JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 9, Number 6, 2002

Pp. 849–864

Predicting CNS Permeability of Drug Molecules: Comparison of Neural Network and Support Vector Machine Algorithms

SCOTT DONIGER,[1] THOMAS HOFMANN,[2] and JOANNE YEH[1,3]

ABSTRACT

Two different machine-learning algorithms have been used to predict the blood–brain barrier permeability of different classes of molecules, to develop a method to predict the ability of drug compounds to penetrate the CNS. The first algorithm is based on a multilayer perceptron neural network and the second algorithm uses a support vector machine. Both algorithms are trained on an identical data set consisting of 179 CNS active molecules and 145 CNS inactive molecules. The training parameters include molecular weight, lipophilicity, hydrogen bonding, and other variables that govern the ability of a molecule to diffuse through a membrane. The results show that the support vector machine outperforms the neural network. Based on over 30 different validation sets, the SVM can predict up to 96% of the molecules correctly, averaging 81.5% over 30 test sets, which comprised of equal numbers of CNS positive and negative molecules. This is quite favorable when compared with the neural network's average performance of 75.7% with the same 30 test sets. The results of the SVM algorithm are very encouraging and suggest that a classification tool like this one will prove to be a valuable prediction approach.

Key words: neural net, support vector machine, machine-learning algorithms, blood brain barrier, central nervous system, predictive methods, kernel methods.

INTRODUCTION

PREDICTING THE ABILITY OF A MOLECULE to enter the central nervous system (CNS) through the blood–brain barrier (BBB) would be an extremely useful tool for designing drug compounds. Designing drugs for targets in the CNS is a difficult task because of the presence of the blood–brain barrier. The BBB is a selective membrane that prevents small molecules from entering the CNS, making drugs that are effective in other parts of the body virtually useless for CNS targets. As a result, bacterial and viral infections can harbor in the CNS, making it difficult to fully eliminate the infection through conventional antibiotic therapies. As a preliminary step to designing more useful drugs that can act on targets in the

[1]Brown University, MCB Department, Box G-J2, Providence, RI 02912.
[2]Brown University, Department of Computer Science, Box 1910, Providence, RI 02912.
[3]Brown University, Department of Chemistry, Providence, RI 02912.

CNS, we are developing a methodology for predicting the permeability and, ultimately, the bioavailability of different classes of molecules into the CNS. While there are in vitro assays to measure the log blood–brain permeation coefficient, these methods are expensive, time consuming, and not very practical when screening large libraries of potential molecules. Hence, the ability to predict blood–brain barrier permeability will be an enormous help in designing drugs that target the CNS.

Most attempts to predict BBB transport have varying degrees of success. Several attempts have been made to correlate the octanol-water partitioning coefficient, log P, with BBB permeability. Pardridge (1998) has shown that log P is well correlated with BBB penetration for molecules below 400 Da. van de Waterbeemd *et al.* (1998) used molecular size, shape, and hydrogen bonding characteristics as descriptors of BBB permeability. They were able to identify a correlation between the molecular size and polar surface area of CNS active and inactive compounds. van de Waterbeemd does not give any form of an equation for predicting BBB permeability, but rather gives guidelines on general properties that make a molecule CNS active. Another prediction method created by Crivori *et al.* (2000) uses three-dimensional structure analysis of small molecules to generate a model of BBB permeability. Crivori used a computer modeling program to transform 3D fields into a descriptor set of the molecules. Through a principal component analysis and partial least squares discriminator analysis they were able to generalize the data and use the results to predict the BBB permeability of novel compounds. Their prediction method was successful in predicting over 90% of unseen CNS active molecules correctly, but only 65% of the CNS inactive molecules correctly. Luco (1999) also used statistical analysis of structural-based descriptors and this study had a success rate similar to Crivori's method. Of the 25 molecules tested, Luco was able to predict 100% of the CNS active molecules correctly and 84.6% of the CNS negative molecules.

In this study, the BBB prediction algorithm is based on physical and structural descriptors. A training set of 324 molecules was used to train a multilayered backpropagation neural network and a support vector machine to predict the molecule's ability to enter the CNS. Neural networks have been used successfully by Ajay *et al.* (1998) to distinguish between drugs and nondrugs, as well to identify CNS active compounds (Ajay *et al.*, 1999). Although the neural network approach may be promising if a large enough database is used for training, this study suggests that support vector machines are capable of outperforming neural networks, particularly in situations utilizing smaller datasets.

MATERIALS AND METHODS

The drug database

The database is comprised of 324 drugs and biologically active molecules that have been accumulated from several sources to make up the training set (Appendix A). These molecules were taken from previous papers discussing BBB transport, primarily Fischer *et al.* (1998), Ajay *et al.* (1999), and van de Waterbeemd *et al.* (1998). These resources did not provide a sufficient number of molecules for training, so additional molecules were identified from the psychotropic database (Lundbeck *et al.*, 2000), the Physicians' Desk Reference, and the National Library of Medicine's Medline Plus Health Information website. The CNS activity of each molecule was determined by one of three general classification factors: previously published results, where the molecule was considered to be CNS active or inactive; whether or not the molecule was listed in the psychotropic database, where all listed molecules are considered to be CNS active; and finally, the medical use, mechanism of action, and contraindications of the molecule.

The chemical and physical descriptors of the molecules were obtained from the NCI structural and physical properties database produced by ChemSW.

The neural network

The neural network algorithms have been implemented using the Matlab Neural Net Toolbox. This software package provides the functions necessary to build and train a backpropagation network. The Neural Net Toolbox contains several training algorithms that are all variations on the general algorithm described above. In this study, the resilient backpropagation algorithm was used to reduce the time necessary to train the network. This algorithm works on the same principles as the gradient descent algorithm; however,

instead of using the product of the activation and the error of the node, it uses only the sign of this value to determine the change in the weight.

The backpropagation network uses a sigmoidal squashing function to provide a continuous activation function. The nature of this function is that its slope approaches zero at the extreme values, so the product of the derivative of the activation and the error is often a very small value. This leads to very small changes in the weight, resulting in very slow training. The resilient backpropagation algorithm uses a separate parameter to determine the size of the step it takes when calculating a new weight. This parameter is slightly increased each time the derivative of the performance function has the same sign for two successive iterations with respect to the weight. The parameter is decreased when the sign of the derivative changes from the previous iteration. This modification from the original gradient descent algorithm reduced the time necessary to train the network by at least an order of magnitude.

The support vector machine

The SVM used in this study was implemented using the SVM Light package available from Thorsten Joachims (*www.ais.gmd.de/~throsten/svm_light*). This software package implements a classification algorithm, which is based on the soft margin algorithm. The SVM Light package provides four different kernel methods: a linear kernel, a polynomial kernel, a radial basis function kernel, and a sigmoidal kernel. In addition to choosing the kernel function, the architecture of the SVM can also be modified by a training parameter (called C) that sets the tradeoff between training error and the margin size. The kernel function and the specificity of the training parameter proved to be the most significant factors when optimizing the SVM algorithm (see results below).

Training times for the SVM varied significantly with the kernel function. The radial basis function (RBF) was the fastest, with training taking less than 0.5 seconds of CPU time on a Sun UltraSparc. The quadratic kernel was significantly slower, taking 350 seconds on average for training with a C value of 5. The SVM using the RBF kernel carries out the training significantly faster than the NN, which had training times that were at least 100 times slower for the identical training sets.

Measuring the predictive performance of the algorithms

For both the NN and the SVM, the performance of the algorithm was measured by counting the number of molecules in the validation set that were correctly classified. One of the standard methods for evaluating machine learning algorithms is the cross validation method, where the data set is split into three equally sized groups and then the training is carried out on two thirds of the data points and the remaining third is used for validation. In this study, the data set is quite small compared to other machine learning problems, so cross validation makes the training set too small to well represent the problem.

A different method for validation has been used here that is based on a "bootstrapping approach," as follows. The validation set is made up of 50 molecules: 25 CNS active and 25 CNS inactive molecules are selected at random from the complete data set. To get an accurate measure of the algorithm's predictive performance, 30 different validation sets were used. The NN or SVM was trained independently for each of the 30 validation sets, and the average performance over all 30 validation sets is taken as the predictive ability of the algorithm.

RESULTS

Designing the neural network

To optimize the performance of the network, several parameters were modified during training of the network including the number of hidden units and the stopping error of the training algorithm. One of the complications in using neural networks is that there is no way to predetermine what the optimal values for these parameters should be. Each data set is unique and each network is unique, so there are no rules defining the optimal level of training or the optimal number of hidden units. There are several common heuristics that can be followed to reach a point where the network's performance is "good enough," even if it is not completely optimized.

In this study, a descriptor set of nine variables is used as inputs to the neural network. The input set consists of molecular weight, molecular volume, surface area, the percent of the surface area that is hydrophilic, the log P (octanol/water partitioning coefficient), the number of hydrogen bond donors, the number of hydrogen bond acceptors, the hydrophilic/lipophilic balance, and a three-dimensional representation of the number of hydrogen bonds. These variables were decided on based on the parameters previously determined to be important in BBB transport (Pardridge, 1998; Crivori *et al.*, 2000; Fischer *et al.*, 1998) as well as on the information available from the ChemSW database. Passive diffusion is the primary method of transport looked at in this study, and each of these variables is important in determining the ability of a molecule to diffuse through a lipid bilayer.

The performance of the network suggests that these input parameters are adequate for predicting the CNS active molecules, but that additional descriptors may be necessary for describing the CNS inactive molecules. This is most likely due to the presence of the efflux proteins, which remove many molecules that are capable of penetrating the BBB.

The values obtained from the ChemSW database are normalized to a mean of zero and a unit standard deviation before being used to train the network. Using the normalized data is advantageous because it prevents one input from dominating the training process. In this data set, the molecular weight has a much larger variance than the log P or hydrogen bonding values, so if the actual data values are used, the network will be heavily weighted for the molecular weight. This could lead the network to view molecular weight as a more important input variable than it actually is, skewing the results.

