Prediction of Human Intestinal Absorption of Drug Compounds from Molecular Structure

Matthew D. Wessel,[†] Peter C. Jurs,*[,†] John W. Tolan,[‡] and Steven M. Muskal*[,‡]

Department of Chemistry, The Pennsylvania State University, 152 Davey Laboratory, University Park, Pennsylvania 16802, and Affymax Research Institute, 3410 Central Expressway, Santa Clara, California 95051

Received March 20, 1998

The absorption of a drug compound through the human intestinal cell lining is an important property for potential drug candidates. Measuring this property, however, can be costly and time-consuming. The use of quantitative structure−property relationships (QSPRs) to estimate percent human intestinal absorption (%HIA) is an attractive alternative to experimental measurements. A data set of 86 drug and drug-like compounds with measured values of %HIA taken from the literature was used to develop and test a QSPR model. The compounds were encoded with calculated molecular structure descriptors. A nonlinear computational neural network model was developed by using the genetic algorithm with a neural network fitness evaluator. The calculated %HIA (cHIA) model performs well, with root-mean-square (rms) errors of 9.4%HIA units for the training set, 19.7%HIA units for the cross-validation (CV) set, and 16.0%HIA units for the external prediction set.

INTRODUCTION

The ability to deliver drugs orally is strongly preferred over alternative routes for systemic administration. This preference is due to the convenience, low cost, and high patient compliance rates for oral dosing. Compounds taken orally, however, must possess several properties to become systemically available: solubility and stability in the gastrointestinal tract, absorption through the intestinal wall, and a low rate of first-pass hepatic metabolism.

Prediction of human intestinal absorption (HIA) is a major goal in the design, optimization, and selection of candidates for development as oral drugs. The growth in drug discovery of combinatorial chemistry methods,[1,2] where large numbers of candidate compounds are synthesized and screened in parallel for *in vitro* pharmacological activity, has dramatically increased the demand for rapid and efficient models for estimating HIA and other biopharmaceutical properties.

In vivo animal studies have long been used, but these models are costly and labor intensive, have low throughput, and consume large amounts of test sample. Both animal and human *ex vivo* intestinal absorption models have also been used, but they are also labor intensive and give variable results.[3−5] Cell membrane models, most often using Caco-2 cells, have emerged over the past 10 years as a medium throughput *in vitro* model of HIA.[6−11] However, these models are also labor intensive and usually require more compound than is produced in a standard combinatorial library syntheses. An alternate computational chemistry approach using quantitative structure−property relationships (QSPRs) could potentially provide useful predictions of %HIA and reduce the need for actual compound synthesis and %HIA measurement. Furthermore, calculated %HIA (cHIA) models would be valuable in the selection and prioritization of building blocks throughout the design and focus of combinatorial libraries.

QSPR methods have successfully been used to model physicochemical,[3,4] chromatographic,[5,6] spectroscopic,[7] and toxicity properties of organic compounds.[8,9] Several computational models have also been reported for such biopharmaceutical properties as %HIA,[19−21] blood-brain barrier,[10,11] skin[12] and occular permeation,[13] pharmacokinetics,[14,15] and metabolism.[16] However, these studies all involved sets of closely related structural analogues, and models based on limited chemical space generally lack predictive value outside their structural classes. Broadly applicable QSPR models of biopharmaceutical properties must be built using compounds which cover both a wide range of the property being modeled as well as of chemical structure space.

The QSPR methodology used in this study consists of three main parts: representation of molecular structure, feature selection, and mapping. The starting point of the study is a set of compounds with known %HIA values, which are used to develop the QSPR relationship. Since the general assumption in QSPR modeling is that molecular structure causes the observed behavior of a compound, linking a series of chemical structures to properties of interest, in this case %HIA, should provide a method for modeling the property. A necessary first step involves encoding the structures. This is done by using calculated structural descriptors, which are mathematical representations of chemical structure. For example, the molecular volume of each structure can provide some information as to the size of each structure. To best encode the structures, it is typically useful to calculate a multitude of descriptors, each helpful in describing the structures in a unique way.

Once the structures have been encoded, the subset of descriptors that best encodes the property of interest must

[†] The Pennsylvania State University.
[‡] Affymax Research Institute.

S0095-2338(98)00029-8 CCC: $15.00
Published on Web 06/19/1998

be found. Feature selection methods, employing the genetic algorithm (GA)[17,18] coupled with computational neural networks[19] are used for this purpose. The GA is effective at finding minima for complex problems without any knowledge of the form of the objective function. Feature selection is necessarily involved in this particular QSPR approach because of the large numbers of descriptors that are calculated (more than 100 per compound). This differs somewhat from traditional QSPR/QSAR methodologies in which a relatively small number of descriptors are utilized.[32−35] Once a subset of descriptors is found, the descriptors are then mapped to the property of interest, using either a linear regression equation or a nonlinear computational neural network. These mapping methods effectively provide a mechanism for linking the chemical structures to their corresponding %HIA values.

EXPERIMENTAL SECTION

The computations for this work, with two exceptions, were performed at Penn State University on a DEC 3000 AXP Model 500 workstation running the OSF/1 V3.0B operating system. Those calculations involving HyperChem[20] (first exception) were performed on a Pentium PC. The three-dimensional model-building, utilizing CORINA (Version 1.6)[21] as well as the molecular fragment extractions and presence/absence determinations (second exception), was performed on a four-processor Silicon Graphics Challenge-L at Affymax Research Institute (ARI). The ADAPT (**A**utomated **D**ata **A**nalysis and **P**attern Recognition **T**oolkit)[22,23] software system was used for all computations except those discussed above and those involving computational neural networks. The neural network software was developed independently at Penn State University.

Data Set. The set of 86 drug and drug-like compounds and their experimentally derived %HIA values used in this study were gathered from literature sources. These compounds are listed in Table 1 with their experimental %HIA values and references.[3,4,11,19,40−147] Table 1 is ordered and split into three separate sections that show the training, cross-validation (CV), and external prediction sets used in this study. Much of the literature uses the term "percent absorbed" imprecisely to mean either percent intestinal absorption (%HIA) or absolute oral bioavailability, which can be lower than %HIA due to first-pass hepatic metabolism. Therefore, each reference was reviewed to ensure that intestinal absorption values were used in this modeling effort, and furthermore that these values were not dose-dependent and involved only healthy clinical populations. The subset of 64 compounds with %HIA less than 100% comprise all the literature examples we were able to find which met these criteria. The remaining 22 compounds, with 100%HIA, were randomly selected from the large number of publications on well-absorbed oral drugs. The proportion of 100%HIA compounds was kept low to minimize, as much as practical, overloading the training set with high %HIA values.

The 86 compounds in the working data set were split into a training set of 76 compounds and an external prediction set of 10 compounds. The external prediction set was chosen in such a way as to cover the range of %HIA values in the data set, and it spans the range of 5−100%HIA. The compounds in the external prediction set were never used during the model development process but were reserved to validate potential models. A cursory examination of the compounds in the data set reveals a large amount of structural diversity. In addition, of the 86 compounds, 22 absorb at 100%, 47 have absorption values at 90% or higher, and 71 compounds (or about 83% of the total data set) absorb at 50% or higher. Only 15 absorb below 50%. While the entire range spanned is 0−100%, this data set is heavily biased toward large values of absorption given the tendency toward successfully-developed orally active drug compounds.

The structures of the 86 compounds were extracted from the ARI database with ISIS[24] and transferred to the DEC Alpha workstation where they were entered into ADAPT. Accurate geometries are necessary for the calculation of certain descriptors thought to be necessary for modeling physical and chemical properties. As a result, CORINA[37] was used to generate accurate three-dimensional geometries.

Descriptor Generation and Analysis. A total of 162 descriptors was generated for each of the 86 compounds using ADAPT descriptor development routines. The descriptors fall into three general categories: topological, electronic, and geometric. Topological descriptors are derived from information about the two-dimensional structure of the molecule. Graph theory can be applied to the 2-D structures to generate a multitude of topological indices.[149−153] Other topological descriptors, such as atom counts, bond counts, and molecular weight, can be derived from the 2-D structural representation. Electronic descriptors were calculated with MOPAC using the AM1 Hamiltonian.[25] Electronic descriptors include partial atomic charges and the dipole moment. Geometric descriptors include moments of inertia, surface area, and volume.[155−158] Accurate three-dimensional geometries of the molecules are necessary to calculate descriptors of this nature. A fourth class of descriptors can be derived by combining electronic and geometric information to form hybrid descriptors. By combining the molecular surface area with partial atomic charges, charged-partial surface area (CPSA) descriptors can be calculated.[26] The same can be done with certain atom types (H, N, O, F, S) to calculate hydrogen bonding specific descriptors. Of the 162 ADAPT descriptors calculated, 84 were topological, 6 were electronic, 29 were geometric, and 43 were CPSA or H-bonding (hybrid) descriptors.

In addition, a large number of substructure fragment descriptors were also generated. These fragment descriptors were binary strings that indicated the presence or absence of 566 important substructure features or fragments. A large pool of over 3200 functional group fragments was excised from over 7000 drug/drug-like molecules found in MDL's Comprehensive Medicinal Chemistry database (CMC 97.1) using the first-order functional group extraction algorithm developed by Sello.[27] These 3200 basis-set functional groups were made more general by changing all single bonds to single or aromatic bonds and all double bonds to double or aromatic bonds, respectively. A total of 566 fragments from the basis-set pool was found in at least one, but not all, of the 86 compounds in the working set. The fragment descriptors were augmented to the pool of ADAPT descriptors. Thus, each compound was represented by 728 descriptors. The next step was to use objective feature selection to discard descriptors which contained redundant or minimal information.

