Using General Regression and Probabilistic Neural Networks To Predict Human Intestinal Absorption with Topological Descriptors Derived from Two-Dimensional Chemical Structures

Tomoko Niwa*

Discovery Research Laboratory, Nippon Shinyaku Co., Ltd., 14, Nishinosho-Monguchi-cho, Kisshoin, Minami-ku Kyoto, 601-8550 Japan

Received March 17, 2002

The objective of this study was to develop rapid and reliable methods to predict the percent human intestinal absorption (%HIA) of compounds based on their 2D descriptors. The analyzed data set included 86 drug and drug-like molecules and was the same as that studied by Wessel and co-workers. Instead of using three-dimensional descriptors such as polar surface area, which require lengthy computations, we employed only two-dimensional topological descriptors derived from information about the two-dimensional structure of molecules. The %HIA values were modeled using a general regression neural network (GRNN) and a probabilistic neural network (PNN), variants of normalized radial basis function networks. Both networks performed well to model the %HIA values. The root-mean square (rms) error was 22.8 %HIA unit for the external prediction set for a GRNN model, and 80% of the external prediction set was correctly classified for a PNN model, indicating the potential of our approach to estimate the %HIA values for a large set of compounds as virtual libraries.

INTRODUCTION

Absorption determines the value of a drug candidate and has an impact on the whole process of drug discovery and development. In fact, achieving satisfactory ADME (absorption, distribution, metabolism, and excretion) properties of a drug candidate is as important as realizing sufficient potency and selectivity. Experimental absorption measurements are expensive and time-consuming. The computational estimation of absorption is thereby an attractive alternative to experimental measurements. Such computational estimation should proceed fast enough to handle at least 100 000 compounds in a reasonable length of time. This is due to the increased use of robotic technologies as high throughput screening and combinatorial synthesis in today's drug discovery and development.

In modeling of intestinal absorption of molecules, the polar surface area (PSA) is one of the widely used descriptors, where PSA is the sum of the van der Waals surfaces arise from nitrogen and oxygen atoms and polar hydrogens connected to nitrogen and oxygen atoms. The PSA values are hence related to the hydrogen-bonding abilities of molecules. Van de Waterbeemd and co-workers have found a correlation of PSA with the apparent permeability through a Caco-2 monolayer.[1] Palm and co-workers have proposed the dynamic polar surface area (PSA_d), the Boltzmann-weighted average value of the polar surface areas of low-energy conformers, to estimate the permeability across Caco-2 cells and rat ileum.[2] The PSA_d values take into account the shape and flexibility of a drug. Clark has shown that the PSA calculated for a representative single conformer performs as well as PSA_d to predict permeability.[3] Wessel and co-workers have reported a neural network model for percent human intestinal absorption (%HIA) prediction.[4] The descriptors they used were topological, electronic, geometric, charged-partial surface area (CPSA) and related ones. These methods, however, are not applicable to large numbers of compounds, because estimations of PSA and CPSA require conformational analysis and molecular orbital calculations, which are computationally impracticable for a large set of compounds.

The objective of this study is to develop a rapid and reliable method to predict %HIA for a large number of compounds without using 3D descriptors. There are so many kinds of 2D descriptors, and finding appropriate descriptors is a difficult task. The polar surface area (PSA) is widely used in analyzing permeability of drugs. The descriptors that can well explain PSA values should be good descriptors to model %HIA values. For these reasons, we searched for such descriptors that worked well to evaluate PSA values and defined 2D topological descriptors derived from information about the chemical structures. The neural network modeling was then performed using newly defined 2D descriptors along with 1D descriptors as ClogP.

Another point to be noted is that general regression neural networks (GRNN)[5] and probabilistic neural networks (PNN)[6,7] were selected instead of widely used Back-Propagation (BP) neural networks. GRNN and PNN are variants of normalized Radial Basis Function (RBF) networks,[8] where the sigmoid activation functions often used in neural networks are replaced by radial basis functions. GRNN is designed for regression and PNN for classification tasks, and both perform well in noisy environments. Selecting training parameters and defining network architectures are time-consuming

* Corresponding author phone: +81-75-321-9171; fax: +81-75-321-9038; e-mail: t.niwa@po.nippon-shinyaku.co.jp.

