Designing Libraries with CNS Activity

Ajay,* Guy W. Bemis, and Mark A. Murcko

Vertex Pharmaceuticals Inc., 130 Waverly Street, Cambridge, Massachusetts 02139

Received January 13, 1999

Library design is an important and difficult task. In this paper we describe one possible solution to designing a CNS-active library. CNS-actives and -inactives were selected from the CMC and the MDDR databases based on whether they were described as having some kind of CNS activity in the databases. This classification scheme results in over 15 000 actives and over 50 000 inactives. Each molecule is described by 7 1D descriptors (molecular weight, number of donors, number of acceptors, etc.) and 166 2D descriptors (presence/absence of functional groups such as NH_2). A neural network trained using Bayesian methods can correctly *predict* about 75% of the actives and 65% of the inactives using the 7 1D descriptors. The performance improves to a prediction accuracy on the active set of 83% and 79% on the inactives on adding the 2D descriptors. On a database with 275 compounds where the CNS activity is known (from the literature) for each compound, we achieve 92% and 71% accuracy on the actives and inactives, respectively. The models we construct can therefore be used as a "filter" to examine any set of proposed molecules in a chemical library. As an example of the utility of our method, we describe the generation of a small library of potentially CNS-active molecules that would be amenable to combinatorial chemistry. This was done by building and analyzing a large database of a million compounds constructed from frameworks and side chains frequently found in drug molecules.

Introduction

Combinatorial chemistry and high-throughput screening (HTS) have frequently been justified based on the premise that the probability of finding a hit in a screening experiment is proportional to the number and variety of molecules screened. However, synthesizing and screening a very large number of compounds is costly. Preparing and testing a random selection of compounds may not be the most efficient way to extract the maximum possible information from an experiment.

Researchers are now beginning attempts to enhance the information content of screening libraries. One method that has seen much use in this context is to enhance the diversity of the compounds screened. The basic rationale behind this is provided by the so-called similarity principle—similar compounds often show similar biological activity.[1] If this is true, then it follows that a screening library covering diverse structural types will generate leads for a large number of biological targets. The concept of enhancing the information content (through diversity and other means) is also important for lead optimization.

Clearly just diversity is not sufficient for all types of screening libraries. If the library is meant for broad use, one may still want to have a diverse set of "drug-like"[2,3] molecules. If the library is meant for a specific target (or set of related targets), one would of course like to incorporate additional target-specific information. For example, if the target is a zinc metalloproteinase, then a diverse set of molecules with zinc binding moeities would be preferred. If the target is an aspartyl protease, then a diverse set of molecules without an aspartate binding hydroxyl group would not be likely to provide useful leads. Again, if the target protein is found in the central nervous system (CNS), then we need molecules that have features which would at least enable them to cross the blood−brain barrier (BBB).

In this paper we demonstrate a method by which molecular feature-based information about BBB penetration and CNS target activity can be obtained. We also demonstrate a computational library design experiment using our results. Our approach is simple. We use a set of descriptors to learn to distinguish CNS-active molecules from CNS-inactive molecules. This is facilitated by the availability of large databases of known CNS-active agents. There are over 80 000 compounds in the CMC[4] and MDDR[5] databases. Confining our attention to compounds with molecular weight less than 600 and calculated log *P* values between −4.0 and 11.0, we are left with roughly 73 000 compounds. Out of these roughly 18 000 compounds are listed as having CNS activity. CNS-active molecules are identified by searching for therapeutic classes such as anticonvulsant, antidepressants, antipsychotic, etc. (see Methods for a detailed discussion). We make the assumption that most molecules in the CMC and MDDR that do not fall into these classes will show no CNS activity (see below for a detailed discussion). This work, therefore, attempts to mimic the characteristics of compounds that were designed and marketed to actively target the central nervous system.

CNS Activity. It is believed that CNS activity (and transcellular permeability in general) is a complex function of physical/chemical properties of molecules such as size, lipophilicity, hydrogen-bonding potential, charge, and conformation.[6] For any given molecule, one of these factors may dominate others.[7]

* To whom correspondence should be addressed. Tel: 617-577-6595. Fax: 617-577-6400. E-mail: ajay@vpharm.com.

10.1021/jm990017w CCC: $18.00
Published on Web 11/10/1999

Table 1. A List of Molecules within These Activity Classes Were Output from both the CMC and the MDDR[a]

activity classes used to define CNS-active molecules
cognition disorders, agent for"
"anxiolytic"
"antipsychotic"
"neuronal injury inhibitor/neuroleptic/neurotropic"
"antidepressant"
"analgesic, non-opioid"
"anticonvulsant"
"antimigraine"
"antiischemic, cerebral"
"analgesic, opioid"
"antiparkinsonian"
"sedative"
"hypnotic"
"stimulant, central"
"antagonist to narcotics"
"centrally acting agent"
"nootropic agent"
"neurologic agent"
"epileptic"

[a] These defined all the molecules with CNS activity.

Drugs with the brain as the site of action should, in general, be able to cross the BBB. Drug delivery to the brain can be enhanced by increasing the lipophilicity of the molecule, by using prodrugs that dissociate after crossing the BBB, or by using passive or active drug targeting that utilizes transport systems at the BBB in the normal or disease states.[8−10] In general, the transendothelial transport of compounds can depend on binding to constituents of the plasma, ionization state, time-dependent plasma concentration, and cerebral flow. It is possible to modify many of these properties with changes in chemical structure.

Previous attempts at understanding CNS activity have resulted in certain rules-of-thumb. For example, Andrews et al.[11] have shown that an aromatic ring−tertiary nitrogen pharmacophore is important for CNS activity. Levin[12] has successfully correlated octanol−water partition coefficient and brain capillary permeability for compounds with molecular weight less than 400. However other, more recent attempts conclude that the octanol−water partition coefficient does not correlate well with blood−brain transport.[13,14] Other criteria, like a limit of 8−10 hydrogen bonding groups per molecule, have also been proposed.[10]

Methods

Historically CNS activity has been studied using standard QSAR approaches on a small series of related compounds.[13,15−17] Recently Norinder et al.[18] built a partial least-squares model on a diverse set of compounds, but their dataset was also small—tens of compounds. This procedure is clearly inappropriate if the aim is to design diverse combinatorial libraries. We need to have a large database to generate models. We have therefore decided to adopt a classification method based on therapeutic use. Two recent publications[19,20] also adopt a therapeutic use-based classification to differentiate between CNS-actives and -inactives. The approach adopted in both of these works, however, is not based on generating regressions.

