorption values (IA%) used in this study were collected from the literature [14]. The absorption values for 86 compounds were used as the ANN's output and 57 calculated molecular descriptors as the inputs (Table 1). A set of 76 compounds was selected for training and testing the ANNs and 10 compounds were used as an external prediction set. Before each training run the 76 compounds in the working data set were split randomly into: the training set containing 67 data sets and nine testing (cross validation) sets. Connections or units were eliminated during training based on sensitivity reports and overall quality of a particular subset of descriptors.

A supervised network with back-propagation learning rule and radial basis transfer function was used.

3. Results and discussion

3.1. Descriptor generation and analysis

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 utilised that account for three aspects of the compounds concerned, namely their hydrophobic, electronic and steric effects. Hydrophobicity is related to the ability of a compound to partition through a membrane. Steric effects relate to their ability to pass through a membrane or to bind to a receptor site. Electronic effects may relate to reactivity or metabolism. A total of 57 calculated structure features including constitutional, topological, chemical, geometrical and quantum chemical descriptors were calculated for each of the 60 compounds (Table 1). Connectivity and topological descriptors were calculated directly from the connection table representation of the structure and employ 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. ChemSketch was used to calculate the composition, polarizability, molecular weight, index of refraction and surface tension. Quantum chemical and geometrical descriptors were calculated using CAChe. Once the molecular structures were encoded, the genetic algorithm (GA) input selection was used to select the subset of descriptors that best encoded the intestinal absorption.

3.2. Neural network analysis

The two forms of network analysis are cross validation and sensitivity analysis. Both methods were used during the training of the network. The cross validation test set was used to determine the level of generalization produced by the training set. Training was stopped at each run when the error in the test set began to rise.

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 non-linear transformations of linear combinations of the inputs. Sensitivity reports show the sensitivity of the output variables, as a percentage, to 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 most significant inputs and the effect that a change in a particular input has on output. In order to reduce the amount of input data and select the best ANN, architecture pruning [15] was applied, this being similar to backward elimination in stepwise regression. The sensitivity analysis feature was used to discard the insignificant portions of data and to focus the training on that which was most impor-

Table 1
Calculated molecular descriptors

Class	Molecular descriptor
Constitutional descriptors	Chemical composition (weight percent of C, H, O, N, S, Cl, F in molecular mass), Atom count (C, H, N, S, Cl, F, S, O), Functional group counts (amine, amide, carbonyl, carboxyl, ether, hydroxyl, methyl, methylene, nitro, sulfide and sulfone)
Topological descriptors	Kier and Hall connectivity indices (Chi0–Chi2) and Valence connectivity indices (Chi0V–Chi2V), Topological shape indices (Kappa0-2)
Chemical descriptors	Molar refractivity, pK_a, pK_a^0, log P, log D, Molecular mass, Parachor, Surface tension, Polarizability, Density, Dielectric constant
Geometrical descriptors	Solvent accessible surface area (SASA), Molar volume
Quantum chemical descriptors	Dipole moment, HOMO and LUMO energies, Steric energy, Heat of formation, Total energy, Minimum energy, Electron affinity

Table 2
Relative sensitivity of IA% on selected molecular descriptors

Descriptor	Average sensitivity
log *P*	+0.13735
Electron affinity	+0.10797
Density	−0.09419
Hydrogen atom count	−0.08855
Steric energy	+0.08448
–OH	−0.08287
HOMO	−0.06539
LUMO	−0.05903
Minimum energy	+0.05777
–COOH	−0.05390
Dielectric constant	+0.04591
$–CH_3$	+0.03869
–CO	+0.03791
Dipole	−0.02931
Parachor	−0.01524

tant. Inputs that produced low sensitivity values were considered insignificant and were removed from the network. This also reduced the size and complexity of the network and thus training time, and improved the network performance.

Initially, a neural network consisting of 62 inputs (molecular descriptors), one hidden layer and one output neuron (target, IA (%)) was used. The number of inputs and hidden neurones were optimized. Connections or units were eliminated during training based on sensitivity report. Following a sensitivity analysis, the number of inputs was reduced to 57, 48, 36, 28, 23, 19, 15, 13, 10 and finally to eight inputs with sensitivity greater than 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5 and 4%, respectively.