10.1021/ci020013r CCC: $25.00
Published on Web 01/03/2003

processes in modeling back-propagation neural networks. There are no training parameters such as learning rate and momentum as in back-propagation modeling. The architectures of GRNN and PNN, namely the numbers of layers and units, are defined by the numbers of compounds and descriptors in the training data. The only weight to be learned is the smoothing factor σ, the width of radial basis functions. The above-mentioned features are favorable to efficiently search for suitable descriptors from a pool of descriptors and to model the %HIA values with certain experimental uncertainties.

METHOD

General Regression Neural Network (GRNN). General Regression Neural Network (GRNN) is Donald Specht's term for the Nadaraya-Watson kernel regression originally developed in the statistics literature.[5] Briefly speaking, GRNN is a memory-based feed forward network, consisting of 4 layers: input, hidden, summation, and output layers. GRNN replaces the sigmoid activation function often used in neural networks with a radial basis function (RBF)[8] and achieves the estimation of the probability density function using Parzen's nonparametric estimator.[9] The predicted value is simply a weighted average of the target values of training patterns close to the given input pattern. The only adjustable parameter is the smoothing factor for the kernel function. The following are details of GRNN.

Assume that $f(\mathbf{x},y)$ represents the joint probability density function of a vector random variable $\mathbf{x}$, and a scalar random variable y. The predicted value is the most probable value of y (called the conditional mean of y given $\mathbf{x}$ or the regression of y given $\mathbf{x}$), as shown by eq 1.

$$E(y/\mathrm{x}) = \hat{y}(\mathrm{x}) = \frac{\int_{-\infty}^{+\infty} y f(\mathrm{x},y)dy}{\int_{-\infty}^{+\infty} f(\mathrm{x},y)dy} \tag{1}$$

The probability density function $f(\mathbf{x},y)$ can be estimated from the training set by using the Parzen's nonparametric estimator:[9]

$$f(\mathrm{x},y) = \frac{1}{n(2\pi)^{(p+1)/2}\sigma^{(p+1)}} \cdot \sum_{i=1}^{n} \exp\left[-\frac{(\mathrm{x}-\mathrm{x}_i)^T(\mathrm{x}-\mathrm{x}_i)}{2\sigma^2}\right] \cdot \exp\left[-\frac{(y-y_i)^2}{2\sigma^2}\right] \tag{2}$$

The number of training patterns and the number of independent features (descriptors) are denoted n and p, respectively. The density function $f(\mathbf{x},y)$ is thereby estimated by a weighted sum of 'kernel functions' also called 'unit functions'. In this work, the kernel function was represented by the normalized Gaussian function (eq 2). Note that the use of the normalized Gaussian function does not imply any normal assumption about the distribution of the data in the feature space. The parameter σ represents the width of the 'kernel function' and is called smoothing factor. A single smoothing factor was used for all the data.

Substituting the joint probability estimate (eq 2) into the conditional mean (eq 1) and interchanging the order of integration and summation leads to eq 3.

$$\hat{y}(\mathrm{x}) = \frac{\sum_{i=1}^{n} \exp\left[-\frac{(\mathrm{x}-\mathrm{x}_i)^T(\mathrm{x}-\mathrm{x}_i)}{2\sigma^2}\right] \int_{-\infty}^{+\infty} y \exp\left[-\frac{(y-y_i)^2}{2\sigma^2}\right] dy}{\sum_{i=1}^{n} \exp\left[-\frac{(\mathrm{x}-\mathrm{x}_i)^T(\mathrm{x}-\mathrm{x}_i)}{2\sigma^2}\right] \int_{-\infty}^{+\infty} \exp\left[-\frac{(y-y_i)^2}{2\sigma^2}\right] dy} \tag{3}$$

Assessing the two indicated integrations and using $\int_{-\infty}^{+\infty} ze^{-z^2}dz = 0$ yield the following.

$$\hat{y}(\mathrm{x}) = \frac{\sum_{i=1}^{n} y_i \exp\left[-\frac{(\mathrm{x}-\mathrm{x}_i)^T(\mathrm{x}-\mathrm{x}_i)}{2\sigma^2}\right]}{\sum_{i=1}^{n} \exp\left[-\frac{(\mathrm{x}-\mathrm{x}_i)^T(\mathrm{x}-\mathrm{x}_i)}{2\sigma^2}\right]} \tag{4}$$

The predictor eq 4 is a weighted sum over all the training patterns. Each training pattern is weighted exponentially according to its Euclidean distance to the unknown pattern $\mathbf{x}$ and also according to the smoothing factors σ.