Compounds in therapeutic classes given in Table 1 were considered to fall into the CNS-active class. Other compounds in MDDR and CMC databases were by default CNS-inactive. It is possible that some of the default CNS-inactive class may actually have CNS activity, but it is reasonable to assume that a majority of these compounds are inactive. This selection procedure results in (after pruning based on molecular weight, etc., and the availability of ISIS keys) 1 050 compounds from the CMC and 16 785 compounds from the MDDR that are classified as CNS-active.

Large databases provide immunity against over-representation of any single class of compounds. For example, the CNS-active set contains about 50 (CMC) and 250 (MDDR) 2-benzylpiperidines (used as a pharmacophore for opioids), about 20 (CMC) and 240 (MDDR) 2,3*H*-1,4-benzodiazepines (used as a pharmacophore for benzodiazepines). Figure 1 in the Supporting Information shows some of the most frequently occurring frameworks[21] in the CNS-active MDDR database. This shows in detail that no nontrivial framework (a benzene ring is an example of a trivial framework), as defined by Bemis and Murcko,[21] dominates the database.

The "automated" classification scheme based on therapeutic use adopted in this work, though useful, can lead to problems in constructing training and test sets. Some compounds will be incorrectly classified. The CNS-inactive classification, in particular, is difficult. For example, the classification of CNS-inactives is confused by noting that crossing the BBB is not necessarily synonymous with the ability to cause CNS effects. Therefore we need to adopt a learning method that can handle noisy classification data. In this data model, called an (a,b) classification noise model, we have a setting in which true positive examples are incorrectly labeled (independently) with probability a and true negative examples are incorrectly labeled (independently) with probability b. The goal of the learning algorithm in this setting is to produce a hypothesis that is ϵ-close to the target function with respect to nonnoisy data. Our experience[22] with Bayesian neural nets shows that they are robust under this classification noise model. For demonstrating that our models perform well under nonnoisy data conditions within the context of this work we have performed three tests:

1. Explore the model in detail to check for reasonableness of the parameter values. We do this in the discriminant analysis models (BNN0) developed by checking for the sign and magnitude of the contribution from the parameters of the 7-descriptor set. In this connection it is important to note that the nonlinear models are built under the same protocol as the linear ones.

2. Explore the performance of the models on a subset of CNS-active MDDR compounds. These compounds are selected based on the information in the database regarding tests on animals.

3. Explore the performance of the models on a dataset where compounds are individually labeled. This dataset has 275 compounds (see Table 1 in Supporting Information for a complete listing). This database was originally constructed by compounds found in the papers by Waterbeemd et al.[19] and Fischer et al.[20] and the compounds used by Norinder et al.[18] along with some collected from various published sources by us. All of these compounds were individually reconfirmed by going back to the original literature. The Supporting Information section offers more details.

Descriptors. There is a wide range of choices available for describing molecules. Here we choose a small set of 1D descriptors, namely, molecular weight (MW), number of hydrogen bond donors (Don), number of hydrogen bond acceptors (Acc), number of rotatable bonds (Rot), ${}^2\kappa_\alpha$, which is a measure of the branching of the molecule, aromatic density (AR), and log P. This 7-descriptor set which contains information about the entire molecule is calculated with internally developed programs. In particular, log P is calculated using our implementation of the Moriguchi method.[23]

We also use a 2D descriptor set based on the ISIS fingerprint. This set of descriptors contains information about the specific set of functional groups within the molecule. The ISIS fingerprint is a bit string (a string of 0's and 1's) of length 166 with a 1/0 indicating the presence/absence of some moiety or "key". There are 166 such keys for each compound. This choice of descriptors is reasonably common in the diversity literature.[24] In fact, Brown and Martin[24] have recently shown that the 166 ISIS keys perform remarkably well for clustering and diversity analysis. This, of course, does not mean that ISIS

Table 2. Two Sets of Keywords Used To Narrow the List of MDDR Active Compounds[a]

action phrases for defining an active MDDR subset
healthy
rat
cat
mice
mouse
rodent
guinea
ferret
dog
monkey
primate
mg/kg
phase phrases for defining an active MDDR subset
preclinical
phase
launched

[a] The first set of phrases is used for searches in the "action" field in the MDDR database and the second in the "phase" field. These lists were then merged resulting in about 3 000 compounds.

keys are useful for all classification problems. The 166 keys for each molecule were written out using the ISIS software. We also explore performance by combining the 166 keys with the 7 calculated descriptors.

Our choice of representation precludes some important characteristics of molecules, e.g., (1) conformational flexibility and (2) stereoisomerism. Conformational preferences of a molecule, in general, are very hard to capture uniquely in any representation. Also, it is not clear that stereospecific information is available for most of the compounds in the databases. Other choices of descriptors are possible, and we are currently working on refining and enhancing our collection of descriptors (results will be reported elsewhere).

Training and Test Sets. We consider 20 complementary train/test set pairs. Each training set consists of roughly 9 000 compounds with active and inactive compounds contributing equally. The remaining compounds fell into the test set. Each test set has roughly 13 000 actives and roughly 53 000 inactives. All train/test set pairs are independently and randomly created. Since roughly 4 500 compounds each are randomly selected from about 17 000 actives and 57 000 inactives, the probability of significant overlap between any two training sets is low, so the models will not show any significant overlap due to overlapping training sets.

Test Sets Based on Subclasses of Compounds. To explore our results in detail we report the performance on different subclasses of compounds. We will analyze the results for 11 subclasses, namely (1) analgesics, (2) anticonvulsants, (3) antidepressants, (4) antipsychotics, (5) anxiolytics, (6) hypnotics, (7) compounds that contain both a tertiary amine and an aromatic ring, (8) compounds with MW $<$ 300, (9) compounds with MW $\geq$ 300, (9) compounds with log P $<$ 4.0, (10) compounds with log $P \geq$ 4.0, and (11) compounds in the MDDR CNS-active set with known animal test results.

The 11th subclass was created in the following manner. As the MDDR database contains some information about the status of human/animal tests done on *some* of the compounds, it is possible to create a set of CNS-active compounds from the MDDR that reflects detailed biological information. This was done by creating a subset of the MDDR CNS-active list based on the keywords in Table 2. This procedure generated about 3 000 CNS-active compounds.