The number of hidden units in the network is generally thought to be an important variable because the hidden units act as the pattern identifiers during the training. A general heuristic is that the number of hidden units should be equal to the number of patterns expected in the data set. The number of hidden units should also be less than the number of data points in the training set. As the number of hidden units approaches the number of data points in the training set, the network becomes merely a lookup table, rather than a network that can generalize from the inputs.

For this specific network, there does not appear to be one optimal value for the number of hidden units. Our results show that performance increases with the number of hidden units, up to 60 hidden units, and then the performance remains consistently in the high 70% range (Fig. 1). In addition to these results, many other tests have been run and results have been highly variable, such that for any given data split there is a different optimal number of hidden units. For this study, the results are given for hidden unit values of 45 and 60. Networks of this size have consistently been some of the best performers and should provide an adequate representation of the predictive powers of the network. Continuing above 60 hidden units was not necessary because there is only a slight increase in performance, and this increase did not merit the substantial increase in training times.

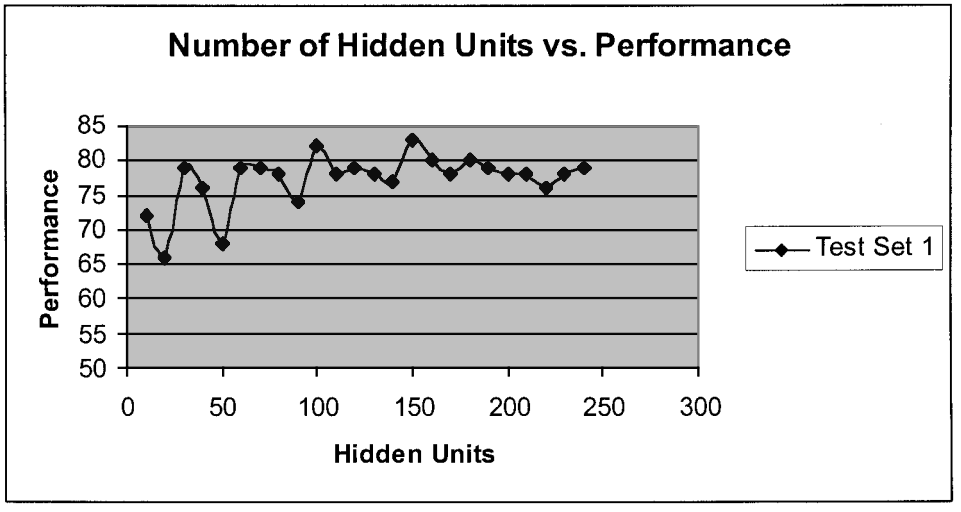

FIG. 1. The number of hidden units versus the performance of the NN. From this graph, it can be seen that the performance of the network increases as the number of nodes increases, but then levels off above 60 hidden units.

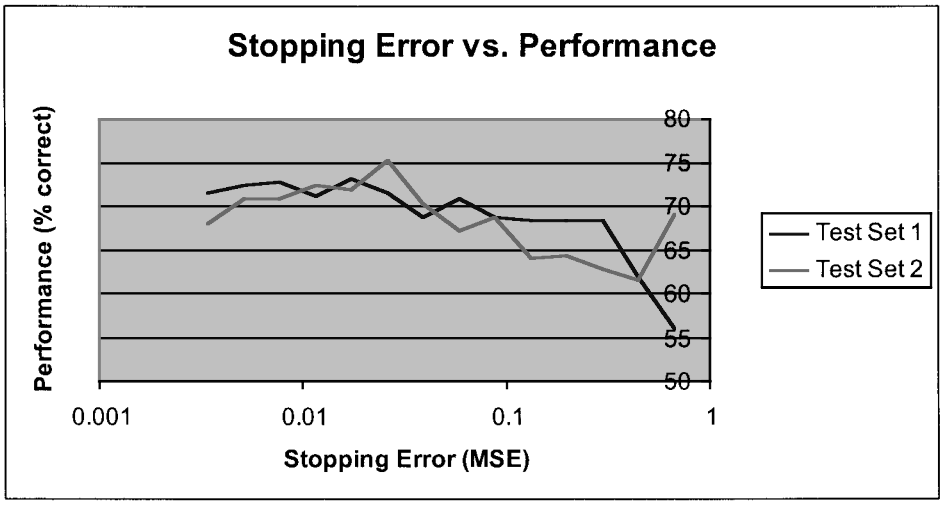

FIG. 2. This graph shows how the performance of the network varies with the stopping error. The best perforance of test set 1 occurred with a MSE of 0.017. Test set 2 had the highest performance at a MSE of 0.026. From the results of this graph, a target MSE of 0.02 was used during network training.

The stopping error, or the point where the network has trained enough, is determined by an error-checking procedure known as early stopping. In early stopping, the data set is split into three sets, the training set, the validation set, and the testing set. As the network is trained, the performance of the network is measured on the validation set. When the error on the validation set begins to increase, which indicates that the network has begun to overfit the data, the training stops. The validation set is then added back into the training set and the network is trained briefly with the combined training set, so that the validation can also be used to generalize for the testing set.

Implementing this early-stopping algorithm actually decreased the performance of the network by an average of nearly 3.5%. This result was very unexpected, as early stopping should typically lead to an improved performance on the test set. One possible explanation for this result is that because of the small size of the data set, removing the validation set from the training set makes the training set too small to properly train the network. Removing an additional 50 molecules for the validation set leaves only 274 molecules to train the network. The validation set is added back into the training set once the early stopping has occurred and the network is briefly trained on this larger training set. It is possible that because the network only sees these additional 50 molecules following the early stopping, this does not provide sufficient training time to fully incorporate these molecules into the weighting of the network.

In order to avoid using a smaller training set, a different form of early stopping that does not split the training set has been implemented. In this method the stopping error is predetermined before the training begins. Using decreasing stopping errors, it was possible to determine the optimal stopping point for the network. The network's performance quickly improves as the mean squared error of the training set decreases, but then begins to plateau and then begin a slow descents as the MSE decreases further (Fig. 2). These results agree with the theory behind the early stopping algorithm, and the peak of this graph was used as the target error when training the network.

The performance of the neural network

Two different networks have been used to measure the predictive ability of the neural network. The networks use the normalized data, a target-training mean squared error of 0.2, and either 30 or 45 hidden units. Each network has one output node, which uses a tangent sigmoidal function to confine the output values between -1 (CNS inactive) and 1 (CNS active). Over 30 different test sets (each test set consists of 25 CNS+ and 25 CNS$-$ molecules selected at random from the data set and are withheld from training) the neural network correctly predicted 75.7% of the molecules when using 30 hidden nodes and 75.0% of the molecules when using 45 hidden nodes. The network correctly classified 81.5% of the CNS active molecules and 69.9% of the inactive molecules.

TABLE 1. A COMPARISON OF THE DIFFERENT KERNEL METHODS AND TRAINING ERRORS

Performance of SVM with specified kernel and C value				
Kernel	*C = 0*	*C = 1*	*C = 5*	*C = 10*
Linear	69.6	71.6	74	73.6
Quadratic	70.4	75.2	73.6	75.6
RBF	77.2	78.4	80.4	76.4

Design of the support vector machine

In this approach, three different kernel methods were compared: (i) a linear kernel that essentially used the 275 dimensional input space as the feature space; (ii) a quadratic function that produced a slightly more complicated decision surface by looking at the relationships between pairs of inputs; and (iii) a radial basis function (RBF) which uses a Gaussian equation to map the inputs to an even more complicated feature space. In addition, higher-degree polynomial functions and a sigmoidal function were also tested as possible kernel functions, but they did not outperform any of the kernels described here.

For each of the kernel functions, four different levels of training specificity were compared to see which led to the best performance. The training error is controlled by the parameter C of the learning algorithm. The initial algorithm, with C equal to 0, led to 30 training points being misclassified. As the penalty for misclassifying data increases, the number of misclassified data points decreases so that for C equal to 5, only 3 data points are misclassified on average, and for C equal to 10, this number drops to 2. Table 1 shows how adjusting the training error affected the performance of the three different kernel methods.

From Table 1 it can be seen that the RBF kernel is the best kernel function for classifying this data set. The results also show that using a C value of 5 leads to the optimal level of training and the best performance of this kernel function. A SVM with a RBF kernel function and C set to 5 was therefore used to predict the ability of the drug molecules to penetrate the BBB.

Performance of the support vector machine

The SVM was trained in parallel with the NN system. The total database of molecules consists of 324 molecules, of which 50 molecules (25 CNS+ and 25 CNS−) were selected at random and used as the validation set, leaving 274 molecules to be used for training. Each molecule was represented by the same nine parameters used in the NN training. The number of molecules correctly classified in the validation set is used to measure the performance of the SVM. In order to obtain a more accurate measure of the performance, 30 different splits of the data were used to calculate the average performance of the SVM.

SVM's overall average in correctly classifying both CNS+ and CNS− molecules was 81.5% over the 30 different data splits. The performance ranged from a low of 66% to a high of 96%, with a median and mode value of 82%. When looking at the performance on the CNS+ and CNS− subsets, the SVM correctly classified 82.7% of the CNS+ molecules and 80.2% of the CNS− molecules. The ability to accurately predict not only CNS+ but also CNS− compounds highlights the strength of the SVM over the NN approach.