Eq 4 was mapped into a neural network of four layers: the input layer, the hidden layer, the summation layer and the output layer (Figure 1). The number of inputs is equal to the number of independent features. Each unit in the hidden layer represents a training pattern. The summation layer includes two units: the first unit sums all the outputs of the hidden layer and evaluates the numerator of eq 4, and the second unit evaluates the denominator of eq 4. Each unit in the hidden layer is connected to each of the two units in the summation layer. The weight of the connection between the unit i in the hidden layer and the first unit of the summation layer is equal to y_i, the target value. The i weight of the connection between any unit i in the hidden layer and the second unit in the summation layer is equal to unity. The output unit merely divides the two outputs of the summation layer to yield the predicted value of the dependent feature.

Probabilistic Neural Network (PNN). GRNN is designed for regression, and PNN is designed for classification. As for GRNN, probability density functions are evaluated using the Parzen's nonparametric estimator.[9] Applicability of PNN to chemometrics problems has been intensively studied by

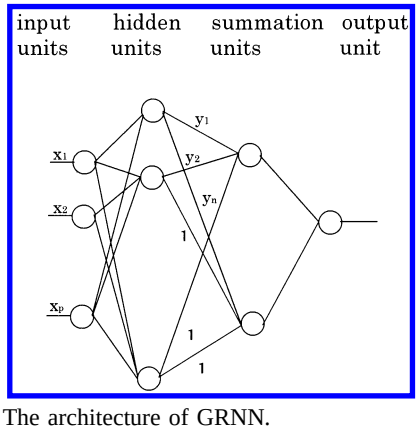

Figure 1. The architecture of GRNN.

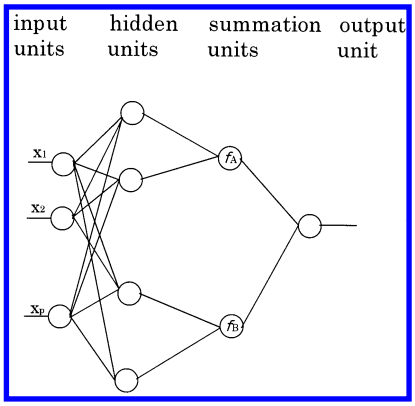

Figure 2. The architecture of PNN.

Shaffer and co-workers.[7] While GRNN uses only one probability density function, PNN utilizes one probability density function for each category, as shown by eq 5

$$f_A(\mathrm{x}) = \frac{1}{m(2\pi)^{p/2}\sigma^p} \cdot \sum_{i=1}^{m} \exp\left[-\frac{(\mathrm{x} - \mathrm{x}_{\mathrm{Ai}})^T(\mathrm{x} - \mathrm{x}_{\mathrm{Ai}})}{2\sigma^2}\right] \quad (5)$$

where m is the total number of training patterns, and $\mathbf{x}_{\mathrm{Ai}}$ is ith training pattern from category θ_A. $f_B(\mathrm{x})$ is similarly defined by eq 5 for category θ_B.

Figure 2 illustrates the architecture of PNN having four layers. For simplicity, PNN with two categories is shown. The input units are merely distribution units that supply the same input values to all of the hidden units. There is one hidden unit for each training pattern. The number of summation units is the number of the categories. The main differences between GRNN and PNN are the hidden-to-output weights. In PNN, the hidden-to-output weights are usually 1 or 0; for each hidden unit, a weight of 1 is used for the connection going to the output that the pattern belongs to, while all other connections are given weights of 0. The output unit performs classification according to the well-known Bayes's theory to minimize the "expected risk".[6,8]

Smoothing Factor Optimization. In the GRNN or PNN developments, the only weight to be learned is the smoothing factor σ, which represents the width of the 'kernel function' (see eqs 4 and 5). We applied only one smoothing factor to all of the input features. This means that all of the input features have the same impact on predicting the output. Training for GRNN and PNN proceeded in two parts. In the first part, the network was trained with the data in the training set. In the second part, a smoothing factor was iteratively optimized using the data in the test set. The mean squared error of the test data was used as the target value for optimization. An upper limit used to test a smoothing factor was set at 0.8; the smoothing factor varied from 0 to 0.8. After optimization, the external prediction set was used to study the predictive power of the obtained model.