Table 1 (in Supporting Information) shows the compounds we have collected from the literature with known CNS activity information along with our predictions. As mentioned, this list was originally culled from the work by Waterbeemd et al.[19] and Fischer et al.[20] We have, however, succeeded in independently confirming the classification of over 95% of the compounds in this table from the literature.

In any prediction method it is useful to obtain an idea about true and false positives and negatives. True positives are known CNS-active compounds that are predicted correctly, and true negatives are known CNS-inactive compounds correctly predicted to be inactives. False positives are CNS-inactives that were wrongly classified as actives, and false negatives are CNS-actives wrongly predicted to be inactives. Since the CNS-active list is fairly well-defined, we can easily get a handle on both true positives and false negatives. But, the CNS-inactives are not based on knowledge about individual compounds that belong to this class; *i.e.*, some of the CNS-inactives may actually cross the BBB and hit some targets in the brain. It is, therefore, much harder to get precise values for false positives and true negatives.

Bayesian Neural Network. We have previously described our Bayesian learning procedure in detail.[2] Neural networks (NN) can be viewed as a flexible regression (classification) technique. A Bayesian approach to NN modeling[25,26] allows for simultaneous and reliable optimization of a large number of control parameters.

Within a Bayesian procedure a large number of models are built. Associated with each model is a probability weighting factor that is high if the classification error of that model is low and vice versa. It is intuitively clear that most choices of NN weights (models) will lead to low probability weightings. It is therefore imperative that a reliable and robust method be found for sampling in NN weight space. This is done in analogy with accepted procedures in molecular dynamics and Monte Carlo methods used in protein simulations. Technically these are called Markov Chain Monte Carlo methods.[27]

Our Bayesian learning procedure enables us to determine the significant descriptors in the model(s). This is done by examining the hyperparameters associated with each weight that connects the inputs in the network (see Ajay et al.[2] for details). Weights with small hyperparameters have small contributions and vice versa, with one important exception: If a descriptor has the same value (in our case 0 or 1) for all the compounds in the training set, the value of the associated hyperparameter is unreliable and should be eliminated from consideration.

We also eliminate the descriptors with small (defined precisely in Results) hyperparameter values for each of the 20 networks constructed. Note that the training set determines the important descriptors. Therefore, not all descriptors will necessarily be considered important by all networks. Next, the descriptors that appear at least in 10 out of the 20 networks are obtained and reported. This produces a "consensus set" of important descriptors.

Results

Comparing the CNS-Active and -Inactive Distributions Based on the 7 Descriptors. Table 3 provides a comparison between CNS-active and -inactive compounds in the database (CMC and MDDR) based on the 7-descriptor set. As expected, CNS-active compounds are somewhat smaller than other biologically active molecules: 90% of the CNS-active compounds have anywhere from 2 to 7 hydrogen-bonding groups while this range is from 2 to 9 for the inactives. Over 90% of the CNS-active compounds have 7 or fewer rotatable bonds, while this number is 10 for the inactives. CNS-active compounds in general have somewhat larger log P (more lipophilic) than other biologically active molecules.

Predictions and Consensus Performance. Table 4 gives the main results for training/testing on compounds from the MDDR and CMC. The best results indicate that we can achieve around 80% predictivity. There are some important themes in these results that we should highlight:

1. The performance of the 7 1D descriptor set is quite remarkable: 75% accuracy on actives and 65% on inactives.

Table 3. Differences between CNS-Active (CNS+) and CNS-Inactive (CNS−) Compounds in Terms of the 7 1D Descriptors

descriptor	description	value (CNS+/CNS−)
MW	mean	354/387
	median	351/385
	90% range	200−540/198−577
$^2\kappa_\alpha$	mean	7.4/8.3
	median	7.3/7.9
	90% range	3.4−12.2/3.4−15.9
log *P*	mean	2.8/2.4
	median	2.9/2.5
	90% range	0.0−5.2/−1.0−5.4
donors	% with zero	16/34
	% with at least one	73/71
	% with at least two	96/90
	% with at least three	99/96
acceptors	% with zero	5/2
	% with at least one	19/8
	% with at least two	43/22
	% with at least three	65/41
	% with at least four	81/60
rotors	% with at least one	20/20
	% with at least three	49/44
	% with at least five	77/65
	% with at least seven	90/79
AR	% with zero	4/12
	% with at least one	27/38
	% with at least two	70/72
	% with at least three	93/91
Don+Acep	% with at least one	4/3
	% with at least two	18/11
	% with at least three	39/24
	% with at least four	62/42
	% with at least five	78/58
	% with at least six	88/73
	% with at least seven	94/82

Table 4. Range of Errors for All 20 Test Sets[a]

method	% error on actives	% error on inactives
BNN0;7des	28−31	37−40
BNN5;7des	25−28	34−38
BNN5;ISIS	19−22	22−25
BNN0;ISIS+7des	21−22	24−26
BNN5;ISIS+7des	17−20	21−23

[a] Notice that even though the training sets (of size roughly 9 000) were chosen completely randomly from the whole database (of size roughly 70 000) the differences in the error rate are quite minor. We have broken down the predictions into CNS-active and -inactive sets as the number of compounds in the inactive set vastly exceed the actives. The results are shown for a neural network with 5 hidden units (BNN5) and with no hidden units (BNN0) trained using a Bayesian procedure. There were roughly 13 000 compounds in each of the CNS-active test set, and roughly 53 000 in the CNS-inactive.

2. The performance of the ISIS keys is even better, a 5% improvement in accuracy for the actives and a 10% improvement in accuracy on the inactives compared to the 7-descriptor set.

3. The ISIS keys and the 7-descriptor set together yield the best results: 81% predictive accuracy on the actives and 78% on the inactives.

4. The accuracy on the inactive sets is lower by about 4% compared to the active sets. This could be due to the much larger (and hence more diverse) number of compounds in the inactive test sets. Another reason could be that we have not assigned all the active compounds correctly by the classification scheme in Table 1 (see below).

The performance of the linear model is very close to the best, though the addition of nonlinearities results in a small (3%) improvement in predictability. Notice the advantages of the Bayesian learning procedure from these results. Despite increasing the number of parameters by almost an order of magnitude, the prediction performance does not deteriorate; i.e., there is no overlearning. In addition the predictive error in the different test sets are not very different from one another. As expected for any learning methodology, the training set errors are lower than the test set errors by about 8% (same for both actives and inactives).

