Prediction of CNS Activity of Compound Libraries Using Substructure Analysis

Ola Engkvist,*,# Paul Wrede, and Ulrich Rester†

CallistoGen AG, Neuendorfstrasse 24b, D-16761 Hennigsdorf, Germany

Received June 18, 2002

An *in silico* ADME/Tox prediction tool based on substructural analysis has been developed. The tool called SUBSTRUCT has been used to predict CNS activity. Data sets with CNS active and nonactive drugs were extracted from the World Drug Index (WDI). The SUBSTRUCT program predicts CNS activity as good as a much more complicated artificial neural network model. SUBSTRUCT separates the data sets with approximately 80% accuracy. Substructural analysis also shows surprisingly large differences in substructure profiles between CNS active and nonactive drugs.

INTRODUCTION

The rapid development of combinatorial chemistry and high throughput screening (HTS) has increased the number of active hits. However, due to ADME/Tox liabilities, it is difficult to turn most of the new hits into drugs. Many promising lead candidates fail in the late stages of drug development due to poor pharmacokinetical properties.[1,2] A plethora of in vitro ADME/Tox screens has been implemented in the pharmaceutical industry during the last years with the aim of eliminating compounds, which are likely to fail in the later stages of development. Drug candidates are evaluated both for intestinal absorption properties and for susceptibility to metabolic transformations. As valuable as these experimental filters can be they do have limitations, e.g. they require physical samples for testing, which is time-consuming and costly.[3]

Thus there is a great demand for fast and reliable *in silico* ADME/Tox prediction tools. Computational methods can be applied to virtual libraries, permitting rapid and cost-effective elimination of poor candidates even prior to synthesis. Drug likeness, human intestinal absorption, and blood-brain barrier penetration have recently been reviewed.[4,5] In several studies supervised Artificial Neural Networks (ANN) and decision tree models have shown to give excellent results for drug-likeness.[6−8] Two databases one with drugs and another with nondrugs have been separated with approximately 80% accuracy.

Blood-Brain Barrier (BBB) penetration is of crucial importance for CNS active drugs. Poor penetration means low bioavailability of the drug in the brain. It is therefore not surprising that a lot of work has been done to predict BBB penetration *in silico*. In several studies a small number of compounds, with known BBB penetration capacity, have been used to develop a Quantitative Structure−Property relationship (QSPR) model for this property. For a detailed review see ref 5. Ajay et al. modeled CNS activity directly by extracting data from a database with known drugs.[9] From the database two data sets are extracted with CNS active and nonactive drugs, respectively. This procedure results in data sets covering a large part of the chemical space. This is important in the lead discovery stage of drug development, since a highly diverse set of molecules will be investigated.

Pharmacokinetical properties of a molecule depend on its chemical composition, i.e., chemical (atom types) and topological information (atom connectivity). Thus it should be possible by analyzing molecular fragment (substructure) distributions of data sets of molecules with known properties to assign a molecule to a certain class according to its specific fragment profile. Substructure analysis has a long history; in 1974 Cramer et al. published a study correlating biological activity with the presence of certain substructures.[10] This work has inspired others groups to perform similar studies.[11−14] However, CNS activity has so far not been investigated with substructure analysis.

The main drawback of supervised ANN is that it is very difficult to interpret the results. Interpretation and understanding of the pattern responsible for certain properties cannot be derived from the optimized parameters of a multilayer feed-forward ANN. To overcome this problem in classification with ANN the tool SUBSTRUCT was developed. The algorithm of SUBSTRUCT is based on substructure analysis and can be used for both classification and selection of structural information to improve the ADME/Tox properties of a given molecule. Detailed comparison of the classification performance of the substructure analysis tool with ANN based on the same data set will be presented. Knowledge about substructure frequencies is important in the design of new CNS active drugs, since the medicinal chemist can estimate the effect of a certain modification of a drug candidate.

Extraction of CNS Active and Nonactive Drugs from the World Drug Index (WDI). CNS active and nonactive drugs were extracted from the World Drug Index (WDI).[15] The molecules in the WDI were divided into three parts: i) drugs that must pass the blood-brain barrier for their action (e.g. psychosedatives, psychostimulants, antidepressants), ii)

* Corresponding author phone: +49 89 740 165 0; fax: +49 89 740 165 20; e-mail: engkvist@axxima.com.

Present address: Axxima Pharmaceuticals, Am Klopferspitz 19, D-821 52 Martinsried, Germany.

† Present address: Merck KGaA, Preclinical R&D, Frankfurter Str. 250, D-64293 Darmstadt, Germany.