Table 1. List of Compounds and Their References

ref	name	exptl %HIA	cHIA
	Training Set		
127	gentamicin	0.00	0.00
122	cromolyn	0.50	0.00
43,49	olsalazine	2.30	4.82
82	ganciclovir	3.80	6.21
64,91	cefuroxime	5.00	0.00
71,72	chlorothiazide	13.00	20.99
4,45	mannitol	15.00	16.05
46	nadolol	34.50	36.00
135,136	norfloxacin	35.00	53.63
64,92	phenoxymethylpenicillinic acid	45.00	43.89
146,147	etoposide	50.00	51.65
46,61	atenolol	50.00	79.62
11	ziprasidone	60.00	70.32
59	sulfasalazine	65.00	65.90
19,93	hydrochlorothiazide	67.00	62.08
55,56	sumatriptan	75.00	74.76
43,47	guanabenz	75.00	85.03
44	propylthiouracil	75.00	100.00
138	quinidine	80.00	95.07
3	acetaminophen	80.00	100.00
88	methylprednisolone	82.00	95.86
74,75	sorivudine	82.00	99.56
106,107	bupropion	87.00	83.05
11	trovoflaxicin	88.00	95.12
105	acrivastine	88.00	100.00
41,42	acebutolol	89.50	93.02
43,47,48	timolol maleate	90.00	77.43
98,19	phenytoin	90.00	91.79
112	betaxolol	90.00	95.10
43,47,48	oxprenolol	90.00	95.35
139,140	scopolamine	90.00	95.58
43,47,48	propranolol	90.00	95.77
141	tenidap	90.00	96.78
116,117,118	chloramphenicol	90.00	100.00
142,67	terazosin	91.00	93.87
4,60	hydrocortisone	91.00	96.30
67,68,69	amoxicillin	93.50	88.82
126	fluconazole	95.00	89.66
43,47,48	metoprolol	95.00	90.89
53,54	sotalol	95.00	91.53
119,120,121	clonidine	95.00	96.02
123	imipramine	95.00	96.30
43,47,48	labetalol	95.00	100.00
113	trimethoprim	97.00	93.34
64,65	cephalexin	98.00	84.95
109	warfarin	98.00	100.00
19,95	prednisolone	98.80	96.50
79,4	naproxen	99.00	100.00
43,50	practolol	100.00	74.79
101,102	loracarbef	100.00	78.23
76,77	fluvastatin	100.00	88.13
66	antipyrine	100.00	91.79
70	caffeine	100.00	92.02
62,86,87	lormetazepam	100.00	92.70
90	bumetanide	100.00	93.89
57	testosterone	100.00	93.97
43,44	corticosterone	100.00	94.14
110,111	felodipine	100.00	95.29
137	prazosin	100.00	95.37
52	ondansetron	100.00	95.65
123	desipramine	100.00	96.24
43,44	dexamethasone	100.00	96.35
128,129	ibuprofen	100.00	97.14
143,144,145	valproic acid	100.00	98.60
40	acetylsalicylic acid	100.00	100.00
94	ketoprofen	100.00	100.00
58	zidovudine	100.00	100.00
	Cross-Validation Set		
78,79,80	enalaprilat	10.00	47.68
74,99	pravastatin	34.00	41.06
51	ranitidine	50.00	76.56
81	furosemide	61.00	89.25
103,104	lamotrigine	70.00	87.27
62,83,84	bromazepam	84.00	87.38
43,47,48	pindolol	90.00	95.11
124	diazepam	100.00	86.70
130,131,132,133,134	methotrexate	100.00	100.00
	External Prediction Set		
125	doxorubicin	5.00	0.00
96,97	lisinopril	25.00	0.00
114,115	cefuroxime axetil	36.00	9.76
3,73	gabapentin	50.00	51.36
85	captopril	67.00	100.00
89,64	cefatrizine	76.00	73.98
62,63	cimetidine	85.00	76.53
4,100	progesterone	91.00	93.99
43,108	alprenolol	93.00	95.95
40	salicylic acid	100.00	100.00

Three methods of objective feature selection were employed. First, for the ADAPT descriptors only, all descriptors that had greater than 80% identical values were removed from the pool. Second, pairs of descriptors were examined for redundancy. If two descriptors were pairwise correlated with an $r > 0.90$, one of them was removed from the pool. Finally, the reduced pool of ADAPT descriptors was added to the 566 fragment descriptors, and steps 1 and 2 were repeated on the current descriptor pool (566 fragments + 75 ADAPT descriptors). These methods effectively reduced the entire (ADAPT plus fragments) descriptor pool to 127 members. This was a manageable number of descriptors to screen using subjective methods of feature selection. Within the 127-member reduced descriptor pool, there were 61 fragment descriptors, 25 topological descriptors, 21 CPSA/H-bonding descriptors, and 20 geometric descriptors.

Multiple Linear Regression Analysis. Multiple linear regression models that link the property of interest to the structures can be developed using subsets of descriptors selected from the reduced descriptor pool. A simulated annealing (SA) feature selection routine and a genetic algorithm (GA) feature selection routine were used to find good descriptor subsets.[161−163] Each method performed a directed search of the descriptor space in order to determine an optimal subset of descriptors to be used in a linear regression model. The driving force behind each algorithm is the continuous reduction of the rms errors of %HIA estimation from subset to subset. Subsets of descriptors that give lower rms errors are favored.

Computational Neural Networks. Neural networks were originally designed to mimic the activity of a system of neurons. There are many different types of neural networks, but the type that has been most useful for quantitative structure−property relationships is the fully connected, feed-forward neural network. This network consists of a multilayer system of neurons, with each neuron in a given layer fully connected to all neurons in the two adjacent levels. The objective of a neural network is to map a set of input data to a particular set of output data. In this case, molecular structure descriptors, linearly scaled to the range (0,1), serve as input data, and the %HIA values serve as output data. The connections between neurons are known as weights. A neural network is trained to map a set of input data to a corresponding set of output data by iterative adjustment of the weights. In this study, our networks were trained using a quasi-Newton optimization algorithm. Detailed discussions of the type of neural network and the training algorithm used in this study have been published previously.[31,28]

Often, a subset of descriptors that supports a good linear model is used to develop a nonlinear neural network model. However, there is no reason to believe that the best linear subset of descriptors is also the best subset for a nonlinear model. In fact, it is likely that this is not the case. Therefore, a feature selection routine which combines the genetic algorithm with a neural network fitness evaluator[29] was used for this study. The GA/neural network routine selects subsets from the reduced descriptor pool that support good nonlinear models by using neural networks to evaluate each potential subset. The genetic algorithm uses the rms error to find a good subset of descriptors. This forces the algorithm to find descriptor subsets that minimize the number of large outliers, at the possible expense of overall model quality. The genetic algorithm used in this study also incorporates the PRESS statistic[30] to improve the chances of finding a general and predictive model. In any optimization procedure similar to the one described here, the starting conditions can greatly influence the final results. This is largely due to the multivariate nature of the problem. Therefore, it should not be surprising that the "best" subset of descriptors found by the GA will differ from run to run. It is also fully expected that as more and more compounds are added to the training set, the GA will find different, but perhaps qualitatively overlapping, subsets of descriptors.

RESULTS AND DISCUSSION

Regression Analysis. The SA and GA feature selection routines were employed to seek good linear models. However, none of the models found was satisfactory, with the best model having a training set rms error of 20.4%HIA units and a prediction set rms error of about 35%HIA units. This model was also shown to be statistically invalid using basic linear regression diagnostics.[166] It became evident that the diversity of this data set and the number of data points possessing greater than 50% absorption values (about 83%) are largely responsible for producing poor quality linear models. Therefore, this data set became a good candidate for developing nonlinear neural network models directly using the GA/neural network feature selection routine.

Neural Network Analysis. The 127-member reduced pool of descriptors was fed to the GA/neural network feature selection routine for the purpose of developing a nonlinear model. The original regression training set was split randomly into a neural network training set of 67 compounds and a cross-validation (CV) set of nine compounds. The original 10-member external prediction set was used to validate any neural network models. The CV set was used to monitor overtraining of the network, and the training set was used to actually train the network. The CV set and training set rms errors are used by the GA to determine a cost function that relates directly to the overall quality of a particular subset. The cost function is calculated with the following equation:

$$\mathrm{COST} = \mathrm{TSET} + \mathrm{CVFC} \times (|(\mathrm{TSET} - \mathrm{CVSET})|) \qquad (1)$$

where TSET is the training set rms error, CVSET is the CV set rms error, and CVFC is the weight factor for the value in parentheses. In this study, a value of 0.4 was used for CVFC because it has been found to produce the best behavior of the genetic algorithm in a number of previous studies.