We have run exploratory analysis on training sets with 5 000, 8 000, and 13 000 compounds. The performance of the networks with 8 000 and 13 000 compounds lies roughly within the same error ranges as shown in Table 4. The predictive accuracy goes down by about 10% when using only a 5 000-compound training set.

Table 5. Results of Consensus Predictions on 6 Subclasses of the CNS-Active Compounds[a]

subclass	total	% not CNS
analgesics	2592	19.0
anticonvulsants	1747	11.4
antidepressants	2896	7.5
antipsychotics	2905	5.6
anxiolytics	3593	10.0
hypnotics	447	13.2

[a] It is important to realize that the subclasses are not all mutually exclusive. "Total" gives the total number of compounds in both the CMC and MDDR.

Predictions on Subclasses. In this section we report on the predictions on the 11 subclasses of CNS-active compounds based on BNN5 with all 174 descriptors. Table 5 indicates that consensus predictions on analgesics produce the largest errors and antipsychotics the smallest. There are some important details that should be kept in mind: (1) there is some overlap of compounds between the classes, and (2) most of the compounds in these subclasses are not part of the training set, and hence the results in Table 5 are mostly *predictions*. It is not clear why analgesics should be harder to predict than the other subclasses. The error in all the six subclasses of CNS-active compounds is at or below the roughly 19% expected from the results in Table 4.

1. Compounds with a Tertiary Nitrogen and an Aromatic Ring. An interesting subclass to explore is the one formed by all the molecules that have a tertiary amine and an aromatic ring moiety in them, as Andrews[11] argues for the importance of this moiety for CNS activity across different targets. Roughly 54% of the CNS-active compounds in the CMC have a tertiary amine and an aromatic ring moiety, while this number is 57% for the CNS-active MDDR database. About 31% of CNS-inactive compounds in the CMC and 36% of MDDR have this moiety. Another reason to look for predictivity on this subclass is due to the lack of any single key that specifies a tertiary amine in the 166 ISIS keys that we use.

Actives: Roughly 6−7% of compounds that are CNS-active in this subclass are incorrectly classified as CNS-inactive (see Table 6). So the false negative rate is quite low and is very similar in both the CMC and the MDDR. This implies that the true positive rate is high.

Inactives: The performance of the network on CNS-inactive molecules with a tertiary nitrogen and an

Table 6. Results of Consensus Predictions on 4 Subclasses of Compounds Based on CNS-Actives (CNS+) and CNS-Inactives (CNS−)[a]

subclass	class	total	% incorrect
tertN+arom	CNS+	10197	6.5 (FN)
	CNS−	20170	36.0 (FP)
MW < 300	CNS+	5950	6.6 (FN)
	CNS−	11108	39.0 (FP)
MW ≥ 300	CNS+	11885	11.2 (FN)
	CNS−	46434	18.0 (FP)
log *P* < 4.0	CNS+	15057	12.5 (FN)
	CNS−	47711	22.2 (FP)
log *P* ≥ 4.0	CNS+	2778	14.1 (FN)
	CNS−	9831	21.6 (FP)
MDDR subset	CNS+	2997	13.4 (FN)

[a] "Total" gives the total number of compounds in both the CMC and MDDR, except for the MDDR subset. Here, FN stands for false negative and FP for false positive.

Table 7. Therapeutic Classes Discovered by the Networks as Belonging to CNS-Actives

new set of CNS-active classes
tranquilizer
antivertigo
anorexic
narcotic antagonist
serotonin antagonist
anti-anxiety
sleep
enhancer
sigma opioid antagonist
antiemetic
antinauseant
antispasmodic/anticholinergic

aromatic ring moiety is more interesting. To explore this in detail we look at the CMC and MDDR subsets separately.

False positive rate on CMC-inactives: There are 1 453 compounds in the CMC-inactives with a tertiary amine and aromatic ring moiety. Out of these, 858 compounds are predicted to be CNS-active (i.e., a 59% false positive rate). The question is, "are these really failures?" 297 out of the 858 compounds have *no* information in the activity class (and hence by default they are classified as CNS-inactive). Removing them from consideration and exploring the activity classes of the remaining 561 compounds, we get about 168 compounds that fall in the categories given in Table 7. These classes should have been included in the list (Table 1) that formed the definition of CNS-active compounds[28] but were inadvertently excluded. Correcting for this brings the false positive rate to 34% (assuming we decline to make any predictions on the 297 compounds with no defined activity class)—a reasonable figure considering the overall prediction accuracy. A further 16% of the compounds left are antihistamines, a majority of which may have CNS activity.[28,29] In addition, some of the compounds (e.g., chloramphenicol, phenoxybenzamine, ergonovine, etc.) that are left are known to have some CNS activity,[28,29] providing further proof that not all compounds classified as active from the inactive list are incorrect. So the false positive ratio is not as bad as it appears at first sight.

True negative rate on CMC-inactives: There are 595 compounds that were classified as CNS-inactives from among the 1453 compounds designated as CNS-inactives. Only 15 of these compounds fall into any of the classes in Table 7. Out of the 580 left, 477 had some information in the "activity class" within the CMC database, out of which only 13 were antihistamines. None of these "activity classes" can be reasonably classified as CNS-active. Therefore a majority of the inactive predictions appear to be correct.

False positive and true negative rates on MDDR-inactives: About 33% (roughly 6 200 out of 18 500 compounds) of the compounds designated as CNS-inactive were classified as CNS-active. Out of these 6 200, 34 fall into classes in Table 7. There is some indication for possible CNS penetration in the MDDR database for another 99 compounds. About 5% of the remaining compounds are antihistaminic. It is gratifying that the false positive and true negative rates for MDDR are not different from the ones found in the CMC.

2. Subclasses Based on Molecular Weight. Actives: Among the CNS-active molecules with molecular weight less than 300 the prediction accuracy is quite high (see Table 6). It worsens somewhat for compounds with molecular weight greater than 300 but is not bad compared to the overall errors. Therefore reassuringly, true positive rates are high while false negative rates are low. However, it is clear that a larger percent of heavier active compounds are predicted to be inactive.