10.1021/ci0102721 CCC: $25.00
Published on Web 12/24/2002

Table 1. Drug Classes Selected as CNS Active (CNS+)[a]

psychosedatives (1134)
psychostimulants (820)
antidepressants (681)
anticonvulsants (661)
neuroleptics (659)
tranquilizers (589)
nootropic (346)
opioids (329)
narcotics (324)
anorectics (201)
MAO−inhibitors (195)
benzodiazepine-agonists (113)
gabaminergics (104)
gaba-antagonists (78)

[a] The number of compounds found in each class is given within the parentheses. Note that many compounds are assigned to more than one class.

drugs that might pass the blood-brain barrier (e.g. antiemetics and antiparkinsonians), and iii) drugs that most likely do not pass the blood-brain barrier (e.g. cytostatics, antibiotics, and hypertensives). Fourteen classes of CNS active compounds have been selected comprising 3892 CNS active drugs. The 14 drug classes are given in Table 1. The data set was further filtered resulting in 3678 molecules used for SUBSTRUCT analysis and supervised artificial neural network (ANN) training.[16] The data set of drugs that might show CNS activity comprises 6218 compounds. These compounds were excluded from the classification. As a negative reference data set, 5000 CNS nonactive drugs were randomly selected out of 50 025 molecules from the WDI. The data set of 3678 molecules (CNS+) and 5000 molecules (CNS−) was used in the ANN model and SUBSTRUCT classification.

The SUBSTRUCT Program. Substructure analysis is an important method in drug discovery. It has been used to understand and predict molecular properties since the pioneering work by Cramer et al.[10] However, most studies have used predefined substructure fragments in their analysis.[11−14] In this article we pursue a new approach based on an exhaustive enumeration of molecular fragments up to a given user defined size, i.e. no predefined scheme for fragmenting the molecule. For this reason no important structural information regarding differences in chemical composition can be lost. In this study fragments up to size 4 have been used; however, for other applications a different maximum fragment size might be optimal. The SUBSTRUCT program analyses data sets by calculating the frequency of all possible fragments of the molecules resulting in a frequency profile, ready for classification. Several algorithms for exhaustive fragment identification have been published.[17,18] However, none of these methods have been developed with the intention of classification of ADME/Tox data sets. While Rücker et al.[18] developed a pure graph theoretical approach, SUBSTRUCT takes the protonation state and the number of hydrogen for each atom into account. Inclusion of the protonation state is motivated by its strong effect on physicochemical properties[1−3] and is calculated by an in-house program.[19] The substructure analysis program SUBSTRUCT has been implemented as a C program.

To classify new molecules with unknown CNS activity, data sets with known CNS active (CNS+) and nonactive (CNS−) molecules were derived as described above from the WDI. The molecules of both data sets are fragmented into all possible fragments up to a user defined fragment size (default fragment size: between 1 and 4 atoms), and the frequency profiles of the two data sets are calculated and compared to each other. The exact fragmentation procedure is given in the appendix. The fragment frequency is defined as the proportion of the molecules in the data set containing a specific fragment. To avoid a strong influence of fragments with low frequency, all fragments are excluded that occur in less than 0.5% of the molecules in both data sets. An advantage of the proposed method is that all substructures will be calculated, thus it is impossible to miss important differences in their substructure frequencies.

For example, one data set with CNS+ and one with CNS− molecules are fragmented, and the *M*olecular *S*um *O*ver all *F*ragments (MSOF) of an unknown molecule is calculated according to

$$MSOF = \sum_{fragments} \frac{F_{CNS+}}{F_{CNS+} + F_{CNS-}} \quad (1)$$

where F_{CNS+} and F_{CNS-} is the fragment frequency for each fragment of the molecule in the CNS+ and CNS− data set, respectively. MSOF ranges between zero and one. If MSOF is above 0.5, the molecule is more similar to the molecules in the CNS+ data set; otherwise it is more similar to the CNS− data set.

Artificial Neural Network (ANN) Preparation. For comparison purposes we have trained an ANN to separate CNS+ and CNS− compounds. Ghose and Crippen established a system of atom types for predicting AlogP.[20] Furthermore, the atom types have successfully been used to predict drug-likeness.[6] In the present study, a subset of 92 out of these 120 atom types were selected as molecular descriptors. The selection criterion was that the descriptors should be present in at least 0.5% of the molecules in the CNS+ and CNS− data sets. For all ANN calculations the SNNS program package was used.[21] The feed forward neural nets consisted of 92 input neurons, five hidden neurons, and one output neuron. All layers were fully connected, resulting in a total of 471 weights including bias terms. The backpropagation momentum method was applied for training. The training was performed over 4000 cycles with a learning rate of 0.2 and a momentum term of 0.1. For technical reasons all input and output values were linearly scaled to values between 0.1 and 0.9. During each epoch the training patterns were presented to the network in randomized order.