To decrease the possibility of chance effects influencing neural network training, the ratio of observations to total adjustable parameters should be at or above 2.0.[31] A neural network consisting of 6 input neurons (descriptors), 4 hidden neurons, and 1 output neuron (target, %HIA), thus producing a 6-4-1 architecture, was used since it produced the maximum number of adjustable parameters recommended for a data set of this size. For this 6-4-1 architecture, the ratio of training set observations to adjustable parameters ρ was 67/33, or 2.03. Other architectures were examined (6-2-1, 6-3-1), but they produced poorer quality neural network models.

Using this 6-4-1 network architecture, the GA routine searched the reduced descriptor pool for subsets that supported good models. Several models with good cost func-

Table 2. The Six Descriptors in the Neural Network Model for cHIA Estimation

Descriptor Label - Definition
NSB - number of single bonds
SHDW-6 - normalized 2D projection of molecule on YZ plane
CHDH-1 - charge on donatable hydrogen atoms
SAAA-2 - surface area of hydrogen bond acceptor atoms/number of hydrogen bond acceptor atoms
SCAA-2 - surface area × charge of hydrogen bond acceptor atoms/number of hydrogen bond acceptor atoms
GRAV-3 - Cube root of gravitational index

Table 3. Linear Correlation-Coefficient Matrix for the Six Descriptors in the Final Network Model[a]

descriptor	1	2	3	4	5	6
1	1.000					
2	−0.169	1.000				
3	0.414	0.036	1.000			
4	−0.021	0.098	0.233	1.000		
5	0.208	−0.079	−0.093	−0.580	1.000	
6	0.631	−0.171	0.205	−0.152	0.029	1.000

[a] In the same order as they appear in Table 2.

tions were found by the GA routine. The best subset of descriptors was then taken and further optimized for network performance. The results of training a feed-forward neural network are dependent upon the initial set of weights and biases used, which are assigned randomly. Because of this dependence, it is often necessary to retrain the same network several times with different initial sets of weights and biases. The goal is to find an optimal starting point that will lead to a quality termination point and a good network model. A generalized simulated annealing (GSA)[32] routine was developed at Penn State for this purpose. The GSA routine searches the complex *n*-dimensional network error surface for optimal starting points. In fact, if several training sessions are conducted on the same network, using a different set of starting weights and biases for each session, a pattern of errors with a Gaussian-like distribution emerges. The GSA routine attempts to find starting points that will lead to termination points in the low error region of this distribution. While it is nearly impossible to find the global minimum, and harder still to verify this, a close approximation to the global minimum can be found with the GSA routine.

After the genetic algorithm runs were completed, several sets of weights and biases were then found in separate CNN trainings. The set that produced the best training set and cross-validation set errors was then validated with the external prediction set. The six descriptors that comprised the best subset found by the GA are shown in Table 2. Of the six descriptors, 1 is a topological descriptor, 3 are hydrogen bonding descriptors, and 2 are geometric descriptors. Table 3 shows the correlation coefficient matrix for all six descriptors. The mean value is 0.21, and the highest correlation coefficient between any two of these six descriptors is 0.63. The gravitational index descriptor is calculated from the following equation

$$\mathrm{GRAV} = \sqrt[3]{\sum_{i=1}^{n} \frac{w_i^1 w_i^2}{D_i}} \qquad (2)$$

where n is the total number of bonds between all non-hydrogen atoms in the structure, w_i is the atomic weight of one of the two atoms joined by the ith bond, and D_i is the length of the ith bond. The six descriptors span the following ranges: NSB (3−35), SHDW-6 (0.36−0.76), CHDH-1 (0.00−1.30), SAAA-2 (3.91−38.23), SCAA-2 (−0.28 to −18.38), GRAV-3 (8.76−15.75). Of the six descriptors in the final model, none were binary fragment descriptors. Initially, the fragment descriptors were included in many preliminary models, but as more compounds were added to the data set over time (arriving at a final set of 86), the fragment descriptors began to appear less frequently. Often, a model that included fragment descriptors would not validate with the external prediction set. The descriptors in this model do not encode a causal relationship between structure and %HIA. However, it is useful to examine qualitatively the possible meaning of each descriptor. The NSB descriptor is encoding single bonds, and this may be an indication of the amount of structural flexibility. The SHDW-6 and GRAV-3 descriptors are encoding molecular size, shape, and bulk properties. These size descriptors may be important with respect to the ability of the drug to penetrate cell membranes. The three remaining descriptors are all hydrogen bonding descriptors. These can be thought of as indicators of the degree of the lipophobic and lipophilic character of a drug compound in biological environments. While the above model is effective at estimating %HIA, a second six-descriptor model, derived from descriptors independent of MOPAC (*i.e.*, no charge information), was also developed. The MOPAC independent model is not as accurate as the model reported above, but it is more effective for use in batch processing of combinatorial libraries.[33]

The training set rms error for this six-descriptor neural network model was 9.4%HIA units. The mean absolute error (mae) was 6.7%HIA units. These values were calculated after all output values from the network greater than 100% or less than 0% were fixed at 100% or 0%, respectively. The CV set rms error was 19.7%HIA units (mae 15.4%HIA units). Figure 1 shows a plot of cHIA vs observed %HIA for the training and CV sets. There is a good fit to the 1:1 correlation line. Validation of the model was performed using the 10-compound external prediction set. The rms error for the external prediction set was 16.0%HIA units (mae 11.0%HIA units), a good validation of the model. A plot of cHIA vs observed %HIA is shown in Figure 2. It is likely that the overprediction of absorption values above 50% is mainly due to the original bias in the training set.

The interpretation of effects of individual descriptors within any QSPR model is a difficult task, and even more so when the model is both multivariate and nonlinear such as with neural network models. However, some insight into the degree of nonlinear behavior of each descriptor can be assessed with functional dependence plots.[170−172] Functional dependence plots assume a fixed set of weights and biases, typically the set producing the best results. The value of a single input is varied through its range, while all other input

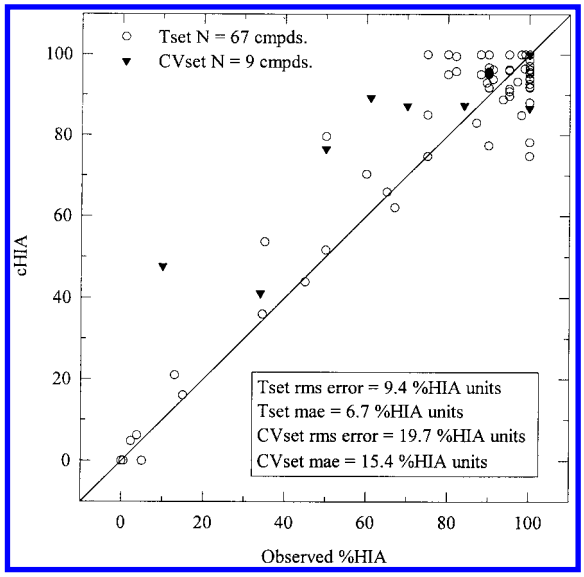

Figure 1. Plot of calculated percent human intestinal absorption (cHIA) vs observed %HIA for the training set and cross-validation set compounds. Compound set membership is shown in Table 1.

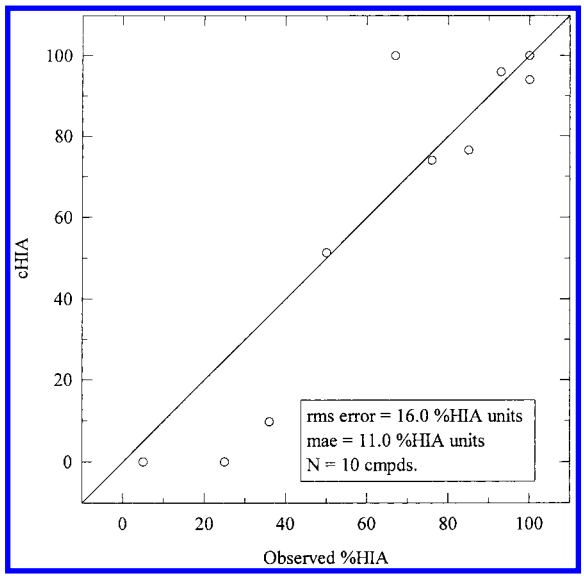

Figure 2. Plot of predicted percent human intestinal absorption (cHIA) vs observed %HIA for the external prediction set compounds. Compound set membership is shown in Table 1.

units are held constant. The network output is plotted against the variable descriptor input to generate a functional dependence plot. Figure 3 displays a functional dependence plot for three of the six descriptors in the final model for predicting %HIA. The three descriptors (labeled as in Table 2) are NSB (○), SAAA-2 (▽), and GRAV-3 (□). They are all scaled to the range (0,1). A constant value of 0.5 was used for the fixed input units as the remaining input unit was varied through its range. The high degree of nonlinearity of the three descriptors is clearly evident in Figure 3. It should be noted that the shape of any given functional dependence plot will be altered (sometimes drastically) if the values of the fixed inputs are changed.[170] This is because of the interdependence of the descriptors within the framework of the neural network. This evidence suggests that the descriptors are acting as a group rather than as individuals for the %HIA neural network model. While these functional dependence plots are suggestive of the complex relationships