Pruning the descriptor set

As the level of complexity increases during machine learning, the resulting decision surface becomes more specific to the particular data set and often the level of generalization decreases with increased complexity. In order to determine if the descriptor set used here was leading to overly specific training and poor generalization, training was attempted in the absence of each descriptor to see if the performance would increase. Removing any of the descriptors actually lowered the performance of the SVM (Table 2). The hydrogen-bonding characteristics have the largest impact on the data, lowering the performance by 7.2%. No other parameter had a significant impact on the performance of the SVM. Several smaller combinations of the parameters were used to train the SVM to see if a better training set could be identified. Using just the hydrogen bonding characteristics, the SVM's performance only decreased by 6.2%, again showing the

TABLE 2. THE EFFECT OF REMOVING EACH OF THE PARAMETERS[a]

	Effect of removing descriptors							
Descriptor	*log P*	*MW*	*Vol*	*SA*	*% hydro SA*	*Hbond*	*HLB*	*log P and Hbond*
% change	−0.96	−0.76	−0.24	−0.96	−1.12	−7.2	−0.4	−4

[a]From left to right, log P is the octanol/water partitioning coefficient, followed by the molecular weight, volume, surface area, percent of hydrophillic surface area, hydrogen bond donors/acceptors and 3D hydrogen bonding, and the hydrophilic–lipophilic balance. The final column shows the results of training with log P and the hydrogen bonding characteristics alone.

TABLE 3. COMMON FALSE POSITIVES AND COMMON FALSE NEGATIVES[a]

False negatives		*False positives*	
Methylpentynol	Arecoline	Fluoridine	Clotrimazole
Etazolate	Uridine	Melphalan	Ibuprofen
Haramalol	Meclofexonate	Ethacrynic Acid	Pheniramine
Phenelzine sulfate	Pimpamperone	Hyoscyamine	Phenylbutazone
Benactyzine	Meprobomate	Indomethacin	Chlorambucil
		Mequitazine	Propranolol
		Coumarin	Hydralazine
		Spironolactone	

[a]These molecules either appeared incorrectly classified in multiple testing sets, or were given very large incorrect scores.

TABLE 4. COMMON FALSE POSITIVE AND NEGATIVE MOLECULES[a]

	MW	*Volume*	*SA*	*% hydro*	*log P*	*HLB*	*H acc*	*H donor*	*H 3d*
False Pos	−53.03	−30.1	−4.27	−13.44	1.46	−2.24	−0.6801	−0.42	−4.55
False Neg	−68.94	−46.51	−5.32	18.13	−1.615	3.8	0.2	0.19	1.42

[a]The chart shows that difference between the average values for the 9 descriptors of the false positive/negative molecules and the overall averages.

importance of these parameters. When log P was added back in, so that the SVM was trained with the hydrogen bonding and log P descriptors only, the performance was decreased by just 4%.

A very interesting result is that when log P and hydrogen bonding were removed from the descriptor set, the learning algorithm did not converge, meaning the SVM could not classify the molecules in the absence of these two parameters. This indicates that the primary descriptors being used for classification are the hydrogen bonding descriptors and the log P descriptors. However, the remaining parameters should not be ignored because they do enhance the performance of the SVM.

Analysis of false positive and false negative outputs

To get a better understanding for why 20% of the molecules are being misclassified, the outputs from 10 different test sets have been analyzed. These test sets give a broad selection of the molecules in the database, and through the repetition of some molecules in multiple testing sets, it is possible to identify common themes among the false positive and false negative results. Molecules that were misclassified at least twice, or molecules that were severely misclassified (i.e., a molecule that is CNS−, but the SVM reports a score of +1 or greater) were identified as false positives or negatives. In the 10 test sets, 10 molecules were identified as false negatives and 15 molecules were identified as false positive (Table 3).

The results show that the false positive molecules are significantly smaller, more lipophilic, and have fewer hydrogen bond donors and acceptors (Table 4). This is all consistent with the expected model that CNS positive molecules are small, lipophilic molecules. One of the most obvious classes of molecules

that appear in the false positive set is the antihistamines. Mepiramine and pheniramine are H1 agonists and are classified as CNS inactive, but other antihistamines such as diphenhydramine are classified as CNS active. The structure of pheniramine and diphenhydramine are quite similar and pheniramine actually has a higher log P value and fewer hydrogen bonds, so this would seem to make it more likely to cross the BBB than diphenhydramine. Another molecule that stands out is hyoscyamine. This molecule is an anticholinergic drug, classifying it as CNS inactive. However, the SVM predicted outputs of 1.69 and 1.62 for this molecule, suggesting that it very easily penetrate the BBB. In this study, BBB permeability is equated with CNS activity, and as these results suggest, this simplified method of classification can be somewhat misleading. There are molecules that have the physical ability to cross the BBB, but then do not interact with any specific receptor in the CNS, making them CNS inactive, but capable of penetrating the BBB.

The false-negative compounds are less lipophilic than the average CNS active compounds. What is interesting is that these molecules are significantly smaller than the rest of the CNS active molecules. The smaller size of these molecules can also account for some of the decrease in the log P values. In general, the algorithm misclassifies fewer CNS active molecules than CNS inactive molecules; this could be due to the fact that the CNS active molecules have less variance in the values for the 9 descriptors than the CNS inactive molecules. This can also be attributed to the fact that there is less ambiguity in classifying CNS active molecules than classifying CNS inactive molecules.

Analysis of the outputs of the NN and SVM

The goal of this study is not only to design a method of classifying drug molecules based on the ability to enter the CNS but to also use this information to classify potential drug molecules in drug design screens. Neural networks and support vector machines are binary classifiers capable of separating a complicated data set into two distinct classes. The outputs from these algorithms range from -1 to $+1$ for the NN algorithm and, while for SVM there are no bounds on the output, in general they range between -1.8 and $+1.8$. If these outputs are to be used for designing molecules specifically targeting the CNS, it would be a great advantage if the outputs could be used as a quantitative measure of BBB permeability.

The test sets of the SVM were first searched for molecules with similar structures. Ideally, if the outputs are proportional to BBB permeability, then molecules with similar structure should have similar output. In test set 1, both apomorphine and morphine are present. Apomorphine is synthesized from morphine and these molecules are similar chemically although dissimilar pharmacologically in that they interact with different receptors. However, looking at the results of the SVM, you would not know this. Apomorphine is assigned an output of 0.03, making it only slightly CNS active. Morphine on the other hand is assigned an output of 0.99, making it a definite CNS active molecule, as one would expect. The output for morphine is encouraging, although we continue to work towards enhancing the performance of the algorithm so that apomorphine can be correctly predicted.

Test set 9 contains both temazepam and medazepam and they are assigned values of 0.80 and 0.58, respectively. These molecules are both anxiolytic compounds and are both CNS active. They have similar structures and differ in that temazepam contains a carbonyl and a hydroxyl group that are not present in medazepam. This raises the level of hydrogen bonding in temazepam and lowers its log P value. Based on these facts, it would seem that temazepam should be less able to penetrate the BBB, but the SVM algorithm scores temazepam higher than medazepam.

In another set of test sets (data not shown) clomipramine and imipramine are assigned outputs of 1.72 and 1.03, respectively. These molecules are both tricyclic antidepressants and differ only in the addition of a chlorine atom in clomipramine. The two molecules have identical hydrogen bonding characteristics and their log P values differ by only 0.5. It is difficult to say if clomipramine is really 70% more able to cross the BBB, but it encouraging that two similar obvious CNS active structures both received high positive outputs. Clomipramine appears in 4 of the 10 test sets and is assigned values of 0.77, 0.92, 1.65, and 1.72. These values show that across different training sets the actual values of the outputs cannot be simply compared because of variations within the training set which leads to different weights and, consequently, varying the outputs.

The important question is how do these outputs for clomipramine compare to the rest of the testing set. Clomipramine is consistently one of the highest outputs of the testing sets. In the two test sets where clomipramine was assigned values below 1, none of the molecules in the test set were assigned values

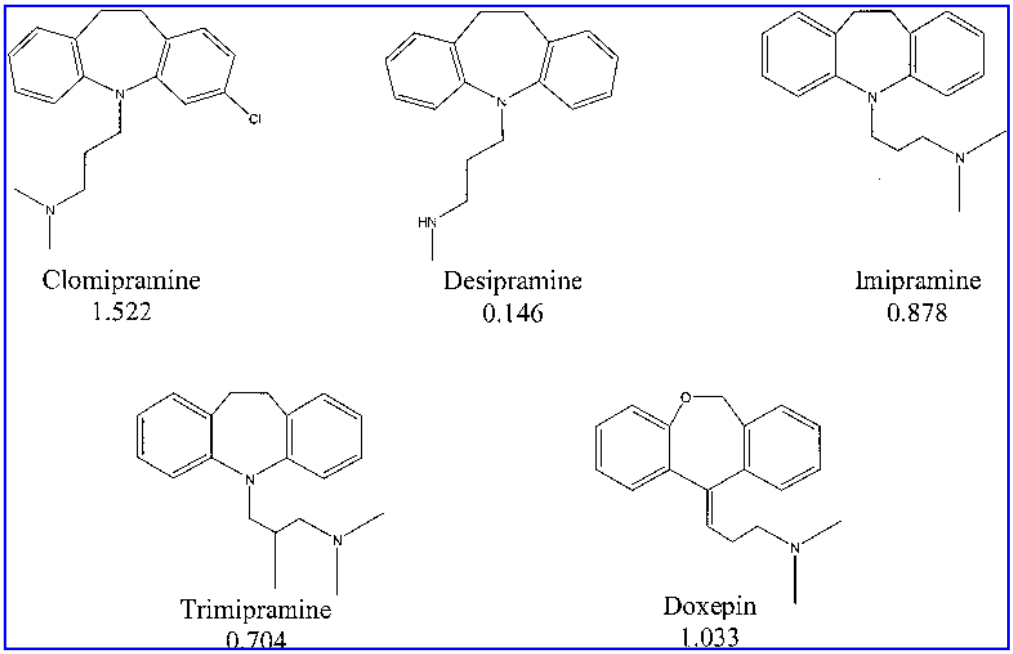

FIG. 3. The tricyclic antidepressants. The molecule and its structure are shown along with the SVM algorithm's predicted BBB permeability. All tricyclic antidepressants are CNS+.

above 1, showing that relative to the other molecules in the test set clomipramine can consistently be considered to easily cross the BBB. Two additional test sets have been created to test the ability of the algorithm to rank closely related molecules. The first test set contains the tricyclic antidepressants, clompiramine, imipramine, desipramine, trimipramine, and doxepin. The second test set contains codeine, hydrocodone, and oxycodone, three narcotic analgesics with related structures. The SVM was trained using all of the remaining molecules in the data set (319 molecules and 321 molecules, respectively), so these test sets can be thought of as novel molecules, simulating a design situation.