The extracted data set of 3678 CNS+ and 5000 CNS− molecules were used as training and validation set. The data set was randomly divided into 10 parts. Cross-validation was performed by using 9 parts as training set and the remaining part as validation set. All 10 training and validation calculations were repeated 10 times, and the average result was taken.

Substructure Analysis CNS Data Sets by SUBSTRUCT. Information about substructures frequencies of CNS active drugs is very useful in lead discovery, combinatorial library design, and lead optimization. Occurrence of interesting functional groups in CNS+ and CNS− drugs are presented in Table 2. It can be seen that a lot of substructures have considerably different frequencies in the two data sets. Protonated nitrogens are much more frequent among CNS+

Table 2. Substructure Analysis of the Frequencies (in Percent) for Some Interesting Fragments of the CNS+ and CNS− Drugs

Fragment	CNS+ (%)	CNS- (%)	CNS+/CNS-
	43.7	14.7	2.97
	37.9	24.9	1.52
	12.2	6.2	1.97
	30.9	34.3	0.90
	11.6	6.4	1.81
	10.4	10.7	0,97
	10.8	19.4	0,56
R−OH	28.3	50.0	0,57
R−F	12.4	5.2	2,38
R−Cl	19.8	9.0	2,2
N(aromatic)	17.6	20.0	0,88
	39.8	58.0	0,68
C(aromatic)	91.2	70.6	1,29

drugs. It is true for protonated tertiary, secondary, and primary amines. Protonated tertiary amines are three times more present among the CNS+ drugs, while primary and secondary are twice as common. Nonprotonated tertiary amines are 50% more common among CNS+ drugs, while secondary and primary are slightly more present in CNS− drugs. Carboxylic acids, alcohols, ethers, and esters are more frequent among CNS− drugs. Over 90% of all CNS+ drugs contain at least one aromatic ring compared to 70% of the CNS− drugs. The substructure analysis also shows that CNS− drugs have slightly higher proportion of aliphatic carbons. It has been suggested that the combination of tertiary nitrogen and an aromatic ring system will increase the CNS activity.[22] Our study supports that conclusion. It can be seen that indeed both tertiary nitrogen and aromatic rings are more frequent among the CNS+ drugs. It is also noted that the probability of being a CNS+ drug increases if the tertiary nitrogen is protonated. It is noted that the prediction of the protonation state is rather crude;[19] however, the trend is clear: Protonated nitrogen are significantly more frequent among CNS+ drugs. Also halogens contribute to CNS activity, both fluorine and chlorine have twice as high frequency among CNS+ drugs compared with CNS−, while sulfur and phosphorus atoms are as frequent in CNS+ as in CNS− drugs. CNS activity is a complex process that is not

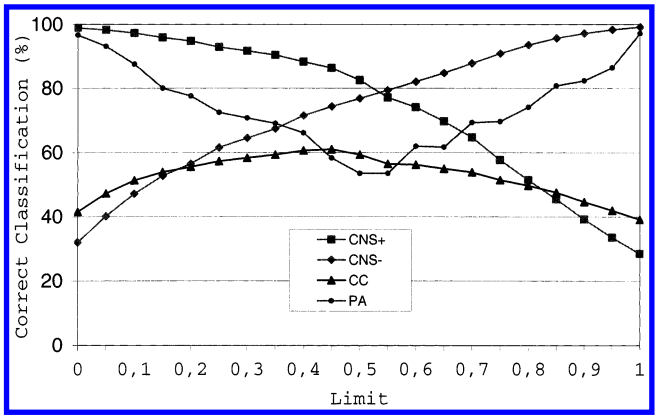

Figure 1. Result of the ANN classification of CNS activity. The correct classification rate is given as a function of the limit separating CNS+ and CNS− molecules, respectively. E.g. if the limit is 0.5 all molecules with a score lower than 0.5 are classified as CNS− and all molecules above are classified as CNS+. The correlation coefficient according to Matthews (CC) multiplied with 100 and the prediction accuracy (PA) for the compounds within a specific scoring interval is included in the figure.

fully understood today. It is therefore very difficult to make an interpretation why certain fragments seem to enhance CNS activity and others do not. However, a lot of tertiary nitrogen has a pK_a value close to 7, meaning that they can easily be deprotonated. Deprotonation might enhance BBB penetration, since a neutral molecule will diffuse easier through lipophilic membranes. A CNS− molecule might pass the blood-brain barrier. A model that predicts CNS activity can also predict blood-brain barrier penetration well (see next section and ref 9). Thus there exists a correlation between CNS activity and BBB penetration. From a pharmacological point of view it is also favorable if a CNS− drug does not pass the blood-brain barrier since the risk of side effects will be reduced. There might also be a higher frequency of negatively charged amino acids in the receptors situated in the brain, compared to most other tissues, which can interact favorably with positively charged ligands.