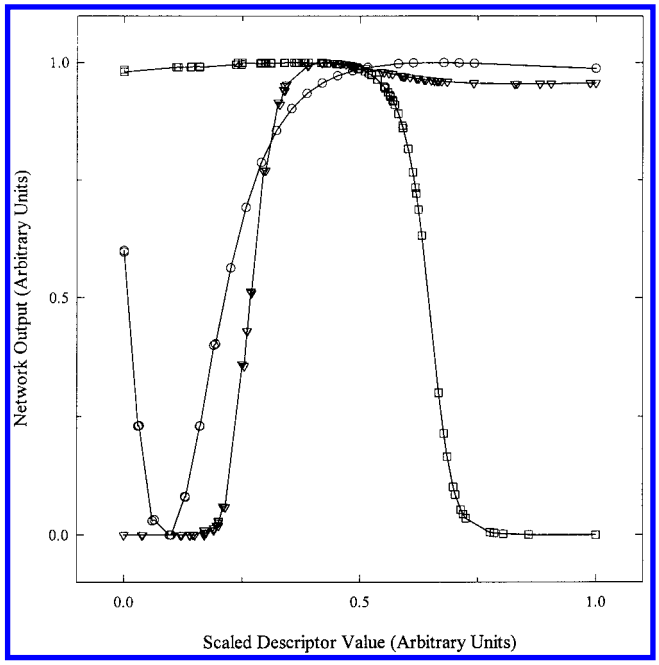

Figure 3. Functional dependence plots for three descriptors from the final model. The three descriptors are (as labeled in Table 2) NSB (○), SAAA-2 (▽), and GRAV-3 (□). The value for all fixed inputs was 0.5.

between the input descriptors and the %HIA, the detailed significance of an individual descriptor and how it relates to %HIA is very difficult to gauge.

In general, the final set of weights and biases used is chosen based on the performance of the CV set and a positive validation with an external prediction set. However, it may be that any single set of weights and biases is not necessarily the best set to use for prediction. In fact, the idea of using a committee of neural networks (or several different sets of weights and biases) has been proposed.[34] For this study, eight different sets of weights and biases, each giving good training set and CV set performance, were used to generate eight different sets of predictions. The same compounds were used for the external prediction set throughout. Average predictions were then generated from the eight individual predictions. This final, average set of predictions had an rms error of 17.7%HIA units and an mae of 13.0%HIA units. This is not quite as good a result as the single best set of weights and biases reported above (rms 16.0%HIA units, mae 11.0%HIA units). However, this approach may produce more reasonable predictions for new sets of data because of the effect of averaging. This may be a useful way to deal with the problem of multiple minima that is encountered when using neural networks.

One of the problems encountered when developing QSPRs is the possibility of chance correlations. It has been shown that performing feature selection on a pool of independent random variables can lead to linear correlations with a given dependent variable if the number of independent variables in the pool is much larger than the total number of observations.[35] Obviously, since the variables are random, there can be no real meaning attached to the correlation. To ensure that chance effects did not influence the current study, a randomized test was performed. The dependent variables of each of the compounds in the training set and cross-validation sets were scrambled randomly, and the GA was

run again. The cost function of eq 1 was 19.3%HIA units. The cost function for the best real model was 13.5%HIA units. This was the expected result, given that the target values were scrambled randomly. Also, the prediction set rms error from the randomized model was 41.7%HIA units, as opposed to 16.0%HIA units from the real model. The cost function for the scrambled data is 50% higher than that for the real data, which indicates that the model built from the real data was not based on chance. Keeping the ρ value above 2.0 also helps to avoid the problem of findings due to chance.

The process of intestinal absorption of drug compounds depends both on complex biological processes (including passive membrane penetration, active transport mechanisms and metabolism in the gastrointestinal tract) and on compound physicochemical properties (including solubility, dissolution rate, and dissociation constants). Therefore, we do not expect that a QSPR model derived using 76 diverse compounds will be a highly precise and rugged predictive tool. A much larger training set, presently unavailable in the published literature, would be required to build a model based not only on structural diversity but also on diverse biological and physicochemical properties. In fact, QSPR models have been demonstrated for such compound characteristics as solubility and pK_a.[36,37] Instead, this model is intended to serve as a valuable tool for both individual and compound library design to significantly improve the likelihood of overall increased %HIA of compounds selected for synthesis. As shown in Figure 2, this model does not produce an exact rank ordering, but it clearly differentiates the well-absorbed compounds from the poorly absorbed ones.

CONCLUSIONS

A six-descriptor nonlinear computational neural network model has been developed for the estimation of %HIA values for a data set of 86 drug and drug-like compounds. The six descriptors in the final model are listed in Table 2. The training set rms error was 9.4%HIA units, and the CV set rms error was 19.7%HIA units. Based on the rms errors of the training and CV sets, it is clear that a link between structure and %HIA does exist. However, the strength of that link is best measured by the quality of the external prediction set. With an rms error of 16.0%HIA units and a good visual plot, the external prediction set ensures the quality of the model. Given the structural diversity and bias of the data set, this is a good first attempt at modeling human intestinal absorption using QSPR methods.

A basic QSPR for estimation of %HIA values of drug and drug-like compounds is presented in this paper. The model can be used as a potential virtual screen or property estimator. With a larger data supply less biased toward the high end values of %HIA, a more successful model could likely be developed. This study illustrates the potential of using QSPR methods to aid in the drug development process.

ACKNOWLEDGMENT

The authors would like to greatly acknowledge Karen Lockhart, Jennifer Irvine, and Margrethe Schibig (ARI) and Anne Hersey (Glaxo-Wellcome) for their efforts with respect to the literature searching and validation of literature data and Russell Grove and Barry Selick (ARI) for their valuable discussions and suggestions. We would also like to thank Stephen Johnson (Penn State University) for expanding the GA/Neural Network algorithm to include the PRESS statistic. This research was supported by the Affymax Research Institute.

REFERENCES AND NOTES

(1) Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P.; Gordon, E. M. Application of Combinatorial Technologies to Drug Discovery. 1. Background and Peptide Combinatorial Libraries. *J. Med. Chem.* **1994**, *37*, 1233−1251.

(2) Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P.; Gallop, M. A. Applications of Combinatorial Technologies to Drug Discovery. 2. Combinatorial Organic Synthesis, Library Screening Strategies, and Future Directions. *J. Med. Chem.* **1994**, *37*, 1385−1401.

(3) Stewart, B. H.; Chan, O. H.; Lu, R. H.; Reyner, E. L.; Schmid, H. L.; Hamilton, H. W.; Steinbaugh, B. A.; Taylor, M. D. Comparison of Intestinal Permeabilities Determined in Multiple In Vitro and In Situ Models: Relationship to Absorption in Humans *Pharm. Res.* **1995**, *13*, 693−699.

(4) Rubas, W.; Jezyk, N.; Grass, G. M. Comparison of the Permeability Characteristics of a Series of Peptides Using an In Vitro Cell Culture Model (Caco-2) and Those Using an In Situ Perfused Rat Ileum Model of the Intestinal Mucosa. *Pharm. Res.* **1993**, *10*, 113−118.

(5) Kim, D. C.; Burton, P. S.; Borchardt, R. T. A Correlation Between the Permeability Characteristics of a Series of Peptides Using an In Vitro Cell Culture Model (Caco-2) and Those Using an In Situ Perfused Rat Ileum Model of the Intestinal Mucosa. *Pharm. Res.* **1993**, *10*, 1710−1714.

(6) Hidalgo, I. J.; Raub, T. J.; Borchardt, R. T. Characterization of the Human Colon Carcinoma Cell Line (Caco-2) as a Model System for Intestinal Epithelial Permeability. *Gastroenterology* **1989**, *96*, 736−749.

(7) Artursson, P. Cell Cultures as Models for Drug Absorption Across the Intestinal Mucosa. *Crit. Rev. Drug Carrier Systems* **1991**, *8*, 305−330.

(8) Gan, L.-S.; Eads, C.; Niederer, T.; Bridgers, A.; Yanni, S.; Hsyu, P.-H.; Pritchard, F. J.; Thackker, D. Use of Caco-2 Cells as an In Vitro Intestinal Absorption and Metabolic Model. *Drug Dev. Ind. Pharm.* **1994**, *20*, 615−631.

(9) Artursson, P.; Palm, K.; Luthman, K. Caco-2 Monolayers in Experimental and Theoretical Predictions of Drug Transport. *Adv. Drug Del. Rev.* **1996**, *22*, 67−84.

(10) Rubas, W.; Cromwell, M. E.; Shahrokh, Z.; Villagran, J.; Nguyen, T. N.; Wellton, M.; Nguyen, T. H.; Mrsny, R. J. Flux Measurements Across Caco-2 Monolayers may Predict Transport in Human Large Intestinal Tissue. *J. Pharm. Sci.* **1996**, *85*, 165−169.

(11) Yee, S. In Vitro Permeability Across Caco-2 Cells (Colonic) can Predict In Vivo (Small Intestinal) Absorption in Man - Fact or Myth. *Pharm. Res.* **1997,** *34*, 1242−1250.

(12) Wang, S.; Milne, G. W. A.; Klopman, G. Graph Theory and Group Contribution in the Estimation of Boiling Points. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1242−1250.

(13) Wessel, M. D.; Sutter, J. M.; Jurs, P. C. Prediction of Reduced Ion Mobility Constants of Organic Compounds from Molecular Structure. *Anal. Chem.* **1996**, *63*, 4237−4243.

(14) Stanton, D. T.; Jurs, P. C. Computer-Assisted Prediction of Gas Chromatographic Retention Indices of Pyrazines. *Anal. Chem.* **1989**, *61*, 1328−1332.