The results from the tricyclic compounds are encouraging (see Fig. 3). First, all five molecules are correctly classified as CNS+ molecules, which shows that the SVM is effective in binary classification. Second, the molecule with the highest hydrogen bonding has the lowest predicted BBB permeability. This result is consistent with the expected negative correlation between hydrogen bonding and BBB permeability. Of the four remaining compounds, clomipramine receives the highest score and trimipramine the lowest. This result is somewhat surprising. One would expect that the polar nature of the chlorine atom in clomipramine would reduce its BBB permeability and the additional methyl group in trimipramine would raise its BBB permeability. Unfortunately, no data is currently available on the actual logBBB (blood–brain barrier partitioning coefficient) of these molecules, so that these results predicted by the algorithm remain to be confirmed.

The results of the narcotic analgesics are also encouraging (see Fig. 4) because again the molecule with the most hydrogen bonding, oxycodone, has the lowest predicted BBB permeability. It is interesting that hydrocodone has a predicted BBB permeability that is half of codeine. This suggests that hydrogen bond acceptors may hinder crossing the BBB more than hydrogen bond donors do. Again, there is a lack of experimentally determined values of BBB permeability, and the predicted results remain to be experimentally confirmed.

The results from comparing the various test sets and from the smaller specific test sets suggest that the outputs given by the SVM can be viewed as preliminary predictions for BBB permeability. As the exact values of predictions are only meaningful within the context of the specific test set, the output for any particular compound cannot be taken as a quantitative value. Values generated between test sets cannot be quantitatively compared because of variations in the training sets that alter the weights of the classification function and thus the scale of the outputs. Nevertheless, this SVM methodology appears to be an excellent starting point for narrowing down a large library of potential drug molecules.

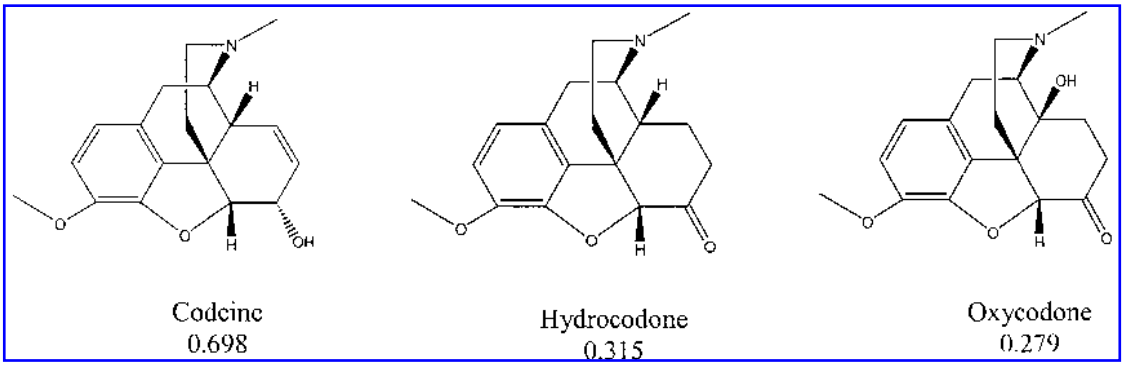

FIG. 4. Narcotic analgesics. The structure and predicted BBB permeability are shown for three CNS+ molecules of related structure.

DISCUSSION

The support vector machine outperforms the neural network

This study has compared the ability of two different machine-learning algorithms to predict the ability of drug molecules to cross the BBB and enter the CNS. The results show that over the same data set, using identical training and validation sets, the support vector machine outperforms the neural network by about 6% (81.5% to 75.7%). The result of the SVM algorithm is very encouraging, as it is comparable to other studies that have been done in this area. Ajay *et al.* (1999) used a database of over 9,000 molecules and achieved an 80% predictivity. The prediction methods of Luco (1999) and Crivori (2000) based on three-dimensional structure analysis both achieved around 90% predictivity. The SVM algorithm is impressive because even with a very small data set it performs comparably to the other methods used.

The SVM algorithm is also advantageous because it accurately classifies both CNS active and CNS inactive compounds. The neural network correctly classified 81.5% of the CNS active and 69.9% of the inactive molecules. The neural network used by Ajay correctly classified 92% of the CNS active molecules, but only 71% of the inactive molecules. The SVM correctly classified 82.7% and 80.2% of the CNS+ and CNS− compounds, respectively, suggesting that the SVM is better able to handle classification problems such as the one described here.

An assumption that has been made for the prediction algorithms is that BBB permeability is equated with pharmacological activity in the CNS. However, this simplified assumption may not be true for some molecules, where CNS inactivity is not necessarily due to an inability to cross the BBB but due to the lack of target proteins such as receptors in the CNS. It is also possible that molecules can facilely penetrate the BBB but are then quickly removed from the CNS by the P-glycoproteins and other efflux proteins present on the BBB. In the parameters used by this study, passive diffusion is the only method of transport considered. This can result in high predictive values even with molecules that transported quickly out of the CNS (hence, low CNS bioavailability) as these algorithms currently focus primarily on permeability.

AZT (azidovudine) is often a common example of a molecule that is able to penetrate the BBB by passive diffusion, but is then removed from the CNS by the efflux proteins (Physician's Desk Reference). When AZT is tested using the SVM approach, an output of +1 is obtained, indicating that the network believes this molecule can penetrate the BBB. In this case, the network was trained correctly because AZT does indeed penetrate the BBB, but AZT's CNS bioavailability is low due to its rapid removal by transport mechanisms. To further increase the usability of the SVM approach, transport properties need to be accounted for, a complicated task as experimental data and fundamental understanding of specific transport mechanisms that operate in the CNS are limited.

CONCLUSION

While SVMs are frequently being used in other disciplines, they are still being explored in the field of medicinal chemistry. The results of this study show that SVMs can be used to improve current prediction

methods for the BBB problem and many similar classification problems that are important in this field. SVMs are advantageous over neural networks because they have faster training times, they are convex problems with no local minima, and most importantly, they appear to be better able to classify small data sets like this one.

The SVM algorithm described here is a good starting point for developing a method of screening potential drug molecules based on their ability to penetrate the BBB. The SVM will reliably indicate whether a molecule will cross the BBB and we are currently developing a method for generating a dependable, quantitative prediction of BBB permeability.