Data Sets. The PSA data used in the present analyses were those calculated by Clark for 74 compounds.[3] The %HIA data set was the same as that studied by Wessel and co-workers: 67 compounds for the training set, 9 compounds for the test set, and 10 compounds for the external prediction set.[4] ClogP and CMR values were computed with the CLOGP ver.4.51 program.[11] 2D topological descriptors were computed with in-house programs. Multiple Linear Regression (MLR) analysis was performed with the StatView program by SAS Institute[12] and GRNN and PNN modeling with the Neuroshell 2 program by Ward Systems.[13] All computations were done on a Pentium II desktop computer.

RESULTS

Multiple Linear Regression Analyses of PSA. The polar surface area (PSA) is widely used in analyzing permeability of drugs as mentioned above, and the descriptors that can well explain PSA values should be good descriptors to model %HIA values. For these reasons, we searched for such descriptors that worked well to evaluate PSA values. The PSA values for 74 drug or drug-like molecules studied here were those calculated by Clark for a representative conformer.[3] This set of compounds was a subset of those studied by Wessel and co-worker to analyze the %HIA values (Table 1).[4]

Because PSA is the sum of the van der Waals surfaces arise from nitrogen and oxygen atoms and polar hydrogens connected to nitrogen and oxygen atoms, PSA could be expressed as the sum of surfaces of appropriately defined fragments. Based on this assumption, we defined the fragments listed in Table 2. Using the number of occurrences of each fragment gave an excellent correlation equation.

$$\begin{aligned}\mathrm{PSA} = {} & 19.129(\pm 3.080)[\mathrm{NH2}] + \\ & 16.375\ (\pm 2.853)[\mathrm{NH}] + 3.374(\pm 2.982)[\mathrm{-N<}] + \\ & 15.582\ (\pm 2.308)[=\mathrm{N-}] + 20.449(\pm 1.967)[\mathrm{OH}] + \\ & 9.477(\pm 1.384)[\mathrm{-O-}] + 19.574(\pm 1.777)[=\mathrm{O}] - \\ & 0.120(\pm 5.269)\ n = 74,\ s = 8.426,\ r = 0.985 \quad (6)\end{aligned}$$

In this equation, [X] is the number of occurrences of fragment X listed in Table 2, n is the number of compounds, s is the standard error of the estimate, and r is the correlation coefficient. Regression coefficients are given with their 95% confidence intervals. Because PSA is the sum of surfaces of polar atoms, multiple linear regression analysis is more suitable than neural network modeling in this case.

Equation 6 is statistically highly significant and could be used as an alternative rapid method to estimate PSA. Although the surface area of a molecule is apparently dependent on the conformations, PSA is computable with our simple 2D topological descriptors at least for drug and drug-like molecules. The result of eq 6 encouraged us to analyze the permeability of drug molecules without using 3D descriptors. Recently, Ertl and co-workers reported a method similar to estimate PSA values using 43 precisely defined 2D descriptors.[14] Because of the limited number of available %HIA data, fewer number of descriptors is more favorable in the present %HIA modeling. The descriptors used in eq 6 were selected for further analyses.

GRNN Modeling of HIA. 2D descriptors similar to those used in eq 6 were calculated for carbon, nitrogen, oxygen, sulfur, phosphorus and halogen atoms. We also calculated pharmacophore-feature like descriptors as the numbers of hydrogen-bond acceptors, hydrogen-bond donors, acidic centers, basic centers, positive ionizable centers, rotatable bonds, and aromatic rings. Molecular weight, ClogP and CMR values were also computed. Starting with various combinations of descriptors, redundant descriptors were excluded based on their modeling abilities. 1D descriptors,