False positive rate on CMC-inactives: Out of the roughly 2 300 CMC molecules with MW < 300 designated as CNS-inactive, about 970 are classified as CNS-active by the networks. This implies an apparent false positive rate of 42%. Only 705 out of these 970 compounds have some description in the activity field of the CMC database. Roughly 125 out of these 705 compounds fall into the categories given in Table 7, reducing the false positive rate to roughly 25%. About 8% of the remaining compounds are antihistamines. For compounds with MW ≥ 300, 638 compounds from the 2 415 CMC CNS-inactive designates were classified as CNS-actives, with 423 having some activity information. This implies an apparent false positive rate of 20%. Removing compounds from Table 7, the false positive rate falls to 13%. Again roughly 8% of the remaining compounds are antihistamines. Therefore, as was the case with the aromatic ring−tertiary nitrogen pharmacophore, the false positive rate is not unreasonably large. Notice that the trend for false positive rate among compounds with MW ≥ 300 is *smaller* than for compounds with smaller molecular weight.

True negative rate on CMC-inactives: Out of the roughly 2 300 CMC inactive designates with MW < 300, 1 326 were classified as inactive, implying an apparent 58% true negative rate. 1 128 compounds has some information about the activity class in the database, and out of these only 25 compounds fall into any of the classes in Table 7. For compounds with MW ≥ 300, there are 1 500 inactives with some information in the activity class. 65 out of these 1 500 belong to the classes in Table 7. Therefore, very few of the known actives are being misclassified as inactives.

False positive and true negative rates on MDDR-inactives: For MDDR inactive designates with MW < 300, the initial false positive rate appears to be a high value of 38%. Only 10 molecules out of about 3 350 compounds predicted to be active fall into the categories

in Table 7. There is some indication of CNS activity for another 80 molecules. About 100 molecules are also central 5-HT receptor agonists. Central 5-HT receptor agonists should also have been included in the definition of CNS-active classes but were again inadvertently missed. Not much can be said about the rest of the molecules. The false positive rate for MW ≥ 300 is only 17%, so as seen for the false positive rate in the CMC-inactives here too the false positive rate on heavier molecules is lower than for lighter ones. The results for true negatives also parallel the ones found for the CMC above.

3. Subclasses Based on log *P*. As reported in Table 6, we again obtain high values for true positives and low for false negatives. As observed earlier the tentative values for false positives are high, but after some corrections they become smaller. The situation with true negative values is again similar to the results observed above for the different subclasses.

428 compounds out of 1 130 CMC CNS-inactives with log $P \geq 4.0$ were classified as CNS-active. Out of these 275 have some activity information available in the CMC database. So tentatively about 28% of compounds are false positives. Removing compounds in Table 7 reduces the false positive ratio to 21.5%. 15% of the remaining false positives are antihistamines. This is a higher false positive ratio than was observed before but is within the average expected error.

1 179 out of 3 580, CMC CNS-inactives with log $P <$ 4.0 were classified as CNS-active. Only 853 compounds have some information in the activity class. So tentatively about 26% of the compounds are false positives. Again, removing compounds in Table 7 reduces the false positive ratio to 21.2%. 6.2% of the remaining compounds are antihistamines. So again, the number of false positives is down to a reasonable number. A cursory perusal of some of the remaining compounds in consultation with the *Drug Facts and Comparisons*[28] indicates that some of these do penetrate the CNS.

4. Active Compounds with MDDR Animal Data. Out of the roughly 3 000 compounds in this subset of MDDR active compounds the false negative rate is quite small at 12%. About 30%of these false negative compounds are neuronal injury inhibitors. This does not mean that the networks have not learned the characteristics of neuronal injury inhibitors (NII) as only 20% of the NII's in this subset are misclassified. As in all the previous results the error in this subset is also small and within the 20% overall error in Table 4.

Database of Compounds with Known CNS Activity. We have collected from the literature a set of 275 compounds with known CNS activity. 139 of these are active and 136 are inactive. We have achieved 93% accuracy among the actives while this number is 72% for the inactives. These results are roughly equivalent to the overall results reported earlier. To our knowledge this is the largest database on which any BBB models have been tested. It also shows that the learning method advocated in this paper is robust to the classification noise. We have been able to collate experimentally determined $BB = C_{brain}/C_{blood}$ data on 80 compounds (starred in the table in Supporting Information). 37 of the 38 actives have been correctly predicted, while 30 of the 42 inactives were correctly predicted.

Another view of the results can be obtained by looking at predictions along with confidence levels of each predictions. Our confidence levels directly incorporate, using Bayesian averaging, the predictions of all of the 20 models we have built. If we look at predictions over a 50% confidence level, the performance of our models improves substantially to around 88% for both actives and inactives. At this level of confidence we miss out on roughly 25% of the molecules in this database. More details can be found in the Supporting Information. The inactive prediction accuracy again rises to over 80% for the 80 compounds with experimental *BB* data.

84 compounds in this database were not present in our original training or test sets. 95% of the actives (38 out of 84 total) were predicted correctly, and 75% of the inactives were predicted correctly. Incorporating confidence levels again improves the predictive accuracy on the inactives to over 85% and decreases it to 88% for the actives.

The 7-Descriptor Set. The performance of a linear network using just the 7 descriptors is surprisingly good, especially considering the simplicity of the description. This network correctly predicts about 70% of the actives and roughly 60% of the inactives. As we have seen above some of the compounds designated as CNS-inactive actually should be CNS-active. This will improve the correct classification rate on the inactives. We will study these linear models in some detail as it is easier to build intuition about the problem based on this descriptor set.

Using the distribution of the median value of the hyperparameters we obtain, in general, the following rank ordering of importance among the 7 descriptors:

$$\text{Acep} > \text{AR} \approx \text{Don} \approx \kappa > \text{MW} \approx \log P > \text{Rot} \quad (1)$$

This rank ordering is seen in almost all the 20 networks trained on independent data.

All twenty networks give the same general results. If the MW, κ, or number of rotatable bonds (Rot), or number of hydrogen bond acceptors (Acep) is increased, the compound will be *less* likely to be CNS-active. On the other hand, if the aromatic density (AR), number of hydrogen bond donors (Don), or log *P* is increased, the compound is *more* likely to be CNS-active. This result is compatible with accepted wisdom on the characteristics of molecules that are CNS-active. There are some important constraints on allowable values for the descriptors. No predictions should be made for molecules if MW is less than 60 or greater than 600, if log *P* is less than −4.0 or greater than 11.0, if $\kappa > 30$, or Don > 20, or Acep > 30, or Rot > 30, or AR > 20.