The training set was used to train the network and the testing set used to monitor over-training. The testing and training set RMS errors were used to evaluate the overall quality of a particular subset of descriptors and network topology (Table 2).

A 15-descriptor, nonlinear GNN model has been developed for the estimation of IA% values for a data set of 85 drugs. The model has 15 selected inputs with sensitivity greater that 3%, one hidden layer with four neurons and percent of intestinal absorption as the output neuron, thus forming a 15-*8*-1 architecture. Other architectures were examined, but they produced poorer quality neural network models (Table 3). The training and testing RMS errors are 0.59 and 0.90, respectively. The strength of the correlation between selected descriptors and gastrointestinal absorption is measured by the quality of the external prediction set. With the RMS error of 0.42 and high correlation between predicted and experimental values ($R^2 = 0.802$), the external prediction set ensures the quality of the model (Fig. 1).

The relative importance of selected descriptors is evaluated by sensitivity analysis and is shown in Table 2. Classic gastrointestinal absorption studies suggested that molecules pass through a mem-

Table 3
The influence of different topology on the ANN's performance[a]

ANNs[b]	*N*	RMStr	RMStest	ERR(%)	S.D.	IS
57-*20*-1	57	0.130	0.170	0.188	27.98	0
48-*18*-1	48	0.112	0.190	0.172	27.24	0.005
36-*14*-1	36	0.105	0.182	0.211	33.63	0.01
28-*12*-1	28	0.105	0.164	0.254	34.26	0.015
23-*10*-1	23	0.105	0.138	0.189	27.93	0.02
19-*9*-1	19	0.100	0.152	0.205	27.72	0.025
15-*8*-1	**15**	**0.105**	**0.141**	**0.142**	**25.01**	**0.03**
13-*3*-1	13	0.110	0.155	0.142	25.01	0.035
10-*2*-1	10	0.105	0.155	0.373	37.45	0.04
8-*4*-1[c]	8	0.094	0.154	0.344	50.24	

[a] *N*, number of inputs; IS, input sensitivity; ERR(%) = [(predicted − actual)/actual)] × 100.
[b] Number of inputs–*hidden neurons*–outputs.
[c] ANN model developed by Wessel et al. [14].

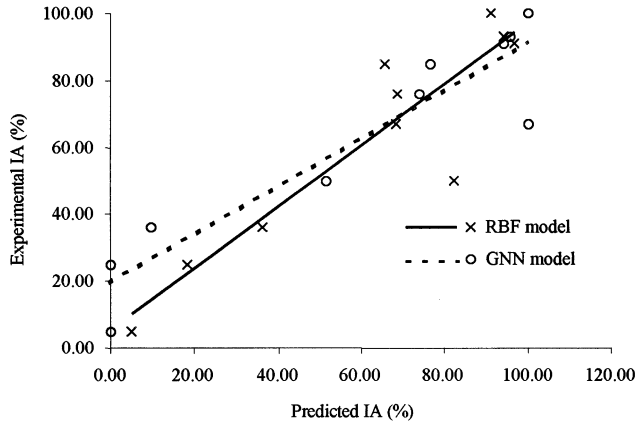

Fig. 1. Predicted IA(%) values for the validation data set obtained with developed RBF model (15/8/1) and GNN model [14] versus experimentally derived IA(%) values. RBF: experimental IA(%) = 5.5 + 0.837 × predicted(%); GNN: experimental IA(%) = 19.5 + 0.719 × predicted(%).