APPENDIX A. THE DATABASE

Name	*MW*	*Volume*	*SA*	*% hydro*	*log P*	*HLB*	*H acc*	*H donor*	*H 3d*	*CNS +/−*
CNS inactive molecules										
acetaminophen	151.165	83.0793	10.72	49.5386	0.494	12.71	0.8284	0.55476	15.379	−1
acetylsalicylate	194.187	115.859	15.27	66.9677	1.62	15.562	0.5441	0.02283	7.60558	−1
albuterolsulfate	337.387	47.262	7.565	100	−1.23	20	0.4003	0.7762	14.7768	−1
allopurinol	136.1128	65.5428	7.909	100	−1.486	20	0.8691	0.60792	21.0878	−1
Alprostadil	354.486	205.543	27.66	27.1819	1.117	7.3946	1.5299	0.76239	11.6683	−1
aminophylline	180.166	90.3217	11.35	100	0.2021	20	0.8354	0.27728	10.6627	−1
Amiodarone	645.318	279.083	34.66	19.8972	8.1676	3.9079	0.4861	0.03712	7.64779	−1
amoxicillin	365.403	193.257	24.5	65.2894	−3.6298	15.504	2.4322	1.3648	14.574	−1
ampicillin	349.404	187.438	23.67	59.6211	−2.9628	14.553	2.0899	1.09898	11.2776	−1
androsan	302.456	207.985	26.72	12.8943	4.048	2.9768	0.5802	0.23101	10.2105	−1
aspirin	180.16	104.856	13.74	70.0384	1.394	16.55	0.7865	0.29767	10.9724	−1
Astemizole	458.577	346.325	45.14	28.6254	5.024	7.8998	0.6774	0.31075	7.50218	−1
atropine	290.382	171.955	22.02	76.4387	2.8666	15.517	0.6934	0.29095	7.76083	−1
Auxeomycin	478.885	287.349	37.59	80.1662	−7.1642	18.407	3.2498	1.8663	20.4471	−1
Betamethasone	392.467	268.728	35.89	30.9187	−1.2688	6.7817	1.7962	0.76406	14.9178	−1
carbenicillin	378.399	215.944	27.82	63.4068	−3.7628	14.97	2.0187	0.88235	11.3847	−1
carbidopa	226.232	168.584	24.07	77.1606	−1.741	16.369	1.8007	1.32903	16.4799	−1
cateolol	292.377	168.782	21.89	42.1378	1.2928	9.8591	1.2196	0.67384	11.1786	−1
Cefazolin	454.496	233.831	29.92	84.7014	−7.1492	18.632	1.4547	0.60159	13.1859	−1
cephapirin	423.458	210.628	26.62	76.5637	−4.4932	16.83	1.8601	0.62797	10.3937	−1
chlorambucil	304.216	160.074	21.02	51.7515	2.911	12.755	0.6781	0.29357	8.16285	−1
chloramphenicol	323.132	180.846	24.57	98.2414	−2.347	19.875	2.3382	0.76558	16.1802	−1
chlorothiazide	295.715	159.193	21.16	98.979	−1.419	19.932	1.4224	1.0417	17.3356	−1
chlorpropamide	276.7373	145.39	19.17	79.8239	4.532	17.754	0.9388	0.58784	11.006	−1
Chlortetracycline	478.885	287.349	37.59	80.1662	−7.1642	18.407	3.2498	1.8663	20.4471	−1
chlorthalidone	338.765	202.923	25.93	73.873	0.2252	16.097	1.5965	1.17766	17.0619	−1
clofibrate	242.7017	144.398	19.13	60.269	3.5082	15.951	0.4354	0.02836	5.40552	−1
clofibricacid	214.648	130.554	17.39	67.4761	2.2736	16.823	0.6726	0.30209	8.62268	−1
clotrimazole	344.843	204.539	23.96	29.1888	5.6497	6.6427	0.1504	0.11901	4.85491	−1
corticosterone	346.466	221.089	28.31	28.8915	0.9666	6.0071	1.2404	0.4937	13.3888	−1
Cortisone	360.449	244.232	32.14	32.7877	−1.3188	7.3282	1.5164	0.52928	13.5909	−1
coumarin	146.145	76.2801	8.836	32.6951	2.552	7.6664	0.2385	0.03588	3.30915	−1
dapsone	248.299	137.855	16.98	62.5959	−0.024	13.546	1.5988	1.06843	13.7996	−1
Dehydrocholic Acid	402.53	221.048	28.23	23.3393	1.051	5.2183	1.1413	0.27306	7.36499	−1
dexamethasone	392.467	268.728	35.89	30.9187	−1.2688	6.7817	1.7962	0.76406	14.9178	−1
dicumarol	336.3	172.036	19.74	50.5239	3.845	12.258	1.1678	0.5861	11.6674	−1
dicyclomine	309.491	192.096	25.64	23.1637	5.511	7.1144	0.4069	0	5.57391	−1
Diethylstilbestrol	268.355	162.747	20.24	18.1899	5.126	4.3254	0.6842	0.58622	13.1931	−1
Digitoxin	764.949	473.895	61.11	55.4598	2.5078	12.665	2.9952	1.13669	12.2671	−1
Digoxin	780.948	481.651	62.27	57.9109	0.2008	13.149	3.3305	1.36234	13.4658	−1
Domperidone	425.917	223.384	26.55	61.3032	2.694	16.845	1.0206	0.56757	8.42175	−1
dopamine	153.18	88.0365	11.47	50.9332	0.099	11.238	1.1862	0.91874	17.3773	−1
doxorubicin	543.5262	277.528	33.99	77.5749	−3.9958	18.192	3.5116	1.66894	17.6936	−1
doxycycline	444.44	240.953	30.48	64.8744	−7.8772	15.536	3.2504	1.86879	20.94	−1
Dyclonine	289.417	171.967	22.07	18.4714	3.714	6.6385	0.4386	0.02234	6.51322	−1
Econazole	381.688	204.875	25.06	74.9159	5.1036	16.431	0.3361	0.09037	7.37159	−1
enkephalin	554.645	404.493	55.96	58.5473	−3.4607	13.137	3.7962	2.332	12.9534	−1
Ephedrine	165.235	101.985	13.34	43.7147	1.0366	8.8473	0.6998	0.46011	11.4119	−1
epinephrine	183.207	143.461	20.44	90.3025	−0.6064	18.359	1.3801	0.99166	18.8934	−1
erythromycin	733.935	451.81	60.86	51.6776	0.143	15.081	3.0751	1.13045	10.337	−1
estradiol	272.386	208.974	27.34	15.1308	4.304	3.3794	0.6798	0.51262	13.2039	−1
estrone	270.3706	203.892	26.43	14.2347	4.062	3.3301	0.5408	0.28628	10.1271	−1
Ethacrynic Acid	303.141	147.976	19.16	68.4178	2.3456	16.232	0.9475	0.3012	8.37772	−1

(*continued*)

Name	MW	Volume	SA	% hydro	log P	HLB	H acc	H donor	H 3d	CNS +/−
ethinylestradiol	296.408	220.109	28.23	14.0364	3.724	3.1055	0.8044	0.74126	13.2087	−1
fenoterol	303.357	189.205	24.25	59.9369	1.2226	13.593	1.7269	1.28282	18.6948	−1
floxuridine	246.1947	132.7	17.46	93.6686	−1.0466	18.457	1.5013	0.73168	15.2239	−1
Flucloxacillin	453.872	224.664	27.69	59.9976	−0.0228	14.925	1.8248	0.60297	9.39797	−1
fluoruoracil	130.0783	60.0415	7.868	86.8184	−0.771	17.079	0.8902	0.56704	25.357	−1
furosemide	330.742	151.203	18.81	69.1891	1.26	15.942	1.5838	1.16672	12.663	−1
ganciclovir	255.233	133.802	17.42	100	−3.7021	20	2.1215	1.26788	20.7611	−1
glycopyrrolate	318.4351	205.249	26.59	52.7151	4.0746	10.816	0.7079	0.28846	6.68983	−1
guanabenz	231.084	109.866	14.21	96.9594	0.386	19.825	1.4731	1.37112	11.4574	−1
guanethidinesulfate	296.384	35.503	5.409	100	−1.23	20	0.4003	0.7762	14.7768	−1
homatropine	275.3468	171.955	22.02	76.4387	2.8666	15.517	0.6934	0.29095	7.76083	−1
hydralazine	160.178	91.6616	11.22	59.2701	−0.5817	13.372	0.8387	0.82897	8.45312	−1
Hydrochlorothiazide	297.731	147.013	19.1	98.8687	−1.0256	19.932	1.4195	1.17945	18.0304	−1
Hydrocortisone	362.465	252.985	33.69	33.1186	−0.8568	7.3431	1.6547	0.75479	15.2257	−1
hydroflumethiazide	331.284	160.676	21.14	76.9524	−0.8556	15.048	1.5797	1.1761	23.1565	−1
hyoscyamine	289.3736	170.149	21.33	41.7763	0.674	12.105	0.729	0.25452	10.7743	−1
ibuprofen	206.284	172.263	24.24	17.2506	4.099	4.3646	0.5389	0.29288	7.53358	−1
indomethacin	357.793	191.765	23.33	56.7074	2.6	15.127	0.8564	0.31856	8.81988	−1
isoniazid	137.141	73.3092	9.547	90.9483	−2.322	19.412	1.1705	0.8383	8.78401	−1
Isoproterenol	211.26	164.361	23.19	69.4358	0.3216	15.635	1.3845	0.99036	17.1127	−1
Isoquercitrin	464.382	228.273	27.91	90.4194	−1.9486	18.748	3.2493	2.01896	21.6959	−1
Ketoconazole	531.438	297.98	36.88	80.2482	4.283	17.361	0.9839	0.08757	7.34864	−1
labetalol	328.41	256.547	35.01	53.5117	1.5986	11.703	1.8369	1.26584	13.9555	−1
levodopa	197.19	148.354	21.12	67.9608	−1.27	14.616	1.8322	1.31758	18.6637	−1
mannitol	182.173	96.213	13.85	100	−4.2124	20	1.9888	1.35894	27.0935	−1
Meclocycline	476.87	255.367	32.08	75.7119	−6.9482	17.478	3.238	1.84933	20.3484	−1
Meclofenamate	296.152	145.113	17.69	56.7662	5.438	13.709	0.7812	0.56931	7.81618	−1
Mefenamic Acid	241.289	136.482	16.71	26.8173	4.661	6.9671	0.7865	0.57028	7.98413	−1
mefloquine	378.317	191.919	24.15	21.2789	2.7836	4.4448	1.1732	0.45315	14.5181	−1
melphalan	305.2034	157.265	20.71	63.6075	0.583	14.617	1.2779	0.78124	9.84388	−1
mepenzolatebromide	340.4413	219.877	27.99	55.2599	4.1106	10.94	0.6919	0.31679	6.46873	−1
mepiramine	285.388	179.772	22.99	35.9972	2.6689	10.941	0.3829	0.05861	9.77709	−1
mequitazine	322.467	237.082	29.97	6.95848	4.0473	2.4823	0.1129	0.0448	6.08535	−1
mestranol	310.435	231.952	29.91	14.2508	4.31	3.6741	0.5609	0.47048	10.1894	−1
methantheline	340.4413	218.495	27.75	66.9543	4.808	13.881	0.3565	0.04273	6.40151	−1
methotrexate	454.4444	247.989	31.56	73.1409	−3.3257	15.901	3.3561	1.8794	15.2169	−1
Methyclothiazide	360.23	186.755	24.84	99.1304	−0.3276	19.944	1.4517	0.88143	15.4832	−1
methyldopa	211.217	162.312	23.23	64.6741	−0.871	14.687	1.8383	1.31424	17.5827	−1
Methylprednisolone	374.476	256.509	33.83	31.7029	−0.8818	7.1075	1.65	0.76379	14.9409	−1
Miconazole	416.133	269.993	35.69	91.5355	5.8166	19.132	0.3356	0.08754	7.20945	−1
minocycline	457.482	273.01	35.27	58.9386	−6.0242	14.437	2.9182	1.60348	18.7531	−1
nandrolone	274.402	204.73	27.12	13.9257	3.58	3.2811	0.5773	0.23138	10.856	−1
neodicoumarol	408.3636	251.939	31.22	57.5524	3.403	12.986	1.4657	0.5862	11.0684	−1
Nicotinyl Alcohol	109.127	64.5981	8.343	41.891	−0.373	10.087	0.5525	0.3082	15.1963	−1
norethindrone	298.424	218.819	28.55	12.8373	2.87	3.017	0.7019	0.45994	11.0255	−1
norfenefrine	153.18	122.17	17.23	57.3852	−0.2074	11.632	1.2516	0.91814	19.1952	−1
nylidrin	299.412	203.422	26.21	24.6265	3.3946	5.8166	1.0465	0.75063	12.427	−1
Oxandrolone	306.444	205.919	26.54	15.8318	4.193	3.9822	0.6523	0.22571	9.43654	−1
Oxaprozin	293.321	187.077	22.96	33.0289	3.448	7.7098	0.7603	0.32959	8.15289	−1
oxytetracycline	460.44	247.031	31.35	70.8926	−10.184	16.822	3.583	2.09517	22.5053	−1
papaverine	339.39	308.184	45.45	40.1976	2.752	10.259	0.5563	0.05124	9.33387	−1
penicillin G	334.389	194.184	24.78	53.4015	−2.4358	13.469	1.4847	0.60827	9.24474	−1
phenacetin	179.218	103.122	13.44	37.2465	1.609	11.948	0.5897	0.28394	8.82604	−1
phenazopyridine	213.2414	160.982	21.34	64.826	−0.8185	13.8	1.3528	1.07707	14.26	−1
pheniramine	240.347	199.753	26.92	17.2961	2.605	5.3295	0.2532	0.06294	6.03587	−1
Phenolphthalein	318.328	171.447	19.83	31.753	3.789	7.9206	0.9816	0.60766	14.8069	−1
phenoxybenzamine	303.831	173.253	21.72	38.0598	4.3146	9.5171	0.2513	0.05553	5.96358	−1
Phenylbutazone	308.379	178.593	21.82	23.7635	0.995	7.8523	0.461	0.05572	6.51322	−1
pirbuterol	240.302	186.505	26.59	68.5973	0.1066	15.079	1.6078	0.97873	17.2623	−1
Polythiazide	439.87	199.357	25.89	79.0971	0.0008	16.179	1.7154	0.88035	16.3174	−1
prazosin	383.406	229.603	28.81	84.1259	−0.0053	18.848	1.4721	0.54714	12.1142	−1
prednisolone	360.449	251.119	33.37	33.2564	−1.4008	7.3841	1.65	0.76387	15.4378	−1
prednisone	358.433	242.872	31.93	32.9522	−1.8628	7.3694	1.5105	0.53829	13.8354	−1
probenecid	285.357	157.535	20.76	62.7364	3.136	15.644	0.9655	0.29972	11.0199	−1
progesterone	314.467	204.948	25.9	11.2853	4.335	2.799	0.4368	0.00495	6.63421	−1
propranolol	259.347	151.092	18.97	36.436	2.4338	8.7231	0.7291	0.436	9.59265	−1
proscillaridin	530.657	361.003	47.18	39.9123	4.7306	9.24	1.944	0.92952	11.8398	−1
puromycin	471.515	289.645	37.44	67.6296	−1.2838	15.792	2.5964	1.25215	14.2959	−1
quinidine	324.422	186.816	22.61	28.6244	2.3116	9.0071	0.8617	0.31523	11.2767	−1
ribavarin	244.207	120.71	15.8	100	−2.9006	20	2.2422	1.22236	18.989	−1
rifampin	822.951	489.992	63.58	59.186	1.0672	14.691	3.4197	1.56462	15.7408	−1