Relevant Descriptors. We have chosen to analyze the linear models due to their simplicity and because the performance of the linear models is comparable to the best results. Any descriptor with a median hyperparameter value (after convergence of the network dynamics; see Ajay et al.[2] for details) less than or equal to 0.1 was deemed irrelevant. This retained roughly 80 descriptors out of the original 173 in all 20 BNN0 models. The descriptors that appear at least 10 times out of the 20 are shown in Table 2 (Supporting Information). Among the 7-descriptor set all except log *P* appear at least 10 times out of 20. Further, MW, Acep, κ, and AR appear in all 20 networks, the remaining two, namely Don and Rot, appear in 16.

Table 8. Performance of the Linear and Nonlinear Networks on a Subset of Compounds That Have a 4-Member Ring[a]

dataset	class	total	% correct (BNN0)	% correct (BNN5)
train set	CNS+	15	20	93.3
	CNS−	274	100	99.6
test set (CMC)	CNS+	6	16.7	66.7
	CNS−	145	100	100
test set (MDDR)	CNS+	48	45.8	79.1
	CNS−	2698	99.7	99.3

[a] The improved results on adding nonlinearities is reflected on both the training set and the test set. "Total" indicates the total number of compounds in each subclass. All the results correspond to using all 173 descriptors.

These results are in keeping with previous studies that have correlated molecular size (roughly correlating with molecular weight, number of rotatable bonds, and $^2\kappa_\alpha$) with transendothelial transport.[30] The case for log *P* is interesting. It appears that octanol−water partition coefficient is useful for modeling CNS potency, but not brain−blood concentration ratio.[15] It is also likely that log *P* is significant only for compounds that do not undergo any facilitated transport into the brain (nicotine serves as an example[13] of a compound with low log *P* that does cross the BBB quite well). Another reason for the relative insignificance of log *P* could be due to the largely unknown errors in its calculation. It is not clear why Rot, the number of rotatable bonds which was the least significant when regressing on the 7-descriptor set, appears to be more significant than log *P*.

We would like to point out that the above procedure only describes the common set of parameters deemed to be important. For any given dataset some of these parameters will not be considered as important while others may get included. We have elected not to give any equation for this descriptor set as the equations are, in general, very different even in the sign of the coefficients. This is not surprising as the contribution of each descriptor (note that only the keys that are present, i.e., have a value of 1, will contribute to the equation) will be highly dependent on the dataset. A little reflection will also indicate that even giving the sign of the coefficients may not make much sense as many of the keys are not independent of one another.

Do We Really Need To Add Nonlinearities? The performance of the nonlinear networks is not significantly different from the linear ones. Do we really need to consider models with so many additional parameters? The answer appears to be an unambiguous, yes.

As we have demonstrated in the drugs versus nondrugs work,[2] a Bayesian procedure is better at avoiding local minima compared to most other methods. Despite this, local minima are a problem and care must be taken when making predictions. As an example of the importance of nonlinearities consider the following. One of the training sets contains 289 compounds that has a four-member ring (ISIS key 8). 15 of these are CNS-active and the rest are inactive. One simple local minima trap (for a linear network) could be to classify all compounds that have a four-member ring to be CNS-inactive, resulting in only a 5% error (well below the 20% error in predictions). The results however are interesting and given in Table 8. The linear network is able to generate a slightly more complicated representation by correctly assigning 20% of the active compounds and maintaining a 100% accuracy on the inactives. The network with five hidden units (BNN5) performs much better by correctly assigning 94% of the active compounds correctly and missing out on only 1 of the inactives. That the improvement in training set results are not overlearning (i.e., a new irrelevant local minima) due the presence of a much larger number of parameters is evident from the predictive performance in Table 8. The conclusion is that even though the addition of nonlinearities only affords a small overall improvement in predictive performance it is very important in predicting "outliers" correctly.

Designing a CNS-Active Virtual Library. Earlier work in our group by Bemis and Murcko[21] enumerated the scaffolds found repeatedly in known drugs. This work has been extended to identify side chains and also frameworks containing information about side chain substitution patterns (G. Bemis, unpublished). For testing our ability to generate a CNS-active library amenable to combinatorial chemistry we performed the following study. The 100 most frequent scaffolds and the 300 most frequent side chains were combined in a random but chemically reasonable manner to produce a set of approximately 1 million molecules. These molecules were then REOS'ed[30] to remove from consideration the ones that had reactive functional groups and other "undesirable" moieties, along with restrictions based on molecular weight, log *P* etc. (see Methods 2). It is reasonable to expect that molecules built using scaffolds and side chains commonly found in drugs provide a good starting point for generating "drug-like" molecules. The script for generating and REOS'ing takes about 10 molecules/s on an 195 MHz, SGI with an R10,000 processor. The script for predictions goes through 10 molecules/s on the same R10,000 processor. Roughly 25% of the molecules are classified as CNS-active and about 7% of the molecules are predicted to be active with very high degree of confidence.

Analysis of the compounds generated by this procedure produced several classes of molecules that would be amenable to combinatorial chemistry based library generation. Figure 1 shows 40 simple aromatic carboxamides from the high probability CNS-active set of molecules that could be broken down in a retrosynthetic manner into aromatic acids and amines. These components could then be used in a forward sense to experimentally generate a combinatorial chemistry library of potentially CNS-active compounds.

The 40 molecules shown are composed of a set of 27 unique aromatic carboxylic acids and 7 unique amine nucleophiles. Recombination of these components could generate a library of 189 compounds that would be available via relatively simple chemistry.

The combination of library generation and detailed analysis of the structures produced could be used to design libraries of many different sizes and different chemical classes—we show this library only as an example.