brane only as uncharged species [16,17]. However, most biologically active molecules are at least partially ionized at biological pH values and absorption could be influenced by the membrane penetration of the neutral form and by the receptor binding of the ionized form. Hence, ionization equilibrium may influence the total concentration of drug available for passive diffusion or binding to the receptor and active transport. However, pH progressively increases throughout the length of the gastrointestinal tract and solubility and thus absorbability of an ionizable compound will vary accordingly. These considerations suggest reasons for the poor quantitative relationship between pK_a and bioavailability [18]. The most important chemical descriptors were log P, density, parachor and dipole moment. The octanol–water partition coefficient (log P) is frequently used in quantitative structure–activity relationship studies as a measure of the lipophilic character of the molecules. Among physico-chemical determinants of drugs, lipophilicity is approximately correlated to passive transport across cell membranes and the ability of a compound to partition through a membrane [19]. Calculated log P can be roughly correlated with drug absorption, but only for homologous series of compounds as it does not account for intramolecular interactions. For example, intramolecular hydrogen bonding [20] can dramatically influence absorption properties.

The density of a substance is the ratio of its mass to its volume. Molecular size is limiting on the absorption through membranes in general. Molecular mass is often used as molecular size descriptor. Molar density estimates diffusion coefficients for hydrocarbon systems [21] and has effects on the enzyme targeting capability and cell binding properties [22]. Compounds with high molar density are not well absorbed from the gastrointestinal tract due to increase in molecular mass.

Parachor [23] is an additive physical property of a substance related to its molar volume, and is determined by the kind and the number of atoms in a molecule as well as their manner of arrangement and binding. It is essentially a molecular volume with the fourth root of the surface tension as a correcting factor. It is evident, from the sensitivity report, that an increase in the parachor decreases absorption. Small lipid insoluble substances penetrate cell membranes via the pores between aqueous phases on both sides of the membrane. The rate of such passive diffusion depends on the size of the pores, the molecular volume of the solute and the solute concentration gradient.

Dipole moment is the measure of polarity of the molecule and intramolecular electronic effect. Electronic effects may relate to molecular reactivity or metabolism [24]. The greater the dipole

moment and the smaller the steric energy of a molecules, the greater its activity. Drugs with high dipole moments are less well absorbed.

Quantum chemical descriptors can be used to establish the conformational stability, chemical reactivity and inter-molecular interactions. They include thermodynamic properties (system energies) and electronic properties (LUMO or HOMO energy). Energy was calculated for an optimized conformation with the most stable geometry (minimum energy) using molecular or quantum mechanics to determine bond strengths, atomic hybridizations, partial charges, and orbitals from the positions of the atoms and the net charge. The Lowest Unoccupied Molecular Orbital (LUMO) energy level is a property of electronic structure and represents the electron affinity of a molecule or its reactivity as an electrophile. The energy difference between the Highest Occupied Molecular Orbital (HOMO) and the LUMO energies is related to the minimum energy needed to excite an electron in the molecule. Increases in electron acceptor properties will increase binding affinity. Good nucleophiles are those where the electrons reside in high-lying orbitals and good electrophiles are those where the LUMO is low-lying.

Electron affinity [25] incorporates electron correlation and relaxation, whereas LUMO does not, and is also a measure of the reduction in activity. Electrophiles are often reducing agents. Since living organisms function at an optimum redox potential range, it is assumed that redox potential of compounds of certain type correlate with biological effect. 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 only for compounds with similar structure and physical properties. Compounds with lower LUMO and HOMO energy (good electrophiles) and higher electron affinity have a higher rate of absorption.

Charge-transfer and hydrophobic interactions of the lateral substituents have indicated the importance of steric parameters within nonplanar congeners. For a given molecule, the atoms will adjust their positions by stretching and bending bonds away from standard values so as to produce a minimum energy configuration. Deviation from those standard values results in steric energy that influences the relative stability and thus reactivity of molecules. Higher steric energy and minimum energy were found to promote absorption. The increase in activity roughly parallels the decrease in water solubility and the increase in lipid solubility (log P) that may be associated with the availability of the compound for penetration through the membrane barrier.

Dielectric setting, i.e. the change in charge rearrangement of the molecule accompanies a change in hydrogen bond strength [26]. The energy of the molecular interaction is affected by its dielectric constant and the RBF model shows that an increase in dielectric constant promotes intestinal absorption.