Table 1. Actual and Calculated Values

name	%HIA[a]	c%HIA[b]	Poor[c]	cPoor[d]	Good[e]	cGood[f]	PSA[g]
gentamicin	0.00	0.00	1	1	0	0	174.4
cromoglicic-acid	0.50	0.50	1	1	0	0	184.6
olsalazine	2.30	2.64	1	1	0	0	
ganciclovir	3.80	11.95	1	1	0	0	151.8
cefuroxime	5.00	5.37	1	1	0	0	188.6
chlorothiazide	13.00	27.30	1	1	0	0	134.3
mannitol	15.00	15.00	1	1	0	0	
nadolol	34.50	36.17	1	1	0	0	85.5
norfloxacin	35.00	42.03	1	1	0	0	81.1
phenoxymethylpenicillin	45.00	64.39	1	1	0	0	107.6
etoposide	50.00	50.00	1	1	0	0	183.3
atenolol	50.00	78.52	0	0	1	1	
ziprasidone	60.00	69.59	0	0	1	1	60.5
sulfasalazine	65.00	65.27	0	0	1	1	
hydrochlorothiazide	67.00	52.94	0	0	1	1	135.2
sumatriptan	75.00	76.18	0	0	1	1	75.0
guanabenz	75.00	78.47	0	0	1	1	72.9
propylthiouracil	75.00	75.86	0	0	1	1	48.9
quinidine	80.00	84.07	0	0	1	1	48.6
acetaminophen	80.00	85.62	0	0	1	1	58.5
methylprednisolone	82.00	93.72	0	0	1	1	98.3
sorivudine	82.00	83.01	0	0	1	1	131.8
bupropion	87.00	87.46	0	0	1	1	26.2
trovafloxacin	88.00	85.23	0	0	1	1	101.0
acrivastine	88.00	87.96	0	0	1	1	60.2
acebutolol	89.50	87.27	0	0	1	1	89.8
timolol maleate	90.00	87.14	0	0	1	1	82.1
phenytoin	90.00	89.92	0	0	1	1	66.3
betaxolol	90.00	90.94	0	0	1	1	56.3
oxprenolol	90.00	90.85	0	0	1	1	
scopolamine	90.00	89.59	0	0	1	1	61.6
propranolol	90.00	90.58	0	0	1	1	
tenidap	90.00	89.32	0	0	1	1	81.9
chloramphenicol	90.00	89.57	0	0	1	1	118.3
terazosin	91.00	91.38	0	0	1	1	102.3
hydrocortisone	91.00	93.72	0	0	1	1	93.7
amoxicillin	93.50	92.95	0	0	1	1	143.4
fluconazole	95.00	93.58	0	0	1	1	75.8
metoprolol	95.00	90.85	0	0	1	1	
sotalol	95.00	90.26	0	0	1	1	90.3
clonidine	95.00	90.45	0	0	1	1	39.5
imipramine	95.00	94.70	0	0	1	1	6.0
labetalol	97.00	95.85	0	0	1	1	100.6
trimethoprim	98.00	97.60	0	0	1	1	110.6
cefalexin	98.00	93.16	0	0	1	1	115.5
warfarin	98.00	98.74	0	0	1	1	62.8
prednisolone	98.80	93.72	0	0	1	1	98.3
naproxen	99.00	98.36	0	0	1	1	53.8
practolol	100.00	78.52	0	0	1	1	
loracarbef	100.00	93.16	0	0	1	1	117.9
fluvastatin	100.00	99.96	0	0	1	1	88.2
phenazone	100.00	96.22	0	0	1	1	
caffeine	100.00	99.81	0	0	1	1	59.2
lormetazepam	100.00	98.47	0	0	1	1	55.9
dimethylbumetanide	100.00	99.45	0	0	1	1	120.8
testosterone	100.00	99.42	0	0	1	1	43.4
corticosterone	100.00	95.58	0	0	1	1	75.9
felodipine	100.00	97.77	0	0	1	1	65.2
prazosin	100.00	97.78	0	0	1	1	103.6
ondansetron	100.00	93.23	0	0	1	1	38.0
desipramine	100.00	97.47	0	0	1	1	16.5
dexamethasone	100.00	93.72	0	0	1	1	90.7
ibuprofen	100.00	96.06	0	0	1	1	42.0
valproic-acid	100.00	99.59	0	0	1	1	44.2
aspirin	100.00	98.24	0	0	1	1	69.4
ketoprofen	100.00	98.95	0	0	1	1	60.9
zidovudine	100.00	98.45	0	0	1	1	142.9
enalapril	10.00	38.49	1	1	0	0	115.1
pravastatin	34.00	86.98	1	1	0	0	123.9
ranitidine	50.00	81.57	1	0	0	1	86.7
furosemide	61.00	97.51	0	0	1	1	127.7
lamotrigine	70.00	92.51	0	0	1	1	96.4
bromazepam	84.00	92.76	0	0	1	1	57.4