(*continued*)

Name	*MW*	*Volume*	*SA*	*% hydro*	*log P*	*HLB*	*H acc*	*H donor*	*H 3d*	*CNS +/−*
salicylicacid	138.123	71.5378	9.124	55.8614	1.218	12.459	0.8654	0.56709	17.073	−1
scopolamine	318.3919	191.136	24.08	75.3614	1.4188	15.157	0.7675	0.25473	10.0616	−1
Spironolactone	416.574	257.6	32.58	28.0484	4.4606	7.4952	0.7958	0.00502	7.31896	−1
Stanozolol	328.497	223.69	28.12	15.573	5.23	3.5336	0.6409	0.52238	9.58619	−1
sulfasalazine	398.392	213.513	26.26	84.9605	−2.9837	18.288	1.477	0.93785	15.0043	−1
Terconazole	532.469	304.611	37.99	68.1939	6.5688	16.198	0.9041	0.09305	6.72487	−1
testolactone	300.3968	205.906	26.67	19.9009	3.433	5.5947	0.5553	0.01389	7.52859	−1
testosterone	288.429	181.713	22.82	13.8033	3.529	3.1216	0.5779	0.23164	10.5197	−1
TetraCycline	444.44	240.953	30.48	64.8744	−7.8772	15.536	3.2504	1.86879	20.94	−1
theophylline	180.166	90.3217	11.35	100	0.2021	20	0.8354	0.27728	10.6627	−1
thioguanine	167.188	77.9088	9.504	94.5837	−0.9021	19.879	1.4832	1.10941	17.3205	−1
tolazamide	311.3982	179.386	23.37	54.1322	4.488	12.858	0.932	0.60188	10.0246	−1
tolbutamide	270.3458	148.793	19.6	50.1809	4.997	12.736	0.9391	0.58071	10.4155	−1
Triamcinolone	394.439	263.878	35.27	39.1234	−3.0376	8.2703	2.1255	0.99188	16.1291	−1
Triamterene	253.266	134.617	16.05	65.5448	−1.0061	13.911	2.2953	1.58537	17.507	−1
Trichlormethiazide	380.648	186.84	24.75	99.1272	−0.7306	19.947	1.417	1.19716	16.4578	−1
valinomycin	1111.33	705.075	97.18	49.8064	0.4756	14.155	4.9882	1.50676	8.46004	−1
verapamil	454.608	284.776	37.21	27.0892	4.186	8.8036	0.6603	0.03466	8.34602	−1
warfarin	308.333	167.236	19.9	31.5172	3.14	8.112	0.7801	0.32007	9.33301	−1
yohimbine	354.448	198.626	23.76	32.0197	1.494	7.962	1.0433	0.50342	11.5456	−1
zidovudine	267.244	151.885	20.12	89.3704	−2.0662	18.875	2.0797	0.50174	13.8477	−1
Average	335.2577	199.936	25.88	55.6157	0.7874	12.561	1.3545	0.71137	12.5491	
Std. deviation	140.8895	90.0684	12.04	26.4116	3.343	5.2466	0.9145	0.53178	4.68233	
CNS active molecules										
acepromazine	442.529	62.8703	9.176	100	−0.886	20	1.0444	0.56601	15.7038	1
acetophenazine	527.634	62.5695	9.128	100	−0.886	20	1.0444	0.56601	15.7038	1
acetozalmide	222.2364	123.695	17.51	92.0981	−2.9723	19.728	1.186	0.88182	22.0383	1
allobarbitol	208.2164	156.977	23.02	50.2567	0.575	12.11	1.3253	0.56084	11.7318	1
amantadine	151.251	87.9156	10.78	15.3762	2.395	2.1187	0.5076	0.3592	6.2357	1
amitriptyline	277.408	172.19	20.88	10.9933	4.961	3.6076	0.1192	0.04253	5.49208	1
amobarbital	226.275	129.632	17.71	37.9871	2.062	11.143	1.1388	0.53447	9.92661	1
amphetamine	135.208	85.8723	11.1	17.0003	1.742	4.1467	0.5044	0.38695	6.99345	1
antipyrine	188.229	127.759	16.25	39.9931	0.944	11.165	0.222	0.03302	6.78571	1
apomorphine	267.327	152.206	17.75	26.4463	2.39	6.2885	0.8159	0.56729	14.3518	1
arecoline	155.196	88.0503	11.89	65.8173	0.626	17.413	0.4373	0	7.22493	1
azaperone	327.401	182.338	22.46	35.4779	3.144	9.6603	0.5714	0.06082	7.9691	1
benactyzine	327.422	234.675	30.97	32.677	3.4146	8.4973	0.7819	0.31679	6.66431	1
benperidol	381.449	207.12	25.01	31.3755	2.741	10.284	0.7986	0.30646	7.65247	1
benzocaine	165.191	92.7985	12.03	45.8409	1.413	10.421	0.8708	0.53142	8.77164	1
benzquinamide	404.505	243.045	31.42	50.58	0.824	14.053	0.9647	0.01019	9.80881	1
benztropine	307.435	187.625	22.49	12.9725	3.22	5.1431	0.2649	0.05349	6.37245	1
biperiden	311.466	197.261	24.4	9.11787	4.341	2.7627	0.587	0.28648	9.91851	1
bromazepam	316.156	144.404	17.06	44.0701	1.8737	8.4205	0.8256	0.31946	11.0138	1
buclizine	433.0351	258.07	31.57	28.1918	9.1943	7.1868	0.2539	0.07246	5.46825	1
bufotenine	204.271	136.623	17.51	31.0225	1.008	9.1118	0.6911	0.55042	12.7753	1
bupropion	239.744	137.071	18.01	35.1429	2.619	10.639	0.6036	0.25361	3.12285	1
buramate	195.218	109.716	14.39	62.0912	0.3784	12.101	0.8788	0.50847	12.6189	1
butabarbitral	212.248	153.978	22.32	49.7435	1.533	11.88	1.1388	0.53444	10.294	1
caffeine	194.193	100.998	12.87	89.2472	1.0932	19.689	0.5951	0.01695	9.59149	1
cannabidiol	310.435	185.973	22.93	12.0194	7.085	3.6741	0.445	0.29416	9.43387	1
capuride	186.253	113.722	16.06	37.3288	0.15	10.638	1.2563	0.78814	10.6441	1
carbamazepine	236.273	145.506	17.28	28.5421	2.28	6.9463	0.7875	0.57573	8.31511	1
carphenazine	541.661	62.5695	9.128	100	−0.886	20	1.0444	0.56601	15.7038	1
cartazolate	290.364	174.615	22.71	51.7355	2.8003	13.925	0.9468	0.27178	6.9458	1
centazolone	211.223	109.69	12.52	47.4486	0.0623	10.329	0.9411	0.5772	10.3247	1
chloralhydrate	165.404	63.5401	9.166	100	0.1774	20	0.7495	0.53197	19.1305	1
chlordiazepoxide	299.759	165.204	20.01	63.5277	2.9099	13.852	0.673	0.29221	8.4304	1
chlorpromazine	318.863	233.39	31.04	28.7324	3.7413	7.7474	0.1064	0.04399	5.88878	1
chlorprothixene	315.86	181.486	21.98	21.6906	4.986	6.1737	0.1191	0.04368	5.42903	1
chlorzoxazone	169.567	75.7854	8.911	92.7263	2.541	19.643	0.4694	0.28394	9.99517	1
choline	104.172	69.9078	10.81	100	0.0484	20	0.3311	0.22657	13.3705	1
clobazam	300.744	172.585	21.12	56.1686	0.191	13.671	0.4582	0.04924	9.15862	1
clomipramine	314.857	222.957	28.71	22.6303	4.566	7.0832	0.1079	0.04214	6.44147	1
clonazepam	315.715	167.385	20.34	61.8423	2.6133	14.227	0.9891	0.30584	9.86385	1
clopenthixol	400.965	216.106	26.02	49.4443	5.1296	9.4099	0.546	0.27016	10.2821	1
clopimozide	495.998	263.959	31.43	33.7902	6.904	9.6621	0.5597	0.32741	7.06992	1
cocaine	303.357	172.319	21.59	50.5463	1.462	13.46	0.691	0.02712	6.69125	1