The most important constitutional descriptors were found to be the methyl, hydroxyl, carbonyl and carboxyl group count. Functional groups exhibit a characteristic reactivity and chemical behavior when present in a compound. Certain functional groups are important in the specific interaction between a drug and transporter receptor. Undoubtedly, the functional group accounts for many of the dipole–dipole, dipole-induced dipole and hydrogen bond interactions of any molecule. 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. Properties associated with hydrogen bonding should be kept to a minimum to promote high absorption. Since the presence of hydroxyl groups facilitates hydrogen bonding, an increase in the number of hydroxyl groups on a molecule will decrease its absorption. High charge-transfer properties (dipole, carbonyl group) hinder absorption. Dipole–dipole interactions are related to the dipole moment of a whole molecule or a part of a molecule, such as a functional group, e.g. carbonyl. In many cases, experimental work and structure/activity studies have suggested that the charged group is essential for biological activity. An increase in the number of carboxyl groups decreases intestinal absorption, suggesting the for-

mation of non-absorbable complexes or hydrogen bonding interaction. The methyl group is an electron donating group that enhances non-polar solubility. The presence of methyl groups on the terminal carbon atom of a conjugated chain stabilizes the positive charge of the main chain. Methyl substitutions have substantial effects on proton affinity and increase intestinal absorption.

The regression coefficients of the IA(%) values predicted with the RBF model versus experimentally-derived values were then compared with those obtained from the GNN model [14] (Table 4). Since nine data sets were selected randomly as the test set out of 76 data sets and 10 were used as the external validation set at the beginning of each neural net analysis, the values of r^2 vary with each run depending on the components included in the training set. Each run was repeated five times to find the best value of r^2.

A strong correlation between the predicted versus experimentally-derived IA(%) values (up to $R^2 = 0.857$) for the validation data set was obtained with the model developed by Wessel et al. [14]. However, the slope was significantly different from unity, indicating proportional error, and the intercept was different from zero. A proportional systematic error leads to a change in a slope, so that the difference between slope and unity gives an estimate of the proportional error. A constant systematic error shows up in a value of the intercept different from zero. As expected, we did not obtain the higher correlation ($R^2 = 0.802$) for the best RBF model (15/8/1). However, the slope was not significantly different from unity ($t = 0.49$) and the intercept was not significantly different from zero ($t = 1.11$), indicating that the method does not show proportional error or method bias (Fig. 1).

4. Conclusion

A 15-descriptor (Table 3) nonlinear computational neural network model has been developed for the estimation of HIA% values for a data set of 86 drugs. The training set RMS error was 0.590 and the testing set RMS error was 0.900. Based on the RMS errors of the training and testing sets, it is clear that a link between structure and IA(%) does exist. The strength of the link is measured by the quality of the external prediction set. With the RMS error of 0.425 and a good visual plot, the external prediction set ensures the quality of the model. The QSPR model described here does not require experimental parameters and could potentially provide useful prediction of gastrointestinal absorption of new drugs. However, the data set employed here is highly biased towards well absorbed compounds and contains compounds that are subjected to various active transport mechanisms. A number of transport proteins are responsible for solute transport in vivo. Although not all of these are present in the intestine, it is likely that a model based on 76 compounds (the training and testing sets) would be sufficient to describe transport by all possible routes. Furthermore, uncritical use of data sets compiled from the literature is associated with the risk of possibly erroneous values.

Until recently, drug properties could be deduced only empirically. Hence developing a new synthesis or improving a drug product required a lot of experimental work and a fair amount of trial and error. The QSPR model described here does not require experimental work and could potentially provide useful predictions of the rate and amount of gastrointestinal absorption of new drugs.

We expect that this research will suggest new therapeutic strategies for efficient oral drug delivery. Since the cost of drug development is many times larger than the cost of drug discovery, predictive methodologies that might aid the selection of orally bioavailable drug candidates could be very valuable. This approach can be applied to estimate the bioavailability of purely hypothetical compounds as there would be no need for experimental determinations or synthesis. Quantum chemistry and molecular modeling techniques provide insight into which aspects of a molecular structure influence intestinal absorption. Such insight can facilitate the systematic approach to the design of new molecules with more desirable properties. It also has the advantage that it can be applied to new chemical entities before they are even developed.