Table 1. (Continued)

name	%HIA[a]	c%HIA[b]	Poor[c]	cPoor[d]	Good[e]	cGood[f]	PSA[g]
pindolol	90.00	92.71	0	0	1	1	
diazepam	100.00	98.46	0	0	1	1	
methotrexic-acid	100.00	80.99	0	0	1	1	225.0
doxorubicin	5.00	1.81	1	1	0	0	199.5
lisinopril	25.00	57.02	1	0	0	1	142.4
cefuroxime-axetil	36.00	5.00	1	1	0	0	176.8
gabapentin	50.00	91.89	1	0	0	1	63.4
captopril	67.00	97.96	0	0	1	1	61.8
cefatrizine	76.00	92.49	0	0	1	1	188.4
cimetidine	85.00	75.87	0	0	1	1	92.1
progesterone	91.00	98.93	0	0	1	1	41.2
alprenolol	93.00	89.22	0	0	1	1	
salicylic-acid	100.00	92.18	0	0	1	1	62.5

[a] Experimental %HIA values. [b] Predicted %HIA values. [c] 1 for compounds less than or equal to 50% HIA, and 0 for compounds larger than 50% HIA. [d] Calculated values for the poor absorption set. [e] 0 for compounds less than or equal to 50% HIA, and 1 for compounds larger than 50% HIA. [f] Calculated values for the good absorption set. *g* Taken from ref 3.

Table 2. Descriptors Used in the PSA Analysis

NH2	nitrogen with two hydrogen atoms
NH	nitrogen with one hydrogen atom
-N<	nitrogen with three single bonds attached to heavy atoms
=N-	nitrogen with one single bond attached to a heavy atom and one double bond attached to a heavy atom
OH	oxygen with one hydrogen atom
-O-	oxygen with two single bonds attached to heavy atoms
=O	oxygen with one double bond attached to a heavy atom

Table 3. Significant Descriptors in the GRNN Modeling

our model	
NH/NH2	nitrogen with one or two hydrogen atoms
-N<	nitrogen with three single bonds attached to heavy atoms
OH	oxygen with one hydrogen atom
-O-	oxygen with two single bonds attached to heavy atoms
=O	oxygen with one double bond attached to a heavy atom
RB	number of rotatable bonds
NAr	number of aromatic rings
Wessel's model	
NSB	number of single bonds
SHDW-6	normalized 2D projection of molecules on YZ plane
CHDH-1	charge on donatable hydrogen atoms
SAAA-2	surface area of hydrogen bond acceptor atoms/number of hydrogen bond acceptors atoms
SCAA-2	surface area x charge of hydrogen bond acceptor atoms/ number of hydrogen bond acceptor atoms
GRAV-3	cube root of gravitational index

pharmacophore-features-like descriptors, and 2D descriptor for non hydrogen-bonding atoms such as carbon were successively excluded.

Training of a GRNN was very fast, and it took less than 10 seconds to train a network. The final network was modeled with [NH/NH2], [-N<], [OH], [-O-], [=O], RB, and Nar, which are listed in Table 3. Although estimation of %HIA seems to be difficult without using 3D descriptors such as PSA, we could model the %HIA values using only seven 2D descriptors. Table 4 shows the correlation coefficient matrix. Correlation among the descriptors was very low; the highest correlation coefficient was 0.348. [NH/NH2], [-N<], [OH], [-O-], and [=O] express the hydrogen-bonding abilities of a molecules, while RB and Nar seem to be related to the size and flexibility of the molecules. It is interesting to note that although RB is derived from 2D chemical structures, it also represents conformational properties of molecules. The descriptors we used are simple and easy to understand. This would be helpful for experimental chemists to design molecules having better permeability properties.