(*continued*)

Name	*MW*	*Volume*	*SA*	*% hydro*	*log P*	*HLB*	*H acc*	*H donor*	*H 3d*	*CNS +/−*
codeine	299.369	172.856	20.66	32.4268	0.3822	9.0915	0.7372	0.2736	10.7167	1
cyclabarbamate	245.662	123.019	16.11	84.8018	0.8238	17.716	1.2767	0.77782	14.1432	1
cyclazocine	271.402	177.395	21.86	10.6642	4.368	3.1706	0.427	0.28611	10.4234	1
cyclophenazine	433.534	246.051	30.03	25.1293	5.6887	5.635	0.3997	0.03814	7.47793	1
deanol	89.137	66.9868	10.18	72.8187	−0.6276	18.643	0.4197	0.22648	15.4077	1
demerol	247.336	179.864	24.09	33.7433	2.767	9.7104	0.3955	0.02541	6.86405	1
deserpidine	578.661	336.773	41.72	51.255	2.5282	13.178	1.4132	0.2886	10.526	1
desipramine	266.385	196.539	24.96	18.3695	3.207	4.2095	0.3093	0.21406	5.95757	1
dextroamphetamine	233.282	85.8723	11.1	17.0003	1.742	4.1467	0.5044	0.38711	6.99345	1
diazepam	284.745	168.154	20.49	47.7964	2.9339	11.629	0.4247	0.04941	8.03532	1
dihydrocodeine	301.385	170.524	20.38	28.9662	0.9262	8.9639	0.6338	0.23673	10.3096	1
dihydromorphine	287.358	160.161	18.96	30.9431	0.5352	8.6356	0.8768	0.50724	12.6188	1
diphenhydramine	255.359	192.981	25.18	18.9023	2.395	5.2512	0.4993	0.31519	10.4002	1
diphenhylhydantoin	252.272	151.169	18.25	37.4159	1.975	9.6787	0.9416	0.57122	10.4501	1
dixyrazine	427.604	257.195	32.15	46.6422	3.9554	9.3193	0.6556	0.27144	9.54018	1
DMT	188.272	131.072	16.73	20.0287	1.675	6.8036	0.349	0.28493	6.93268	1
doxepin	279.381	164.651	19.89	13.9193	3.507	4.7275	0.2354	0.04452	7.18404	1
doxylaminesuccinate	388.463	62.2583	9.077	100	−1.162	20	1.0968	0.54613	14.2901	1
droperidol	379.433	203.616	24.5	35.5821	2.197	10.392	0.8341	0.30807	8.15658	1
ectylurea	156.184	105.965	15.39	52.6818	−1.322	12.686	1.2711	0.79014	12.8509	1
emylcamate	145.201	95.4426	13.74	33.029	1.681	9.9232	0.8529	0.52306	9.3318	1
estazolam	294.743	159.59	18.44	48.2643	2.7737	11.303	0.2812	0.06794	7.95567	1
etaqualone	264.326	149.514	17.78	19.7002	2.8803	6.9654	0.3903	0.0446	6.24421	1
etazolate	289.336	170.358	22.16	61.6652	1.0123	15.843	0.9276	0.32207	7.71333	1
ethchlorvynol	144.601	103.093	15.13	49.6211	0.261	12.379	0.3765	0.50471	13.6983	1
ethinamate	260.333	156.239	22.09	49.5004	1.757	13.225	1.4394	0.76783	9.84144	1
ethosuximide	141.169	87.6587	11.99	52.4714	−0.025	13.753	0.7122	0.26296	10.9839	1
etymemazine	326.499	261.86	35.25	13.113	4.6053	3.9231	0.1071	0.0372	5.51404	1
flupentixol	434.519	227.108	27.62	38.8096	5.2996	6.4987	0.7018	0.26429	11.2269	1
fluphenazine	437.522	267.313	34.12	37.1967	4.375	7.0943	0.6896	0.26463	11.2734	1
flurazepam	387.884	250.781	32.33	46.6485	3.6849	11.324	0.5207	0.0451	8.20493	1
glaziovine	297.353	164.047	19.92	30.5935	1.296	8.2775	0.8087	0.29459	12.4982	1
glutethimide	217.267	163.989	22.19	34.0599	2.053	8.7503	0.7052	0.29167	9.22464	1
haloperidol	375.87	270.545	36.28	31.0324	3.494	8.2226	0.7581	0.3102	9.98504	1
harmalol	200.24	109.031	12.82	26.8718	1.6254	6.9967	0.8672	0.54452	14.3915	1
harman	182.224	99.3402	11.15	7.30304	2.252	3.0746	0.4269	0.30308	6.66988	1
heroin	369.416	205.587	25	42.5371	0.8096	11.647	0.8267	0.0118	13.9433	1
hexapropymate	181.234	140.304	20.08	32.9202	1.586	7.9503	0.9061	0.59818	10.0079	1
homophenazine	451.549	277.833	35.55	33.2463	4.5454	6.7847	0.692	0.26463	12.1826	1
hydrocodone	299.369	179.654	21.61	32.495	−0.0194	9.7593	0.5889	0.01048	9.93584	1
hydroxyphenamate	209.244	152.052	21	44.5235	1.1696	9.8523	1.2594	0.81031	13.0764	1
hydroxyzine	374.909	227.157	28.71	60.1868	4.305	12.414	0.6705	0.2799	9.4921	1
imiclozapine	486.074	291.413	36.42	59.6982	4.4926	12.211	0.5645	0.04399	8.82098	1
imipramine	280.412	227.275	30.04	14.1515	3.853	4.568	0.1084	0.04291	6.60928	1
isopromethazine	284.418	230.497	31.09	23.0597	3.3967	5.5608	0.1066	0.04485	7.24286	1
ketamine	237.729	156.113	20.39	46.4298	2.4678	10.899	0.6322	0.25462	4.2637	1
kynurenicacid	189.17	95.6217	11.46	62.6617	0.939	13.224	1.0214	0.5761	16.1781	1
lenperone	371.426	202.214	24.89	18.4245	3.371	5.711	0.5788	0.0465	6.53951	1
levorphanol	257.375	205.009	27.42	16.8759	3.769	4.2767	0.4376	0.28613	11.077	1
lidocaine	234.341	158.423	21.16	27.3174	2.9087	8.1147	0.6069	0.27629	8.56035	1
lorazepam	321.162	160.445	19.2	65.7033	2.9545	15.073	1.0313	0.57608	13.3371	1
lorcainide	370.921	210.522	26.06	19.6138	4.084	7.5229	0.3316	0.05474	6.88609	1
meclofenoxate	257.716	138.598	18.22	71.3573	1.8912	17.354	0.494	0.02832	5.94603	1
mecloqualone	270.718	138.826	16.24	37.4893	2.4953	11.195	0.3902	0.04858	6.54733	1
medazepam	270.761	167.524	20.49	27.7755	4.3124	8.2368	0.3425	0.04925	8.03051	1
meprobamate	218.252	125.619	17.92	61.4513	0.283	14.674	1.6926	1.04613	12.3139	1
mepyramine	285.388	179.772	22.99	35.9972	2.6689	10.941	0.3829	0.05861	9.77709	1
mescaline	181.234	107.217	14.08	41.0439	0.881	11.927	0.6999	0.37737	10.2346	1
mesoridazine	386.569	279.941	36.97	14.6678	3.0103	4.5574	0.5847	0.08202	7.56684	1
methamphetamine	149.235	96.1664	12.51	24.7682	1.888	4.0271	0.2926	0.19734	4.13086	1
methaqualone	250.299	139.651	16.45	21.2949	2.3513	7.3557	0.3903	0.04464	6.48041	1
methocarbamol	241.243	129.589	17.15	68.4088	−0.3602	15.432	1.375	0.77272	14.8402	1
methopromazine	430.518	65.8211	9.659	100	−0.886	20	1.0444	0.56601	15.7038	1
methotrimeprazine	328.471	272.477	37.89	22.7892	3.3463	6.3364	0.2057	0.03949	7.05142	1
methylpentynol	98.1444	65.146	9.561	14.705	0.123	3.4658	0.4475	0.45425	14.053	1
methylphenidate	233.31	156.825	20.46	42.3279	1.581	9.6102	0.5884	0.19723	6.86468	1
methyprylon	183.25	119.109	16.28	40.6042	−0.718	11.905	0.8218	0.24308	5.38228	1
midazolammaleate	441.845	58.6805	8.522	100	−0.886	20	1.0444	0.56601	15.7038	1
moperone	355.451	275.474	37.27	22.3986	3.43	6.0243	0.7581	0.30299	9.91125	1
morphine	287.358	160.161	18.96	30.9431	0.5352	8.6356	0.8768	0.50724	12.6188	1

(*continued*)