Table 4
Predicted IA(%) values obtained with developed RBF model (15/8/1) and GNN model [14] versus experimentally-derived IA(%) values

	IA(%)	RBF	Recovery	GNN	Recovery
Training and testing data sets					
Gentamycin	0.00	14.82	*	0.00	*
Cromolyn	0.50	0.00	0.00	0.00	0.00
Olsalazin	2.30	0.00	0.00	4.82	209.57
Ganciclovir	3.80	4.53	119.24	6.21	163.42
Cefuroxime	5.00	12.90	257.98	0.00	0.00
Chlorothiazide	13.00	24.69	189.91	20.99	161.46
Mannitol	15.00	17.71	118.08	16.05	107.00
Nadolol	34.50	78.73	228.19	36.00	104.35
Norfloxacin	35.00	82.87	236.76	53.63	153.23
PenicillinV	45.00	74.76	166.14	43.89	97.53
Etoposide	50.00	50.00	100.00	51.65	103.30
Atenolol	50.00	70.63	141.25	79.62	159.24
Ziprasidone	60.00	100.00	166.67	70.32	117.20
Sulfasalazine	65.00	89.68	137.98	65.90	101.38
Hydrochlorothiazide	67.00	69.67	103.98	62.08	92.66
Sumatriptan	75.00	89.15	118.87	74.76	99.68
Guanabenz	75.00	100.00	133.33	85.03	113.37
Propylthiouracil	75.00	91.78	122.37	100.00	133.33
Quinidine	80.00	100.00	125.00	95.07	118.84
Acetaminophen	80.00	96.66	120.83	100.00	125.00
Methylprednisolone	82.00	96.52	117.71	95.86	116.90
Sorivudine	82.00	80.77	98.50	99.56	121.41
Bupropion	87.00	87.00	100.00	83.05	95.46
Trovafloxacin	88.00	100.00	113.64	95.12	108.09
Acrivastine	88.00	98.79	112.26	100.00	113.64
Acebutolol	89.50	74.17	82.87	93.02	103.93
Timolol maleate	90.00	80.99	89.98	77.43	86.03
Phenytoin	90.00	99.68	110.75	91.79	101.99
Betaxolol	90.00	77.38	85.98	95.10	105.67
Oxprenolol	90.00	84.27	93.64	95.35	105.94
Scopolamine	90.00	100.00	111.11	95.58	106.20
Propranolol	90.00	99.30	110.33	95.77	106.41
Tenidap	90.00	100.00	111.11	96.78	107.53
Chloramphenicol	90.00	97.26	108.07	100.00	111.11
Terazosin	91.00	92.03	101.13	93.87	103.15
Hydrocortizone	91.00	85.31	93.75	96.30	105.82
Amoxicillin	93.50	69.76	74.61	88.82	94.99
Fluconasole	95.00	96.18	101.24	89.66	94.38
Metoprolol	95.00	75.37	79.34	90.89	95.67
Sotalol	95.00	82.17	86.50	91.53	96.35
Clonidine	95.00	97.84	102.99	96.02	101.07
Imipramine	95.00	96.57	101.66	96.30	101.37
Labetalol	95.00	89.89	94.62	100.00	105.26
Trimethoprim	97.00	93.17	96.05	93.34	96.23
Cephalexin	98.00	76.98	78.55	84.95	86.68
Warfarin	98.00	100.00	102.04	100.00	102.04
Prednisolone	98.00	92.86	94.75	96.50	98.47
Naproxen	99.00	96.46	97.44	100.00	101.01
Practolol	100.00	66.97	66.97	74.79	74.79
Loracerbef	100.00	77.08	77.08	78.23	78.23
Flavustatin	100.00	91.15	91.15	88.13	88.13

Table 4 (*Continued*)