Conclusion

We have generated examples of CNS-actives and -inactives based on therapeutic classes. We have shown that it is possible to distinguish between these classes of molecules (with about 80% predictive accuracy) on a

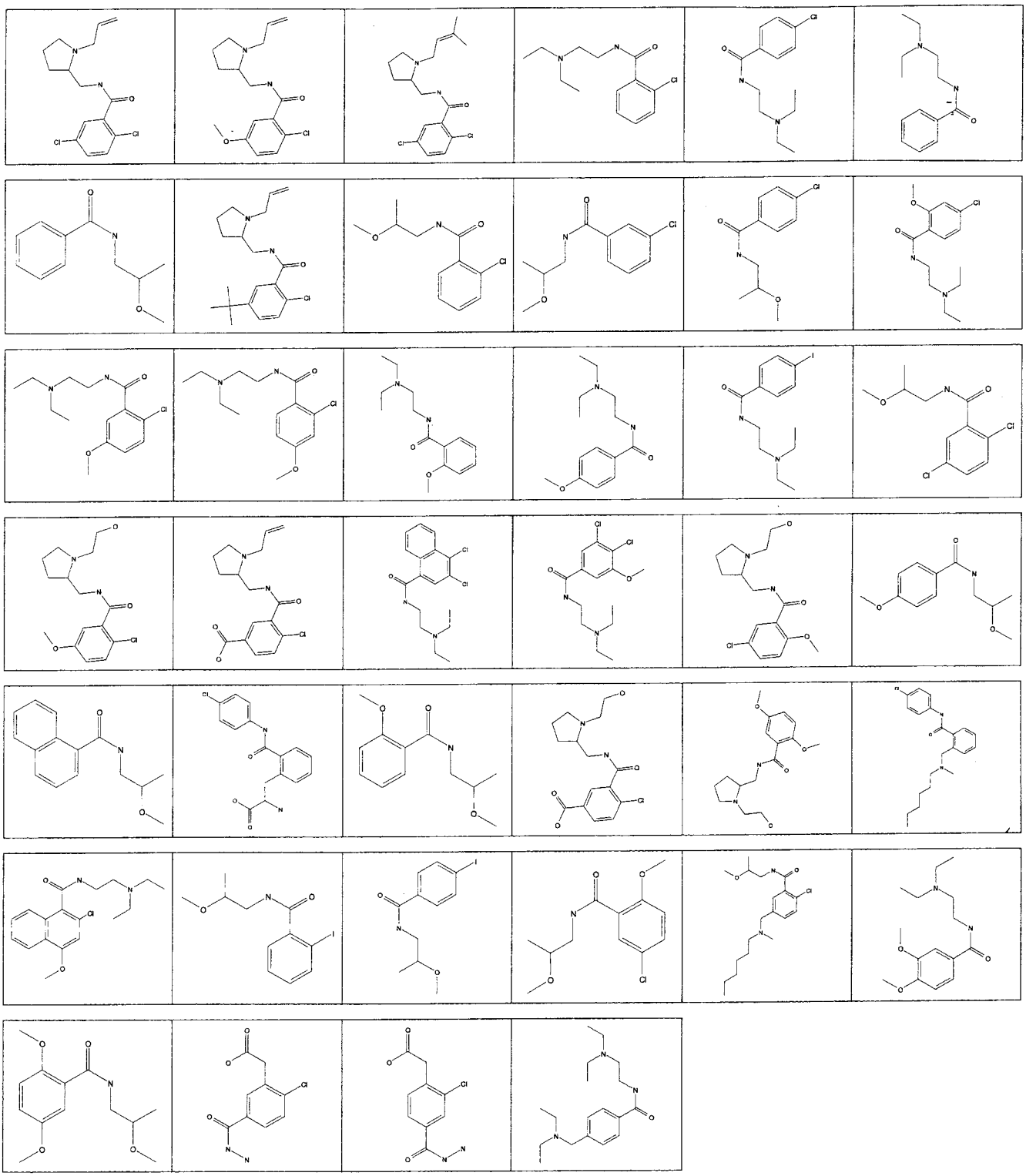

Figure 1. Set of 40 structures selected from the set classified as CNS-active with a very high degree of confidence. We selected the subset of aromatic carboxyamides because they can be broken down retrosynthetically into substituted aromatic acids and amines. These constituents could be used to create a 27 × 7 combinatorial chemistry amenable library.

wide range of structural classes. As noted earlier a therapeutic class based classification scheme is error-prone. We have shown that our modeling approach, BNN, is capable of handling data with classification noise. This was demonstrated by (1) examining the consistency or otherwise of the equations built by the models, (2) examining parameters in the models and their consistency with "received wisdom" in the field, (3) predicting on a subset of CNS-active compounds with known animal data in the MDDR database, and (4) constructing the largest published (to our knowledge) database of CNS-active and CNS-inactive compounds from the literature. The serendipitous identification of "new" CNS-active therapeutic targets (Table 7) also strengthens our belief in the validity of the models. These findings also suggest that our models are "close to" the models that can be built on corresponding nonnoisy data.

As in our previous work[2] we have shown that confidence levels on predictions can be enhanced by model-averaging,[32] i.e., combining predictions from the 20 models built. It is likely that such model-averaging will work better if the individual models use different descriptor sets. We are currently working on enhancing predictability using this and other related approaches. There is some indication that our methodology correctly predicts actively transported molecules (see, e.g., amantadine, baclofen, diphenhydramine, fentanyl, and valproic acid, which is partly carrier-mediated, in Table 1 in Supporting Information). However, more work needs to be done to study this question thoroughly.

A comment on novel methodology development for dealing with classification noise data models is relevant here. As is well-known, high-throughput screens for, say biological activity, produce experimental results that contains data with varying degrees of classification noise. Methods need to be developed to handle these kinds of errors while retaining predictability. We have shown that our learning methodology is capable of handling a limited amount of errors (details will be reported elsewhere).

One of the striking results of our work is the remarkable predictivity of the models built using the 7 1D descriptors. It is possible to rationalize the utility of hydrogen bond acceptor and donor counts in terms of their relationship with solubility and permeability, respectively. It is intuitively reasonable to expect log *P* to be a relevant descriptor but further work is needed to delineate its importance. Another important 1D descriptor that merits study (not done in this work) is the surface area (polar or total) of the molecule. The relevance and importance of 2D topological descriptors such as the ones used in this work also require more analysis.

This study is a departure from standard QSAR-type models built for the study of CNS activity in that we use a therapeutic use-based classification. This approach is open to criticism, and for this reason we have set-up and explored the performance of our models in a lot of detail (see the Methods section). The assignment of analgesics as CNS-active is another potential source of misclassification. In an attempt to explore this issue, we trained networks after completely eliminating all analgesics from our database; i.e., they were not used for either training or testing. Our results (not shown) show that there is no effect on predictive performance of the networks (BNN0/BNN5, 1D/2D descriptors) and the weights in the BNN0 plus 1D models are very close to the original model.

The major arena for application of the models developed in this work is in designing combinatorial libraries that would show preferential CNS activity. We anticipate this to be an extremely useful method in constraining the size of combinatorial libraries. In general, this filter will be most useful in conjunction with the drug/nondrug filter reported by us in previous work.[2] The possibility of predicting CNS activity based on structural and other easily calculatable information will help speed the development of new CNS-active drugs.