Name	*MW*	*Volume*	*SA*	*% hydro*	*log P*	*HLB*	*H acc*	*H donor*	*H 3d*	*CNS +/−*
MPTP	173.257	107.589	13.37	14.7393	2.785	5.7763	0.1341	0.02881	6.56756	1
naloxone	327.379	201.168	24.79	46.9643	−1.6421	13.755	1.2768	0.52363	15.431	1
nicotine	162.234	101.521	12.77	26.2634	1.473	9.3762	0.2873	0.04545	7.28322	1
nitrazepam	281.27	157.342	19.06	48.4895	1.9003	10.886	0.9893	0.30819	10.2049	1
nordazepam	270.718	145.246	17.19	50.7234	2.8703	11.418	0.6704	0.31137	9.86222	1
nortriptyline	263.382	161.454	19.41	18.7092	4.475	3.1939	0.3202	0.21399	4.58309	1
orphenadrine	269.386	199.61	25.97	24.5862	3.7506	6.094	0.2514	0.04726	5.8184	1
oxaflumazine	625.702	60.3941	8.791	100	−1.162	20	1.0968	0.54613	14.2901	1
oxazepam	286.717	150.951	18	55.2121	2.2415	12.735	1.0313	0.57813	13.7901	1
oxiperomide	337.421	187.493	22.46	43.1283	3.0476	12.577	0.6579	0.31122	7.89392	1
oxycodone	315.368	195.3	24.12	50.8671	−1.5701	14.977	0.9242	0.2365	13.7278	1
oxymorphone	301.341	184.139	22.56	53.6107	−2.1561	14.944	1.1672	0.50703	16.1407	1
paracetamol	151.165	83.0793	10.72	49.5386	0.494	12.71	0.8284	0.55476	15.379	1
pecazine	310.456	254.964	34.54	11.5856	4.1623	3.3521	0.11	0.0448	5.93168	1
pemoline	176.174	92.6105	11.38	54.6107	−1.0854	12.61	1.1665	0.80916	12.4517	1
pentazocine	285.428	193.357	24.64	9.64329	4.837	3.0148	0.4924	0.2867	10.0761	1
pentobarbital	226.275	163.529	23.58	46.8634	2.062	11.143	1.1388	0.53444	9.90168	1
pergolide	410.589	51.948	8.152	100	−1.234	20	0.5652	0.49024	12.2374	1
perlapine	291.395	185.583	22.41	41.0578	3.759	9.2081	0.3338	0.0453	8.04815	1
perphenazine-HCl	403.969	253.486	32.04	51.0623	4.205	10.628	0.5335	0.27047	10.3598	1
pethidine	247.336	179.864	24.09	33.7433	2.767	9.7104	0.3955	0.02541	6.86405	1
phencyclidine	243.391	153.121	19.07	0.67309	5.68	1.151	0.1437	0.02702	4.81765	1
phenelzinesulfate	234.27	36.1536	5.508	100	−1.23	20	0.4003	0.7762	14.7768	1
phenobarbital	232.238	123.282	15.47	44.5549	−0.158	11.891	0.9056	0.295	9.24007	1
phenprobamate	179.218	134.638	18.58	38.3615	1.922	8.2647	0.8461	0.54947	9.38845	1
phenytoin	252.272	151.169	18.25	37.4159	1.975	9.6787	0.9416	0.57122	10.4501	1
physostigmine	275.35	188.546	24.69	55.9847	2.2267	12.868	0.7265	0.26351	10.0907	1
piflutixol	451.521	232.483	28.25	5.4911	5.843	1.3737	0.6146	0.25976	10.7778	1
pimozide	461.553	251.433	29.84	19.624	6.191	6.7651	0.5599	0.32795	7.19875	1
pinoxepin	398.932	243.458	30.29	53.7794	4.3636	11.464	0.639	0.26992	10.7227	1
pipamperone	375.485	274.46	37.71	49.4381	2.0657	10.877	1.2832	0.54071	10.1961	1
prazepam	324.809	212.776	26.39	39.1636	3.6669	9.4546	0.4248	0.04983	7.6124	1
pregnanolone	318.498	209.197	26.6	9.95329	4.721	2.0726	0.5329	0.22565	8.93537	1
primidone	218.255	119.763	15.28	42.3905	−0.616	10.271	1.0721	0.51986	10.6684	1
procaine	236.313	139.987	18.5	43.4299	1.492	11.69	0.9641	0.53142	8.12057	1
procyclidine	287.444	187.197	23.71	14.4465	4.351	4.665	0.5176	0.28635	9.75009	1
promazine	284.418	216.969	28.5	14.8042	3.0283	4.5036	0.1068	0.0448	6.05603	1
promethazine-HCl	284.418	230.497	31.09	23.0597	3.3967	5.5608	0.1066	0.04485	7.24286	1
propiomazine	340.482	247.903	32.4	23.7624	3.5847	6.2312	0.3474	0.03941	6.93135	1
propofol	178.274	116.336	15.21	9.03332	4.889	1.908	0.3458	0.28513	10.2098	1
protriptyline	263.382	158.974	19.07	18.8672	4.475	3.1939	0.2918	0.22597	4.54289	1
pyridazinone	96.0884	50.4336	6.528	96.6904	−1.0223	19.79	0.3505	0.30636	18.5579	1
quazepam	386.794	228.69	29.49	39.3772	3.8889	8.7701	0.5309	0.0451	10.2628	1
rescinnamine	634.725	371.09	46.39	50.1884	2.6512	13.211	1.5113	0.29714	10.5466	1
reserpine	608.687	353.556	44.07	53.1579	2.4472	13.843	1.5121	0.28373	10.7812	1
secobarbital	238.286	176.646	25.53	43.4828	2.047	10.581	1.2319	0.54742	10.1464	1
spiclomazine	446.024	274.397	34.35	48.0574	2.7829	11.331	0.7356	0.29261	8.12072	1
tacrine	198.267	112.643	13.13	15.9304	2.194	3.0292	0.7586	0.53184	8.97614	1
temazepam	300.744	174.32	21.39	52.4673	2.3051	12.873	0.7862	0.31592	12.2162	1
tetrabenzine	317.427	190.534	24.08	29.121	2.309	10.718	0.5279	0.01022	8.51075	1
thc	314.467	192.345	24.23	9.59071	7.363	2.8631	0.4991	0.28011	9.37395	1
thebaine	311.38	184.428	22.22	34.534	1.2244	10.926	0.4588	0.02081	13.6424	1
thiethylperazine	515.684	58.3352	8.47	100	−0.886	20	1.0444	0.56601	15.7038	1
thioridazine	370.57	271.972	35.73	7.03598	5.1603	2.1601	0.1325	0.03904	5.64073	1
thiothixene	443.621	248.61	30.59	52.6173	4.9813	10.833	0.6549	0.04203	9.64377	1
tranylcypromine	133.193	86.6879	10.64	19.5667	1.468	4.2095	0.5816	0.4595	6.73272	1
trazodone	371.869	201.11	24.49	57.8872	4.7223	15.257	0.5787	0.05406	8.53395	1
trifluperidol	409.423	292.255	39.93	20.9061	3.664	5.2302	0.9138	0.30407	10.8759	1
triflupromazine	352.417	246.146	32.88	12.8346	3.9113	3.6346	0.2625	0.03814	6.53479	1
trifluroperazine	407.496	250.872	31.82	28.3067	5.2447	5.995	0.3576	0.03814	7.59384	1
trihexyphenidyl	301.471	198.789	25.31	9.24874	4.91	2.8543	0.5184	0.28635	9.43299	1
trimetozine	281.308	161.476	20.96	80.2022	−1.2456	19.355	0.7119	0.01126	10.8836	1
trimipramine	294.439	227.348	29.51	12.8434	4.252	4.3504	0.1087	0.04291	6.44024	1
tybamate	274.359	164.732	23.23	49.7897	2.416	12.622	1.4372	0.76845	9.49963	1
uridine	244.204	134.605	17.77	100	−1.5786	20	1.8322	0.96965	17.3567	1
valproicacid	144.213	89.6601	12.76	22.3394	2.81	6.2432	0.5501	0.27299	8.70361	1
zimelidine	317.228	156.492	19.31	14.0994	3.084	4.0379	0.2772	0.07459	6.9381	1
Average	293.4856	168.639	21.58	41.358	2.2865	10.017	0.6748	0.30354	9.77175	
Std. deviation	103.8409	65.049	8.222	24.747	2.0674	4.9788	0.3587	0.23549	3.27175	

REFERENCES

Ajay, A., Walters, W.P., and Murcko, M.A. 1998. Can we learn to distinguish between "drug-like" and "nondrug-like" molecules? *J. Medicinal Chemistry* 41, 3314–3324.

Ajay, A., Bemis, G.W., and Murcko, M.A. 1999. Designing libraries with CNS activity. *J. Medicinal Chemistry* 42, 4942–4951.

Crivori, *et al.* 2000. Predicting blood brain barrier permeation from three dimensional molecular structure. *J. Medicinal Chemistry* 43, 2204–2216.

Fischer, H., Gottschlich, R., and Seelig, A. 1998. Blood-brain barrier permeation: Molecular parameter governing passive diffusion. *J. Membrane Biology* 165, 201–211.

Luco. J. 1999. Predicition of the brain blood distribution of a large set of drugs from structurally derived descriptors using partial least-squares modeling. *J. Chem. Inf. Comput. Sci.* 39, 396–404.

Lundbeck, H. 2000. Psychotropics Database 2000–2001.

Matlab Neural Net Toolbox, The Mathworks Inc.

NCI database, ChemSW. 420 F Executive Ct., N. Fairfield, CA 94585. *www.chemsw.com.*

Pardridge, W. 1998. CNS drug design based on principles of blood-brain barrier transport. *J. Neurochemistry* 70(5), 1781–1792.

Physicians' Desk Reference, Medical Economics Company. *www.pdr.net.*

van de Waterbeemd, H., Camenisch, G., Folkers, G., Chretien, J.R., and Raevsky, O.A. 1998. Estimation of blood brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. *J. Drug Targeting* 6(2) 151–165.

www.nlm.nih.gov/medlineplus/druginformation.html. Supported by the National Library of Medicine.

Address correspondence to:
Joanne I. Yeh
Brown University
MCB Department
Box G-J2
69 Brown Street
Providence, RI 02912

E-mail: Joanne_Yeh@Brown.edu