	IA(%)	RBF	Recovery	GNN	Recovery
Antipyrine	100.00	100.00	100.00	91.79	91.79
Caffeine	100.00	91.48	91.48	92.02	92.02
Lormetazepam	100.00	100.00	100.00	92.70	92.70
Bumetanide	100.00	82.96	82.96	93.89	93.89
Testosterone	100.00	96.80	96.80	93.97	93.97
Corticosterone	100.00	88.60	88.60	94.14	94.14
Felodipine	100.00	88.69	88.69	95.29	95.29
Prazosin	100.00	95.27	95.27	95.37	95.37
Ondansetron	100.00	95.66	95.66	95.65	95.65
Desipramine	100.00	94.78	94.78	96.24	96.24
Dexamethasone	100.00	96.35	96.35	96.35	96.35
Ibuprofen	100.00	93.86	93.86	97.14	97.14
Valproic acid	100.00	95.91	95.91	98.61	98.61
Aspirin	100.00	94.01	94.01	100.00	100.00
Ketoprofen	100.00	96.93	96.93	100.00	100.00
Zidovudine	100.00	100.00	100.00	100.00	100.00
Enalapril	10.00	4.89	48.94	47.68	476.80
Pravastatin	34.00	51.36	151.05	41.06	120.76
Ranitidine	50.00	78.42	156.84	76.56	153.12
Furosemide	61.00	61.00	100.00	89.25	146.31
Lamotrigine	70.00	98.94	141.34	87.27	124.67
Bromazepam	84.00	100.00	119.05	87.38	104.02
Pindolol	90.00	86.83	96.48	95.11	105.68
Diazepam	100.00	100.00	100.00	86.70	86.70
Methotrexate	100.00	97.31	97.31	100.00	100.00
Average			107.93		109.68
S.D.			37.76		49.85
Validation data set					
Doxorubicin	5.00	5.00	100.00	0.00	0.00
Lisinopril	25.00	18.26	73.06	0.00	0.00
Cefuroxime axetil	36.00	36.00	100.00	9.76	9.76
Gabapentin	50.00	82.25	164.49	51.36	31.22
Captopril	67.00	68.23	101.84	100.00	98.19
Cefatrizin	76.00	68.55	90.20	73.98	82.02
Cimetidine	85.00	65.47	77.02	76.53	99.36
Progesterone	91.00	96.87	106.45	93.99	88.30
Alprenolol	93.00	94.02	101.09	95.95	94.91
Salicylicacid	100.00	91.28	91.28	100.00	109.55
Average			100.54		61.33
S.D.			25.01		45.34

* Cannot be calculated.

References

[1] U. Norinder, T. Osterberg, P. Artursson, Theoretical calculation and prediction of intestinal absorption of drugs in humans using MolSurf parametrization and PLS statistics, Pharm. Res. 14 (1997) 1786–1791.

[2] K. Palm, P. Stenberg, K. Luthman, P. Artursson, Polar molecular surface properties predict the intestinal absorption of drugs in human, Pharm. Res. 14 (1997) 568–571.

[3] P. Artursson, K. Palm, K. Luthman, Caco-2 monolayers in experimental and theoretical predictions of drug transport, Adv. Drug Deliv. Rev. 22 (1996) 67–84.

[4] P. Stenberg, K. Luthman, P. Artursson, Prediction of membrane permeability to peptides from calculated dynamic molecular surface properties, Pharm. Res. 16 (1999) 205–212.

[5] K. Palm, K. Luthman, A.L. Ungell, G. Strandlund, F. Beigi, P. Lundahl, P. Artursson, Evaluation of dynamic polar molecular surface area as predictor of drug absorption: comparison with other computational and experimental predictors, J. Med. Chem. 41 (1998) 5382–5392.

[6] D.C. Pang, G.L. Amidon, P.C. Preusch, W. Sadée, Second AAPS–NIH Frontier Symposium : Membrane Transporters and Drug Therapy, Pharm. Sci. 1 (1999) 48k.

[7] M. Thamotharan, S.Z. Bawani, X. Zhou, S.A. Adibi, Hormonal regulation of oligopeptide transporter pept-1 in a human intestinal cell line, Am. J. Physiol. 276 (1999) C821–C826.