(15) Hasan, M. N.; Jurs, P. C. Prediction of Gas and Liquid Chromatographic Retention Indices of Polyhalogenated Biphenyls. *Anal. Chem.* **1990**, *62*, 2318−2323.

(16) Anker, L. S.; Jurs, P. C. Prediction of Carbon-13 Nuclear Magnetic Resonance Chemical Shifts by Artificial Neural Networks. *Anal. Chem.* **1992**, *64*, 1157−1164.

(17) Johnson, S. R.; Jurs, P. C. Prediction of Acute Mammalian Toxicity from Molecular Structure for a Diverse Set of Substituted Anilines Using Regression Regression Analysis and Computational Neural Networks. In *Computer-Assisted Lead Finding and Optimization*; van de Waterbeemd, H., Testa, B.; Folkers, G., Eds.; Wiley-VCH: New York, 1997; pp 29−48.

(18) *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: New York, 1995; Vol. 2.

(19) Dressman, J. B.; Amidon, G. L.; Fleisher, D. Absorption Potential: Estimating the Fraction Absorbed for Orally Administered Compounds. *J. Pharm. Sci.* **1985**, *74*, 588−589.

(20) Hamilton, H. W.; Steinbaugh, B. A.; Stewart, B. H.; Chan, O. H.; Schmid, H. L.; Schroeder, R.; Ryan, M. J.; Keiser, J.; Taylor, M. D.; Blankley, C. J.; Kalttenbronn, J. S.; Wright, J.; Hicks, J. Evaluation of Physicochemical Parameters Important to the Oral Bioavailability of Peptide-Like Compounds: Implications for the Synthesis of Renin Inhibitors. *J. Med. Chem.* **1995**, *38*, 1446−1455.

(21) Palm, K.; Luthman, K.; Ungell, A. L.; Strandlund, G.; Artursson, P. Correlation of Drug Absorption with Molecular Surface Properties. *J. Pharm. Sci.* **1996**, *85*, 32−39.

(22) Abraham, M. H.; Chadha, H. S.; Mitchell, R. C. Hydrogen Bonding. 33. Factors that Influence the Distribution of Solutes Between Blood and Brain. *J. Pharm. Sci.* **1994**, *83*, 1257−1268.

(23) Basak, S. C.; Gute, B. D.; Drewes, E. R. Predicting Blood-Brain Transport of Drugs: A Computational Approach. *Pharm. Res.* **1996**, *13*, 775−778.

(24) Potts, R. O.; Guy, R. H. A Predictive Algorithm for Skin Permeability: The Effects of Molecular Size and Hydrogen Bond Activity. *Pharm. Res.* **1995**, *12*, 1628−1663.

(25) Yoshida, F.; Topliss, J. G. Unified Model for the Corneal Permeability of Related and Diverse Compounds with Respect to Their Physicochemical Properties. *J. Pharm. Sci.* **1996**, *85*, 819−823.

(26) Kaul, S.; Ritschel, W. A. Quantitative Structure- -Pharmacokinetic Relationship of a Series of Sulfonamides in the Rat. *Eur. J. Drug Metab. Pharmacokinet.* **1990**, *15*, 211−217.

(27) Gobburu, J. V. S.; Shelver, W. H. Quantitative Structure-Pharmacokinetic Relationships (QSPR) of Beta Blockers Derived Using Neural Networks. *J. Pharm. Sci.* **1995**, *84*, 862−865.

(28) Cupid, B. C.; Beddell, C. R.; Lindon, J. C.; Wilson, I. D.; Nicholson, J. K. Quantitative Structure-Metabolism Relationships for Substituted Benzoic Acids in the Rabbit: Prediction of Urinary Excretion of Glycine and Glucuronide Conjugates. *Xenobiotica* **1996**, *26*, 157−176.

(29) Lucasius, C. B.; Kateman, G. Understanding and Using Genetic Algorithms Part 1. Concepts, Properties and Context. *Chemom. Intell. Lab. Sys.* **1993**, *19*, 1.

(30) Hibbert, D. B. A Hybrid Genetic Algorithm for the Estimation of Kinetic Parameters. *Chemom. Intell. Lab. Sys.* **1993**, *19*, 319−329.

(31) Xu, L.; Ball, J. W.; Dixon, S. L.; Jurs, P. C. Quantitative Structure−Activity Relationships for Toxicity of Phenols Using Regression Analysis and Computational Neural Networks. *Environ. Toxicol. Chem.* **1994**, *13*, 841−851.

(32) Hammet, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96−103.

(33) Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. Correlation of Biological Activity of Phenoxyacetic Acids with Hammet Substituent Constants and Partitition Coefficients. *Nature* **1962**, *194*, 178−180.

(34) Hansch, C. A. Quantitative Approach to Biochemical Structure−Activity Relationships. *Acc. Chem. Res.* **1969**, *2*, 232−239.

(35) Hansch, C. Quantitative Structure−Activity Relationships and the Unnamed Science. *Acc. Chem. Res.* **1993**, *2*, 147−153.

(36) Hypercube, Inc. 1115 NW 4th Street, Gainsville, FL 32601-4256.

(37) Sadowski, J.; Gasteiger, J. From Atoms to Bonds to Three-Dimensional Atomic Coordinates: Automatic Model Builders. *Chem. Rev.* **1993**, *93*, 2567−2581.

(38) Stuper, A. J.; Brugger, W. E.; Jurs, P. C. *Computer-Assisted Studies of Chemical Structure and Biological Function*; Wiley-Interscience: New York, 1979.

(39) Jurs, P. C.; Chou, J. T.; Yuan, M. In *Computer-Assisted Drug Design*; Olson, E. C., Christoffersen, R. E., Eds.; The American Chemical Society: Washington, D.C., 1979; pp 103−129.

(40) Latini, R.; Cerletti, C.; de Gaetano, G.; Dejana, E.; Galletti, F.; Urso, R.; Marzot, M. Comparative Bioavailability of Aspirin from Buffered, Enteric-Coated and Plain Preparations. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **1986**, *24*, 313−318.

(41) Gabriel, R.; Kaye, C. M.; Sankey, M. G. Preliminary Observations on the Excretion of Acebutolol and its Acetyl Metabolite in the Urine and Faeces of Man. *J. Pharm.* **1981**, *33*, 386−387.

(42) Foster, R. T.; Carr, R. A.; Acebutolol. In Analytical Profiles of Drug Substances; Florey, K., Ed.; Academic: New York, 1990; Vol. 19, pp 1−26.

(43) Artursson, P.; Karlsson, J. Correlation between Oral Drug Absorption in Humans and Apparent Drug Permeability. *Biochem. Biophys. Res. Commun.* **1991**, *175*, 880−885.

(44) Harvey, S. C.; Winthrow, C. D.; Hormones. In Remington's Pharmaceutical Sciences; Gennaro, A. R. Ed.; Mack Publishing Company: Easton, PA, 1985; Vol. 17, pp 951−1001.

(45) Ukabam, S. O.; Cooper, B. T. Small Intestinal Permeability to Mannitol, Lactulose, and Polyethylene Glycol 400 in Celiac Disease. *Dig. Dis. Sci.* **1984**, *29*, 809−816.

(46) Cruickshank, J. M. The Clinical Important of Cardioselectivity and Lipophilicity in Beta Blockers. *Am. Heart. J.* **1980**, *100*, 160−178.

(47) Ruffalo, R. L.; Garabedian-Ruffalo, S. M.; Garret, B. N. A Rational Therapeutic Approach to the Treatment of Essential Hypertension Part I. Cardiovas. *Rev. Rep.* **1986**, *7*, 692−700.

(48) Meier, J. Pharmacokinetic Comparison of Pindolol with Other Beta-Adrenoceptor-Blocking Agents. *Am. Heart J.* **1982**, *104*, 364−73.

(49) Ryde, E. M.; Ahnfelt, N. O. The Pharmacokinetics of Olsalazine Sodium in Healthy Volunteers After a Single I.V. Dose and After Oral Doses With and Without Food. Eur. *J. Clin. Pharmacol.* **1988**, *34*, 481−488.

(50) Bodem, G.; Chidsey, C. A. Pharmacokinetic Studies of Practolol, a Beta Adrenergic Antagonist, in Man. *Clin. Pharmacol. Ther.* **1973**, *14*, 26−29.

(51) Miller, R. Pharmacokinetics and Bioavailability of Ranitidine in Humans. *J. Pharm. Sci.* **1984**, *73*, 1376−1379.

(52) Roila, F.; Del Favero, A. Ondansetron Clinical Pharmacokinetics. *Clin. Pharmacokinet.* **1995**, *29*, 95−109.

(53) Anttila, M.; Arstila, M.; Pfeffer, M.; Tikkanen, R.; Vallinkoski, V.; Sundquist, H. Human Pharmacokinetics of Sotalol. Acta. Pharmacol. *Toxicol. (Copenh)* **1976**, *39*, 118−128.

(54) Hanyok, J. J. Clinical Pharmacokinetics of Sotalol. *Am. J. Cardiol.* **1993**, *72*, 19A−26A.

(55) Fowler, P. A.; Lacey, L. F.; Thomas, M.; Keene, O. N.; Tanner, R. J.; Baber, N. S. The Clinical Pharmacology, Pharmacokinetics and Metabolism of Sumatriptan. *Eur. Neurol.* **1991**, *31*, 291−294.