Acknowledgment. We thank Paul Charifson for help with the definition of CNS-active molecules and Pat Walters for help with generating the 7-descriptor set. We also thank Ulf Norinder for providing us with a computer-readable file of his dataset.

Supporting Information Available: Frameworks, predictive performance, and important descriptors as mentioned in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Johnson, M. A.; Maggiora., G. M. *Concepts and applications of molecular similarity*; Wiley: New York, 1990.
(2) Ajay; Walters, W. P.; Murcko, M. A. Can we learn to distinguish between "Drug-like" and "Nondrug-like" molecules? *J. Med. Chem.* **1998**, *41*, 3314−3324.
(3) Sadowski, J.; Kubinyi, H. A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* **1998**, *41*, 3325−3329.
(4) Comprehensive Medicinal Chemistry Release 94. 1 is available from MDL Information Systems Inc., San Leandro, CA 94577. An electronic database of vol. 6 of *Comprehensive Medicinal Chemistry*, Pergammon Press, 1990, contains drugs already in the market.
(5) MACCS−II Drug Data Report is available from MDL Information Systems Inc., San Leandro, CA 94577. An electronic database version of the Prous Science Publishers journal *Drug Data Report*, extracted from issues starting mid-1988, contains biologically active compounds in the early stages of drug development.
(6) Burton, P. S.; Conradi, R. A.; Ho, N. F.; Hilgers, A. R.; Borchardt, R. How structural features influence the biomembrane permeability of peptides. *J. Pharm. Sci.* **1996**, *85*, 1336−1340.
(7) Conradi, R. A.; Hilgers, A. R.; Ho, N. F.; Burton, P. S. The influence of peptide structure on transport across caco-2 cells. *Pharm. Res.* **1991**, *8*, 1453−1460.
(8) Abbott, N. J.; Romero, I. A. Transporting therapeutics acroos the blood-brain barrier. *Mol. Med. Today* **1996**, March, 106−113.
(9) Spector, R. Drug transport in the Central Nervous System: Role of carriers. *Pharmacology* **1990**, *40*, 1−7.
(10) Pardridge, W. M. CNS drug design based on principles of blood-brain barrier transport. *J. Neurochem.* **1998**, *70*, 1782−1792.
(11) Lloyd, E. J.; Andrews, P. R. A common structural model for central nervous system drugs and their receptors. *J. Med. Chem.* **1986**, *29*, 453−462.
(12) Levin, V. A. Relationship of octanol/water partition coefficient and molecular weight to rat brain capillary permeability. *J. Med. Chem.* **1980**, *23*, 682−684.
(13) Young, R. C.; Mitchell, R. C.; Brown, T. H.; Ganellin, C. R.; Griffiths, C. R.; Jones, M.; Rana, K. K.; Saunders: D. Development of new physicochemical model for brain penetration and application to the design of centrally acting H_2 receptor histamine antagonists. *J. Med. Chem.* **1988**, *31*, 656−671.
(14) Seelig, A.; Gottschlich, R.; Dervent, R. M. A method to determine the ability of drugs to diffuse through the blood-brain barrier. *PNAS* **1994**, *91*, 68−72.
(15) Abraham, M.; Chadha, H. S.; Mitchell, R. C. Hydrogen-bonding. Part 36. Determination of blood-brain distribution using octanol−water partition coefficient. *Drug Des. Discovery* **1995**, *13*, 123−131.
(16) Calder, J. D.; Ganellin, C. R. Predicting the brain-penetrating capability of histaminergic compounds. *Drug Des. Discovery* **1994**, *11*, 259−268.
(17) Lombardo, F.; Blake, J. F.; Curatolo, W. J. Computation of brain-blood partitioning of organic solutes via free energy calculations. *J. Med. Chem.* **1996**, *39*, 4750−4755.
(18) Norinder, U.; Sjöberg, P.; Österberg, T. Theoretical calculation and prediction of brain-blood partitioning of organic solutes using molsurf parametrization and pls statistics. *J. Pharm. Sci.* **1998**, *87*, 952−959.
(19) van de Waterbeemd, H.; Camenisch, G.; Folkers, G.; Chretien, J.; Raevsky, O. Estimation of blood-brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. *J. Drug Target.* **1998**, *6*, 151−165.
(20) Fischer, H.; Gottschlich, R.; Seelig, A. Blood-brain barrier permeation: molecular parameters governing passive diffusion. *J. Membr. Biol.* **1998**, *165*, 201−211.
(21) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. molecular frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.
(22) Ajay. Unpublished results.
(23) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple method of calculating octanol/water partition coefficient. *Chem. Pharm. Bull.* **1992**, *40*, 127−130.
(24) Brown, R. D.; Martin, Y. C. Use of structure−activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572−584.

(25) Neal, R. M. An improved acceptance procedure for the hybrid Monte Carlo algorithm. *J. Comput. Phys.* **1994**, *111*, 194−203.
(26) Buntine, W. L.; Weigand, A. S. Bayesian back-propagation. *Complex Systems* **1991**, *5*, 603−643.
(27) Tierney, L. Markov chains for exploring posterior distributions (with discussion). *Ann. Stat.* **1994**, *22*, 1701−1762.
(28) *Drug Facts and Comparisons*; St. Louis, MO, 1996.
(29) *Physicians' Desk Reference*; Medical Economics Co. Inc.: Oradell, NJ, 1990.
(30) Lien, E. J. Structure−activity relationships and drug disposition. *Annu. Rev. Pharmacol. Toxicol.* **1981**, *21*, 31−61.
(31) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening−an overview. *Drug Discuss. Today* **1998**, *3*, 160−178.
(32) Ajay. On better generalization by combining 2 or more models−a quantitative structure activity relationship example using neural networks. *Chem. Intell. Lab. Sys.* **1994**, *24*, 19−30.
(33) Pardridge, W. M.; Sakiyama, R.; Fierer, G. Blood-brain barrier transport and brain sequestration of propranolol and lidocaine. *Am. J. Physiol.* **1984**, *247*, R582−R588.
(34) Skinner, J. E. Neurocardiology shows that the central, not peripheral, action of propranolol reduces mortality following acute coronary artery occlusion in the conscious pig. *Integr. Physiol. Behav. Sci.* **1991**, *26*, 85−97.

JM990017W