[8] P.W. Swaan, B.C. Koops, E.E. Moret, J.J. Tukker, Mapping the binding site of the small intestinal peptide carrier (PepT1) using comparative molecular field analysis, Receptors Channels 6 (1998) 189–200.

[9] S. Klee, I. Baumung, K. Kluge, F.R. Ungemach, E. Horne, M. O'Keeffe, I. De Angelis, A.L. Vignoli, F. Zucco, A. Stammati, A contribution to safety assessment of veterinary drug residues: in vitro/ex vivo studies on the intestinal toxicity and transport of covalently bound residues, Xenobiotica 29 (1999) 641–654.

[10] G.L. Amidon, P.J. Sinko, D. Fleisher, Estimating human oral fraction dose absorbed: a correlation using rat intestinal membrane permeability for passive and carrier-mediated compounds, Pharm. Res. 5 (1988) 651–654.

[11] K.I. Inui, M. Yamamoto, H. Saito, Transepithelial transport of oral cephalosporins by monolayer of intestinal cell line CaCO-2; specific transport system in apical and basoteral membranes, J. Pharmacol. Exp. Ther. 261 (1992) 195–201.

[12] S. Bjorge, K.L. Halelehle, R. Homan, S.E. Rose, D.A. Turluk, D.S. Wright, Evidence for glucuronide conjugation of p-nitrophenol in the CaCO-2 cell model, Pharm. Res. 8 (1991) 1441–1443.

[13] P. Artursson, R.T. Borchardt, Intestinal drug absorption and metabolism in cell cultures: Caco-2 and beyond, Pharm. Res. 14 (1997) 1655–1658.

[14] M.D. Wessel, P.C. Jurs, J.W. Tolan, S.M. Muskal, Prediction of human intestinal absorption of drug compounds from molecular structure, J. Chem. Inf. Comput. Sci. 38 (1998) 726–735.

[15] R. Reed, Pruning algorithms — a survey, IEEE Trans. Neural Networks 4 (1993) 740–747.

[16] L.S. Schanker, D.J. Tocco, B.B. Brodie, C.A.M. Hogben, Absorption of drugs from the rat small intestine, J. Pharmacol. Exp. Ther. 123 (1958) 81–88.

[17] C.A.M. Hogen, D.J. Tocco, B.B. Brodie, L.S. Schanker, On the mechanism of intestinal absorption of drugs, J. Pharmacol. Exp. Ther. 125 (1959) 275–282.

[18] D.C. Taylor, R. Pownall, W. Burke, The absorption of beta-adrenoceptor antagonists in rat in-situ small intestine; the effect of lipophilicity, J. Pharm. Pharmacol. 37 (1985) 280–283.

[19] B.B. Brodie, C.A.M. Hogen, Some physico-chemical factors in drug action. Review article, J. Pharmacol. 3 (1957) 345–380.

[20] U. Norinder, T. Osterberg, P. Artursson, Theoretical calculation and prediction of intestinal absorption of drugs in humans using MolSurf parametrization and PLS statistics, Eur. J. Pharm. Sci. 8 (1999) 49–56.

[21] M.R. Riazi, C.H. Whitson, Estimating diffusion-coefficients of dense fluids, Ind. Eng. Chem. Res. 32 (1993) 3081–3088.

[22] M.J. Fonseca, H.J. Haisma, S. Klaassen, M.H. Vingerhoeds, G. Storm, Design of immuno-enzymosomes with maximum enzyme targeting capability: effect of the enzyme density on the enzyme targeting capability and cell binding properties, Biochim. Biophys. Acta 1419 (1999) 272–282.

[23] A. Leo, C. Hansch, C. Church, Comparison of parameters currently used in the study of structure–activity relationships, J. Med. Chem. 12 (1969) 766–771.

[24] M.A. Farnum, C.F. Zukoski, The effect of the dipole moment on protein–protein interactions, Biophys. J. 76 (1999) A127–A127.

[25] S. Trohalaki, R. Pachter, The utility of electron affinity as a QSAR descriptor in predictive toxicology, Abstr. Pap. Am. Chem. Soc. 217 (1999) U674–U674.

[26] 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.