Table 1 shows the calculated %HIA and the observed %HIA. The root-mean square (rms) errors were 6.5, 27.7, and 22.8%HIA units for the training set, the test set, and the external prediction set, respectively. The corresponding values by Wessel and co-workers were 9.4, 19.7, and 16.0%HIA units.[4] The predictive power of our model is a little worse than that by Wessel and co-workers. However, our topological descriptors are much simpler than those used by Wessel and co-workers (Table 3). According to Specht, a GRNN trains 100 000 times faster than back-propagation. These facts show the excellent modeling abilities of our descriptors and GRNN.

PNN Modeling of HIA. Like GRNN, PNN is a variant of normalized Radial Basis Function (RBF) networks.[8] Applicability of PNN to chemometrics problems has been intensively studied by Shaffer and co-workers.[7] While GRNN is designed for regression, PNN is designed for classification tasks. The %HIA values vary from 0% to 100%, and we set the threshold value as 50%. The poor class includes the compounds with less than or equal to 50%, and the good class is larger than 50%. Table 5 lists the number of compounds for each set. Modeling of PNN using the same descriptors as those used in the GRNN development gave excellent results. Of course, we tried to search other descriptors to improve PNNs but could not find a better set of descriptors.

In Table 5, the figures in the parentheses are the numbers of compounds misclassified by the optimized PNN. The percentages of the correctly classified compounds were 100%, 88.9%, and 80% for the training set, the test set, and the external prediction set, respectively. The well absorbed compounds belonging to the good class were perfectly classified, whereas the compounds belonging to the poor class were partly misclassified. This misclassification was due to the limited number of the poorly absorbed compounds. Modeling abilities are largely governed by the quality of the data in the training set. The most important thing in the real-world drug development is that the compounds belonging to the good absorption can be correctly classified. In this sense, the present PNN model proved to work very well.

Table 4. Correlation Coefficient Matrix for the Descriptors in Modeling %HIA

	[NH2/NH]	[>NX-]	[OH]	[-O-]	[=O]	RB	Nar	HIA
[NH2/NH]	1.000							
[NX]	−0.129	1.000						
[OH]	−0.197	−0.288	1.000					
[-O-]	0.082	−0.067	0.145	1.000				
[= O]	0.123	0.059	0.102	0.042	1.000			
RB	0.320	−0.027	0.161	0.348	0.153	1.000		
Nar	0.028	0.312	−0.212	0.076	−0.055	0.113	1.000	
HIA	−0.231	0.072	−0.415	−0.323	−0.299	−0.296	0.125	1.000

Table 5. Training, Test, and Prediction Sets Used for Model Development and the Prediction Results by the Optimized PNN Model

	poor class	good class
training set	11 (0)	56 (0)
test set	3 (1)	6 (0)
prediction set	4 (2)	6 (0)

[a] The figures in this table are the number of data for each set and the figures in the parentheses are the numbers of compounds misclassified by the optimized PNN.

DISCUSSION

The PSA is the sum of the van der Waals surface areas from polar atoms and is widely used to express the hydrogen-bonding capabilities of drug molecules when penetrating the intestinal membranes. According to the definition of the PSA, however, the contributions of each atom are treated equally; the hydrogen-bonding capabilities of the nitrogen of an amino group and the oxygen of a hydroxyl group are the same when the surface areas are the same. In our model, however, the contributions of each descriptor were optimized to get the best model for %HIA values. For example, [=N-] is absent in our model and does not participate in the penetration through membranes. The selected 2D descriptors in our models are more reasonable than the PSA values. In addition to the PSA related fragment descriptors, we could elucidate interesting factors affecting the %HIA values, namely Nar and RB. It should be noted that Veber and co-workers have reported the relationship between bioavailability and the number of rotatable bonds.[16]

There are some disadvantages of using 3D descriptors in %HIA estimation. First, 3D descriptors demand lengthy and expensive computations, and calculation of 3D descriptors for a large set of compounds as in a combinatorial library is actually impracticable. Second, building programs as CONCORD or CORINA cannot perfectly convert all compounds to 3D structures from 2D structures;[15] sometimes they fail to convert compounds with complicated structures, which are often valuable in drug development. Third, 3D descriptors as surface areas and charges depend on the programs or methods. For example, the charges calculated with the semiempirical AM1 method differ from those with the PM3 method, and different van der Waals values of atoms give different PSA values.