(56) Dixon, C. M.; Saynor, D. A.; Andrew, P. D.; Oxford, J.; Bradbury, A.; Tarbit, M. H. Disposition of Sumatriptan in Laboratory Animals and Humans. *Drug Metab. Dispos.* **1993**, *21*, 761−769.

(57) Shinohara, Y.; Baba, S.; Kasuya, Y. Absorption, Metabolism, and Excretion of Oral Testosterone in Humans by Mass Fragmentography. *J. Clin. Endocrinol. Metab.* **1980**, *51*, 1459−1462.

(58) Acosta, E. P.; Page, L. M.; Fletcher, C. V. Clinical Pharmacokinetics of Zidovudine. *Clin. Pharmacokinet.* **1996**, *30*, 251−262.

(59) Pieniaszek, H. J., Jr.; Resetarits, D. E.; Wilferth, W. W.; Blumenthal, H. P.; Bates, T. R. Relative Systemic Availability of Sulfapyridine from Commercial Enteric-Coated and Uncoated Sulfasalazine Tablets. *J. Clin. Pharmacol.* **1979**, *1*, 39−45.

(60) Peterson, R. E.; Wyngaarden, J. B.; Guerra, L. S.; Brodie, B. B.; Bunim, J. J. The Physiological Disposition and Metabolic Fate of Hydrocortisone in Man. *J. Clin. Invest.* **1955**, *34*, 1778−1794.

(61) Riley, S. A.; Kim, M.; Sutcliffe, F.; Rowland, M.; Turnberg, L. A. Absorption of Polar Drugs Following Caecal Instillation in Healthy Volunteers. Aliment. *Pharmacol. Ther.* **1992**, *6*, 701−706.

(62) Macheras, P.; Reppas, C.; Dressman, J. B. Estimate of Volume/Flow Ratio of Gastrointestinal (GI) Fluids in Humans Using Pharmacokinetic Data. *Pharm. Res.* **1990**, *7*, 518−522.

(63) Schentag, J. J.; Cerra, F. B.; Calleri, G. M.; Leising, M. E.; French, M. A.; Bernhard, H. Age, Disease, and Cimetidine in Healthy Subjects and Chronically Ill Patients. *Clin. Pharmacol. Ther.* **1981**, *29*, 737−743.

(64) Amidon, G. L.; Sinko, P. J.; Fleisher, D. Estimating Human Oral Fraction Dose Absorbed: A Correlation Using Rat Intestinal Membrane Permeability for Passive and Carrier-Mediated Compounds. *Pharm. Res.* **1988**, *5*, 651−654.

(65) Nicholas, P.; Meyers, B. R.; Hirschman, S. Z. Cephalexin: Pharmacologic Evaluation Following Oral and Parenteral Administration. *J. Clin. Pharmacol.* **1973**, *13*, 463−468.

(66) Eichelbaum, M.; Ochs, H. R.; Roberts, G.; Somogyi, A. Pharmacokinetics and Metabolism of Antipyrine (Phenazone) after Intravenous and Oral Administration. *Drug. Res.* **1982**, *32*, 575−578.

(67) Sennello, L. T.; Sonders, R. C.; Glassman, H. N.; Jordan, D. C.; Luther, R. R.; Tolman, K. G. Effect of Age on the Pharmacokinetics of Orally and Intravenously Administered Terazosin. *Clin. Ther.* **1988**, *10*, 600−607.

(68) Sinko, P. J.; Leesman, G. D.; Amidon, G. L.; Predicting Fraction Dose Absorbed in Humans Using a Macroscopic Mass Balance Approach. *Pharm. Res.* **1991**, *8*, 979−988.

(69) Spyker, D. A.; Rugloski, R. J.; Vann, R. L.; O'Brien, W. M. Pharmacokinetics of Amoxicillin: Dose Dependence After Intravenous, Oral, and Intramuscular Administration. Antimicrob. *Agents Chemother.* **1977**, *11*, 132−141.

(70) Blanchard, J.; Sawers, S. J. The Absolute Bioavailability of Caffeine in Man. Eur. *J. Clin. Pharmacol.* **1983**, *24*, 93−98.

(71) Osman, M. A.; Patel, R. B.; Irwin, D. S.; Craig, W. A.; Welling, P. G. Bioavailability of Chlorothiazide from 50, 100, and 250 mg Solution Doses. *Biopharm. Drug Dispos.* **1982**, *3*, 89−94.

(72) Shah, V. P.; Prasad, V. K.; Cabana, B. E.;Sojka, P. Thiazides II, New Pharmacokinetic Findings on Chlorothiazide Using HPLC Analysis. *Curr. Ther. Res.* **1981**, *29*, 802−814.

(73) Stewart, B. H.; Kugler, A. R.; Thompson, P. R.; Bockbrader, H. N. A Saturable Transport Mechanism in the Intestinal Absorption of Gabapentin is the Underlying Cause of the Lack of Proprtionality Between Increasing Dose and Drug Levels in Plasma. *Pharm. Res.* **1993**, *10*, 276−281.

(74) Chongg, S.; Dando, S. A.; Soucek, K. M.; Morrison, R. A. In Vitro Permeability Through Caco-2 Cells is not Quantitatively Predictive of In Vivo Absorption for Peptide-Like Drugs Absorbed via the Dipeptide Transporter System. *Pharm. Res*. **1996**, *13*, 120−123.

(75) Everette, D. W.; Olsen, S. J.; Sugarman, A. A.; Ita, C.; Stewart, M. B.; Sherman, J. W.; Kripalani, K. L. Disposition of SQ-32,756 (BV-araU) in Healthy Male Subjects. *Pharm. Res*. **1992**, *9*, S316.

(76) Yu, L. X.; Crison, J. P.; Amidon, G. L. Unpublished results.

(77) Tse, F. L.; Jaffe, J. M.; Troendle, A. Pharmacokinetics of Fluvastatin After Single and Multiple Doses in Normal Volunteers. *J. Clin. Pharmacol*. **1992**, *32*, 630−638.

(78) Ranadive, S. A.; Chen, A. X.; Serajuddin, A. T. M. Relative Lipophilicities and Structural-Pharmacological Considerations of Various Angiotensin-Converting Enzyme (Ace) Inhibitors. *Pharm. Res*. **1992**, *9*, 1480−1483.

(79) Lennernäs, H.; Ahrenstedt, Ö.; Ungell, A. L. Intestinal Drug Absorption During Induced Net Water Absorption in Man; A Mechanistic Study Using Antipyrine Atenolol and Enalaprilat. *Br. J. Clin. Pharmacol*. **1994**, *37*, 589−596.

(80) Lennarnäs, H.; Palm, K.; Fagerholm, U.; Artursson, P. Comparison Between Active and Passive Drug Transport in Human Intestinal Epithelial(Caco-2) Cells In Vitro and Human Jejunum In Vivo. *Int. J. Pharm*. **1996**, *127*, 103−107.

(81) Benet, L. Z. Pharmacokinetics/Pharmacodynamics of Furosemide in Man: A Review. *J. Pharmacokinet. Biopharm*. **1979**, *7*, 1−27.

(82) Jacobson, M. A.; de Miranda, P.; Cederberg, D. M.; Burnette, T.; Cobb, E.; Brodie, H.; R.; Mills, J. Human Pharmacokinetic and Tolerance of Oral Ganciclovir. *Antimicrob. Agents Chemother*. **1987**, *31*, 1251−1254.

(83) Ascalone, V.; Cisternino, M.; Sicolo, N.; De Palo, E. Bioavailability of Bromazepam in Man After Single Administration of an Oral Solution. *Arzneimitttelforschung* **1984**, *34*, 96−98.

(84) Raaflaub, J.; Speiser-Courvoisier, J. Zur Pharmakokinetik Von Bromazepam Beim Menschen. *Arzneimitttelforschung* **1974**, *24*, 1841−1844.

(85) Kripalani, K. J.; Mckinstry, D. N.; Singhvi, S. M.; Willard, D. A.; Vukovich, R. A.; Migdalof, B. H. Disposition of Captopril in Normal Subjects. *Clin. Pharmacol. Ther*. **1980**, *27*, 636−641.

(86) Hümpel, M.; Nieuweboer, B.; Milius, W.; Hanke, H.; Wendt, H.; Kinetics and Biotransformation of Lormetazepam. *Clin. Pharmacol. Ther*. **1980**, *28*, 673−679.

(87) Hümpel, M.; Illi, V.; Milius, W.; Wendt, H.; Kurowski, M. The Pharmacokinetics and Biotransformation of the New Benzodiazepine Lormetazepam in Humans. *Eur. J. Drug Metab. Pharmackinet*. **1979**, *4*, 237−243.

(88) Lew, K. H.; Ludwig, E. A.; Milad, M. A.; Donovan, K.; Middleton, E. Jr.; Ferry, J. J.; Jusko, W. J. Gender-Based Effects on Methylprednisolone Pharmacokinetics and Pharmacodynamics. *Clin. Pharmacol. Ther*. **1993**, *54*, 402−414.

(89) Pfeffer, M.; Gaver, R. C.; Ximenez, J. Human Intravenous Pharmacokinetics and Absolute Oral Bioavailability of Cefatrizine. *Antimicrob. Agents Chemother*. **1983**, *24*, 915−920.