Seemingly, 3D descriptors are much better than 2D descriptors in predicting various properties of drugs. It is true that 3D conformational information is much more beneficial than 2D structural information, as in the case of drug design or quantitative-structure activity relationship (QSAR) studies. This is largely because drug-receptor interactions are very specific. For example, the drug molecules should locate on the right position of the binding site of the target receptor. However, during penetration, the interactions between drug molecules and lipophilic membrane molecules are not so specific as the interactions between drug molecules and the target receptors. Hence, it is reasonable to use 2D descriptors in %HIA predictions. Another good application of 2D descriptors is the prediction of 1-octanol/water partition coefficient (log P) values. A widely used prediction program of log *P* values uses 2D descriptors only.[11] Again, there exists no strictly specific interaction in partitioning processes.

The qualities of good predictive methods depend on the qualities of the input data. Also the mechanism of action should be unique. As pointed out by Clark,[3] the data set analyzed by Wessel and co-workers[4] contains some unsuitable data. One way to obtain a reasonable data set is to exclude doubtable compounds. However, we used all compounds in the data set for the following reasons. First, one of the main purposes of the present study is to validate the usefulness of 2D descriptors for evaluation of %HIA and to compare the predictive power of our model with that by Wessel and co-workers. For this purpose, the data set must be the same. Second, we could not exclude all doubtful data because of the limited number of available data. Third, GRNN and PNN are robust, because they estimate the output values based on the probability density function; the output values are not seriously affected by some unsuitable data. Good models were not given when back-propagation neural networks and 2D descriptors were used. We are well aware that the data set is insufficient and that further analyses using more comprehensive data are strongly demanded. Nevertheless, the predictive methods developed here clearly show the power of our 2D descriptors and GRNN and PNN models in predicting complex properties such as %HIA values. Our approaches could be also applicable to model other complex problems.

REFERENCES AND NOTES

(1) Van de Waterbeemd, H.; Camenisch, G.; Folkers, G.; Raevsky, O. A. Estimation of Caco-2 Cell Permeability Using Calculated Molecular Descriptors. *Quant. Struct.-Act. Relat.* **1996**, *15*, 480−490.

(2) Palm, K.; Stenberg, P.; Luthman, K.; Artursson, P. Polar Molecular Surface Properties Predict the Intestinal Absorption of Drugs in Humans. *Pharm. Res.* **1997**, *14*, 568−571.

(3) Clark, D. E. Rapid Calculation of Polar Molecular Surface Area and its Application to the Prediction of Transport Phenomena. 1. Prediction of Intestinal Absorption. *J. Pharm. Sci.* **1999**, *88*, 807−814.

(4) Wessel, M. D.; Jurs, P. C.; Tolan, J. W.; Muskal, S. M. Prediction of Human Intestinal Absorption of Drug Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 726−735.

(5) Specht, D. F. A General Regression Neural Network, *IEEE Transactions on Neural Networks* **1991**, *2*, 568−576.

(6) Specht, D. F. Probabilistic Neural Networks. *Neural Networks* **1990**, *3*, 109−118.

(7) Shaffer, R. E.; Rose-Pehrsson, S. L. Improved Probabilistic Neural Network Algorithm for Chemical Sensor Array Pattern Recognition. *Anal. Chem.* **1999**, *71*, 4263−4271.
(8) Bishop, C. M. In *Neural Networks for Pattern Recognition*; Oxford University Press: New York, 1995; pp 164−193.
(9) Parzen, E. On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.* **1962**, *33*, 1065−1076.
(10) Clark, D. E. Rapid Calculation of Polar Molecular Surface Area and its Application to the Prediction of Transport Phenomena. 2. Prediction of Blood-Brain Barrier Penetration. *J. Pharm. Sci.* **1999**, *88*, 815−821.
(11) ClogP version 4.51. Daylight Chemical Information Inc., 27401 Loa Altos, Suite #370, Mission Viejo, CA 92691.
(12) StatView 5.0. SAS Institute Inc., 100 SAS Campus Drive Cary, NC 27513-2414.
(13) NeuroShell 2. Ward Systems Group, Inc., Executive Park West 5 Hillcrest Drive Frederick, MD 21703.
(14) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714−3717.
(15) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008.
(16) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45*, 2615−2623.

CI020013R