(90) Pentikäinen, P. J.; Penttilä, A.; Neuvonen, P. J.; Gothoni, G. Rate of [14C]-Bumetanide in Man. *Br. J. Clin. Pharmacol*. **1977**, *4*, 39−44.

(91) O'Callaghan, C. H.; Sykes, R. B.; Ryan, D. M.; Foord, F. D.; Muggleton, P. W. Cefuroxime-A New Cephalosporin Antibiotic. *J. Antibiot. (Tokyo)* **1975**, *29*, 29−37.

(92) Hellström, K.; Rosén, A.; Swahn, A. Absorption and Decomposition of Potassium-35S-Phenoxymethyl Penicillin. *Clin. Pharmacol. Ther*. **1974**, *16*, 826.

(93) Beermann, B.; Groschinsky-Grind, M. Pharmacokinetics of Hydrochlorothiazide in Man. *Eur. J. Clin. Pharmacol*. **1977**, *12*, 297−303.

(94) Jamali, F.; Brocks, D. R. Clinical Pharmacokinetics of Ketoprofen and its Enantiomers. *Clin. Pharmacokinet*. **1990**, *19*, 197−217.

(95) Garg, V.; Jusko, W. J. Biopharm. Bioavailability and Reversible Metabolism of Prednisone and Prednisolone in Man. *Drug Dispos*. **1994**, *15*, 163−172.

(96) Thwaites, D. T.; Cavet, M.; Hirst, B. H.; Simmons, N. L. Angiotensin-Converting Enzyme (Ace) Inhibitor Transport in Human Intestinal Epithelial (Caco-2) Cells. *Br. J. Pharmacol*. **1995**, *114*, 981−986.

(97) Armayor, G. M.; Lopez, L. M. Lisinopril: A New Angiotensin-Converting Enzyme Inhibitor. *Drug Intell. Clin. Pharm*. **1988**, *22*, 365−372.

(98) Jung, D.; Powell, J. R.; Walson, P.; Perrier, D. Effect of Dose on Phenytoin Absorption. *Clin. Pharmacol. Ther*. **1980**, *28*, 479−485.

(99) Singhvi, S. M.; Pan, H. Y.; Morrison, R. A.; Willard, D. A. Disposition of Pravastatin Sodium, A Tissue-Selective HMG−Coa Reductase Inhibitor, in Healthy Subjects. *Br. J. Clin. Pharmacol*. **1990**, *29*, 239−243.

(100) Schedl, H. P. Absorption of Steroid Hormones from the Human Small Intestine. *J. Clin. Endocrinol. Metab*. **1965**, *25*, 1309−1316.

(101) Hu, M.; Chen, J.; Zhu, Y.; Dantzig, A. H.; Stratford, R. E., Jr.; Kuhfeld, M. T. Mechanism and Kinetics of Transcellular Transport of a New Beta Lactam Antibiotic Loracarbef Across an Intestinal Epithelial Membrane Model System (Caco-2). *Pharm. Res*. **1994**, *11*, 1405−1413.

(102) DeSante, K. A.; Zeckel, M. L. Pharmacokinetic Profile of Loracarbef. *Am. J. Med*. **1992**, *93*, 165−195.

(103) Walker, M. C.; Patsalos, P. N. Clinical Pharmacokinetics of New Antiepileptic Drugs. *Pharmacol. Ther*. **1995**, *67*, 351−384.

(104) Elwes, R. D.; Binnie, C. D. Clinical Pharmacokinetics of Newer Antiepileptic Drugs. *Clin. Pharmacokinet*. **1996**, *30*, 403−415.

(105) Balasubramanian, R.; Klein, K. B.; Pittman, A. W.; Liao, S. H.; Findlay, J. W.; Frosolone, M. F. Pharmacokinetics of Acrivastine after Oral and Colonic Administration. *J. Clin. Pharmacol*. **1989**, *29*, 444−447.

(106) Schroeder, D. H. Metabolism and Kinetics of Bupropion. *J. Clin. Psychiatry* **1983**, *44*, 79−81.

(107) Goodnick, P. J. Pharmacokinetics of Second Generation Antidepressants: Bupropion. *Psychopharmacol. Bull*. **1991**, *27*, 513−519.

(108) Johansson, R.; Regårdh, C. G.; Sjögren, J. Absorption of Alprenolol in Man from Tablets with Different Rates of Release. *Acta Pharm. Suec*. **1971**, *8*, 59−70.

(109) Holford, N. H. Clinical Pharmacokinetics and Pharmacodynamics of Warfarin. *Clin. Pharmacokinet*. **1986**, *11*, 483−504.

(110) Edgar, B.; Regardh, C. G.; Johnsson, G.; Johansson, L.; Lundborg, P.; Löfberg, I.; Rönn, O. Felodipine Kinetics in Healthy Men. *Clin. Pharmacol. Ther*. **1985**, *38*, 205−211.

(111) Edgar, B.; Regardh, C. G.; Lundborg, P.; Romare, S.; Nyberg, G.; Rönn, O. Pharmacokinetics and Pharmacodynamics Studies of Felodipine in Healthy Subjects After Various Single Oral and Intravenous Doses. *Biopharm. Drug Dispos*. **1987**, *8*, 235−248.

(112) Frishman, W. H.; Tepper, D.; Lazar, E. J.; Behrman, D.; Betaxolol: A New Long-Acting Betal-Selective Adrenergic Blocker. *J. Clin. Pharmacol*. **1990**, *30*, 686−92.

(113) Schwartz, D. E.; Ziegler, W. H. Assay and Pharmacokinetics of Trimethoprim in Man and Animals. *Postgrad. Med. J*. **1969**, *45*, Suppl:32−37.

(114) Finn, A.; Straughn, A.; Meyer, M.; Chubb, J. Effect of Dose and Food on the Bioavailability of Cefuroxime Axetil. *Biopharm. Drug Dispos*. **1990**, *8*, 519−526.

(115) Lang, C. C.; Moreland, T. A.; Davey, P. G. Bioavailability of Cefuroxime Axetil: Comparison of Standard and Abbreviated Methods. *J. Antimicrob. Chemother*. **1990**, *25*, 645−650.

(116) Glazko, a. J.; Wolf, L. M.; Dill, W. A.; Bratton, C. Biochemical Studies on Chloramphenicol (Chloromectin). *J. Pharmacol. Exp. Ther*. **1949**, *96*, 445−459.

(117) Glazko, A. J.; Dill, W. A.; Kazenko, A.; Wolf, L. M.; Carnes, H. E. Physical Factors Affecting the Rate of Absorption of Chloramphenicol Esters. *Antibiot. Chemother*. **1958**, *8*, 516−526.

(118) Ambrose, P. J. Clinical Pharmacokinetics of Chloramphenicol and Chloramphenicol Succinate. *Clin. Pharmacokinet*. **1984**, *9*, 222−238.

(119) Davies, D. S.; Wing, A. M.; Reid, J. L.; Neill, D. M.; Tippett, P.; Dollery, C. T. Pharmacokinetics and Concentration Effect Relationships of Intravenous and Oral Clonidine. *Clin. Pharmacol. Ther*. **1977**, *21*, 593−601.

(120) Lowenthal, D. T.; Matzek, K. M.; MacGregor, T. R. Clinical Pharmacokinetics of Clonidine. *Clin. Pharmacokinet*. **1988**, *14*, 287−310.

(121) Lowenthal, D. T. Pharmacokinetics of Clonidine. *J. Cardiovasc. Pharmacol*. **1980**, *2*, S29−S37.

(122) Walker, S. R.; Richards, A. J.; Paterson, J. W. The Absorption, Excretion and Metabolism of Disodium [14C] Cromogylate in Man. *Biochem. J*. **1971**, *125*, 27P.

(123) Sallee, F. R.; Pollock, B. G. Clinical Pharmacokinetics of Imipramine and Desipramine. *Clin. Pharmacokinet*. **1990**, *18*, 346−64.

(124) Mandelli, M.; Tognoni, G.; Garattini, S. Clinical Pharmacokinetics of Diazepam. *Clin. Pharmacokinet*. **1978**, *3*, 72−91.

(125) Jauhiainen, K.; Eksborg, S.; Kangas, L.; Perilä, M.; Alfthan, O. The Absorption of Doxorubicin and Mitomycin C in Perioperative Instillation. *Eur. Urol*. **1985**, *11*, 269−272.

(126) Fabris, A.; Pellanda, M. V.; Gardin, C.; Contestabile, A.; Bolzonella, R. Pharmacokinetics of Antifungal Agents. *Perit. Dial. Int*. **1993**, *13*, S280−S282.

(127) Finland, M. Gentamicin: Antibacterial Activity, Clinical Pharmacology and Clinical Applications. *Med. Times* **1969**, *97*, 161−174.

(128) Lee, E. J. D.; Williams, K.; Day, R.; Graham, G.; Champion, D. Stereoselective Disposition of Ibuprofen Enantiomers in Man. *Br. J. Clin. Pharmacol*. **1985**, *19*, 669−674.

(129) Regazzi, B. M.; Rondanelli, R.; Ciaroelli, L.; Rampini, A. Evaluation of the Absorption from Three Ibuprofen Formulations. *Int. J. Clin. Pharmacol. Res*. **1986**, *6*, 469−473.

(130) Chabner, B. A.; Donehower, R. C.; Schilsky, R. L. Clinical Pharmacology of Methotrexate. *Cancer Treat. Rep.* **1981**, *65*, 51−54.
(131) Bleyer, W. A. The Clinical Pharmacology of Methotrexate. *Cancer* **1978**, *41*, 36−51.
(132) Henderson, E. S.; Adamson, R. H.; Oliverio, V. T. The Metabolic Fate of Titrated Methotrexate. *Cancer Res.* **1965**, *25*, 1018−1024.
(133) Delmonte, L.; Jukes, T. H. Folic Acid Antagonists in Cancer Chemotherapy. *Pharmacol. Rev.* **1962**, *14*, 91−135.
(134) Freeman, M. V. The Fluorometric Measurement of the Absorption, Distribution and Excretion of Single Doses of 4-Amino-10-Methyl Pteroylglutamic Acid (Amethopterin) in Man. *J. Pharmacol. Exp. Ther.* **1956**, *15*, 154−62.
(135) Stein, G. E. Review of the Bioavailability and Pharmacokinetics of Oral Norfloxacin. *Am. J. Med.* **1987**, *82*, 18−21.
(136) Holmes, B.; Brogden, R. N.; Richards, D. M. Norfloxacin: A Review of its Antibacterial Activity, Pharmacokinetic Properties and Therapeutic Use. *Drugs* **1985**, *30*, 482−513.
(137) Fourtillan, J. B.; Brisson, A. M.; Couet, W. Pharmacokinetics of Prazosin Administered as Gastro-Intestinal-Therapeutic-Systems to 24 Healthy Volunteers. *Therapie* **1993**, *48*, 115−118.
(138) Guentert, T. W.; Upton, R. A.; Holford, N. H.; Bostrom, A.; Rigelman, S. Gastrointestinal Absorption of Quinidine from Some Solutions and Commercial Tablets. *J. Pharmacokinet. Biopharm.* **1980**, *8*, 243−255.
(139) Beermann, B.; Hellström, K.; Rosén, A. The Gastrointestinal Absorption of Atropine in Man. *Clin. Sci.* **1971**, *40*, 95−106.
(140) Golding, J. F.; Gosden, E.; Gerrell, J. Scopolamine Blood Levels Following Buccal Versus Ingested Tablets. *Aviat. Space Environ. Med.* **1991**, *62*, 521−526.
(141) Gardner, M. J.; Wilner, K. D.; Hansen, R. A.; Fouda, H. G.; McMahon, G. F. Single and Multiple Dose Pharmacokinetics of Tenidap Sodium in Healthy Subjects. *Br. J. Clin. Pharmacol.* **1995**, *39*, 11S−15S.
(142) Sonders, R. C. Pharmacokinetics of Terazosin. *Am. J. Med.* **1986**, *80*, 20−24.
(143) Klotz, U.; Antonin, K. H. Pharmacokinetics and Bioavailability of Sodium Valproate. *Clin. Pharmacol. Ther.* **1976**, *21*, 736−743.
(144) Bruni, J.; Wilder, B. J. Valproic Acid: Review of New Antiepileptic Drug. *Arch. Neurol.* **1979**, *36*, 393.
(145) Gugler, R.; von Unruh, G. E. Clinical Pharmacokinetics of Valproic Acid. *Clin. Pharmacokinet.* **1980**, *5*, 67−83.
(146) Clark, P. I.; Slevin, M. L. The Clinical Pharmacology of Etoposide and Teniposide. *Clin. Pharmacokinet.* **1987**, *12*, 223−252.
(147) Nissen, N. I.; Hansen, H. H.; Pedersen, H.; Stroyer, I.; Dombernowsky, P.; Hessellund, M. Clinical Trial of the Oral Form of a New Podophyllotoxin Derivative, VP-16-213 (NSC-141540), in Patients with Advanced Neoplastic Disease. *Cancer Chemother. Rep.* **1975**, *59*, 1027−1029.
(148) MDL Information Systems, Inc. 14600 Catalina Street, San Leandro, CA 94577.
(149) Kier, L. B. Shape Indexes of Orders One and Three from Molecular Graphs. *Quant. Struct.-Act. Relat.* **1986**, *5*, 1−7.
(150) Randić, M.; Brissey, G. M.; Spencer, R. B.; Wilkins, C. L. Search for all Self-Avoiding Paths for Molecular Graphs. *Comput. Chem.* **1979**, *3*, 5.
(151) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17.
(152) Kier, L. B.; Hall, L. H. Molecular Connectivity in Structure-Activity Analysis; Research Studies: Hertfordshire, England, 1986.
(153) Randić, M. On Molecular Identification Numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164.
(154) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P., AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(155) Goldstein, H. *Classical Mechanics*; Addison-Wesley: Reading, MA, 1950; pp 144−156.
(156) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441.
(157) Pearlman, R. S. In *Physical Chemistry Properties of Drugs*; Sinkula, A. A., Valvani, S. C., Eds.; Quantum Chemistry Program Exchange No. 413. Marcel Dekker, Inc.: New York, 1980; Chapter 10.
(158) Stouch, T. R.; Jurs, P. C. A Simple Method for the Representation, Quantification, and Comparison of the Volumes and Shapes of Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 4−12.
(159) Stanton, D. T.; Jurs, P. C. Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure−Property Relationship Studies. *Anal. Chem.* **1990**, *62*, 2323−2329.
(160) Sello, G. A New Definition of Functional Groups and a General Procedure for Their Identification in Organic Structures. *J. Am. Chem. Soc.* **1992**, *114*, 3306.
(161) Bohachevsky, I. O.; Johnson, M. E.; Stein, M. L. Generalized Simulated Annealing for Function Optimization. *Technometrics* **1986**, *28*, 209.
(162) Sutter, J. M.; Dixon, S. L.; Jurs, P. C. Automated Descriptor Selection for Quantitative Structure-Activity Relationships Using Generalized Simulated Annealing. *J. Chem. Int. Comput. Sci.* **1995**, *35*, 77.
(163) Leardi, R. Application of a Genetic Algorithm to Feature Selection Under Fall Validation Conditions and to Outlier Detection. *J. Chemom.* **1992**, *6*, 267.
(164) Wessel, M. D.; Jurs, P. C. Prediction of Reduced Ion Mobility Constants from Structural Information Using Multiple Linear regression Analysis and Computational Neural Networks. *Anal. Chem.* **1994**, *66*, 2480.
(165) *Computer-Assisted Development of Quantitative Structure−Property Relationships and Design of Feature Selection Routines*; Wessel, M. D. Ph.D. Dissertation, The Pennsylvania State University, 1997; Chapter 3.
(166) Belsley, D. A.; Kuh, E.; Welsch, R. E. *Regression Diagnostics*; Wiley: New York, 1980.
(167) Livingstone, D. J.; Manallack, D. T. Statistics Using Neural Networks: Chance Effects. *J. Med. Chem.* **1993**, *36*, 1295.
(168) Sutter, J. M.; Jurs P. C. Selection of Molecular Descriptors for Quantitative Structure−Activity Relationships. In *Data Handling in Science and Technology (Vol. 15) Adaption of Simulated Annealing to Chemical Optimization Problems*; Kalivas, J. H., Ed.; Elsevier Science B. V.: Amsterdam, 1995; Chapter 5.
(169) The six descriptors in the MOPAC independent model are as follows: 1. valence corrected chain (ring) size six molecular connectivity index;[152] 2. number of sixth order path-clusters appearing in molecule;[152] 3. number of aromatic bonds; 4. sum of all path weights starting from oxygen atoms;[153] 5. geometric moment in *X* axis (including H atoms); 6. two-dimensional projection (shadow area) on XZ plane.[158] This model employs a 6-4-1 neural network architecture. The rms error of prediction for this model is between 20 and 25%HIA units for various neural network training sessions.
(170) Andrea, T. A.; Kalayeh, H. Applications of Neural Networks in Quantitative Structure−Activity Relationships of Dihydrofolate Reductase Inhibitors. *J. Med. Chem.* **1991**, *34*, 2824−2836.
(171) So, S.-S.; Richards, W. G. Application of Neural Networks: Quantitative Structure−Activity Relationships of the Derivatives of 2,4-Diamino-5-(substituted-benzyl)pyrmidines as DHFR Inhibitors. *J. Med. Chem.* **1992**, *35*, 3201−3207.
(172) Suzuki, T.; Ishida, M. Application of Neural Networks in Quantitative Structure−Activity Relationships of Cytotoxicity of Antioxidants. *Pharm. Sci.* **1995**, *1*, 297−300.
(173) Bishop, C. M. *Neural Networks for Pattern Recognition*; Oxford University Press: New York, 1995.
(174) Topliss, J. G.; Edwards, R. P. Chance Factors in Studies of Quantitative Structure−Activity Relationships. *J. Med. Chem.* **1979**, *22*, 1238−1244.
(175) Sutter, J. M.; Jurs, P. C. Prediction of Aqueous Solubility for a Diverse Set of Heteroatom-Containing Organic Compounds Using a Quantitative Structure−Property Relationship. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 100−107.
(176) Dixon, S. L.; Jurs, P. C. Estimation of pK_a for Organic Oxyacids using Calculated Atomic Charges. *J. Comput. Chem.* **1993**, *14*, 1460−1467.

CI980029A