Prediction of CNS Activity by ANN and SUBSTRUCT. CNS activity prediction was performed both with ANN and substructure analysis. The prediction accuracy of the validation set of the ANN is given in Figure 1. The ANN classifies 82.5% of the CNS+ and 76.5% of the CNS− molecules correctly. This result is comparable to those obtained by other methods.[4,7]

One advantage of the SUBSTRUCT program is the capability to analyze CNS activity data. As described above, 3678 CNS active (CNS+) and 5000 nonactive (CNS−) compounds were selected from the WDI. Both data sets were fragmented with the default fragment sizes of one up to four atoms. For the CNS+ and CNS− molecules 7417 and 10779 different fragments were generated. All fragments present in at least 0.5% of the molecules in either one of the two data sets were kept for further analysis (in total 1724 fragments). For all molecules in the data sets fragments were generated. The classification result is presented in Figure 2. 83.3% and 71.2% of the CNS+ and CNS− molecules are correctly classified, respectively. To test the validity of the classification with fragment distributions constructed from all the molecules, we divided both data sets into 10 parts and used 90% of the data for calculating fragment distributions. The missing part was used as a validation set. Thus

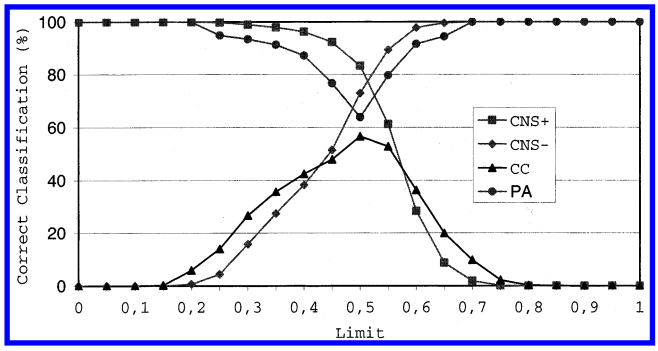

Figure 2. Result of the SUBSTRUCT classification of CNS activity. The correct classification rate is given as a function of the limit separating CNS+ and CNS− molecules, respectively. The correlation coefficient according to Matthews (CC) multiplied with 100 and the prediction accuracy (PA) for the compounds within a specific scoring interval is included.

Table 3. Overlap in Classification between ANN and SUBSTRUCT

		overlap between the methods (%)	overlap if not correlated (%)
CNS+	total	85	68
	correct classification	92	85
	wrong classification	55	20
CNS−	total	79	67
	correct classification	86	79
	wrong classification	59	21

10 calculations were performed. The same division of the data set as for the ANN was used. The results from 10 calculations did not deviate significantly and are also comparable with the results of a classification with fragment distributions constructed from the complete data set. To further analyze the data the correlation coefficient according to Matthews (CC) was calculated[23]

$$CC = \frac{PN - OU}{\sqrt{(N + U)(N + O)(P + U)(P + O)}}$$

where **P** is the number of positive correct predictions, **N** is the number of negative correct predictions, **O** is the number of false-positive predictions (overprediction), and **U** is the number of false-negative predictions (underprediction). Matthews's correlation coefficients are displayed in Figures 1 and 2. For the ANN maximal CC is 0.61 and for SUBSTRUCT 0.57. The prediction accuracy (PA) for compounds within a specific scoring range was also calculated (Figures 1 and 2). As expected the lowest prediction accuracy is for compounds with a score around 0.5.

As shown in Table 3 the predictions of SUBSTRUCT and ANN are correlated, i.e., if a molecule is classified as CNS+ with the ANN then it is also likely to be classified as CNS+ by the substructure analysis. The result is averaged over all cross-validation calculations described above for the ANN and SUBSTRUCT. Both ANN and substructure analysis have their advantages and disadvantages. The ANN is faster than SUBSTRUCT in classifying unknown molecules making it more suitable in conjunction with virtual synthesis[24] or virtual screening[25] applications. Classification of 1000 molecules takes 2 s with ANN and 18 s with SUBSTRUCT on a Pentium III (1 GHz) processor. One main drawback with ANN is that is a "black box", and it is impossible to identify the crucial features in the model. In contrast the advantages of the SUBSTRUCT analysis are as follows: i) interpretation of the results is possible ("transparent box"), ii) detection of differences in chemical composition and property of data sets, iii) collection of relevant substructure information concerning certain ADME/Tox properties, iv) generation of ideas how to modify new chemical entities for improving ADME/Tox properties, v) no time-consuming training step is required, and vi) very easy to include further experimental data when available. Thus in our opinion the methods complements each other and both can be used for *in silico* ADME/Tox classifications.

Table 4. Comparison of SUBSTRUCT and ANN Prediction with Experimentally Measured BBB Penetration Data[c]

compound	exp. BBB penetration	SUBSTRUCT	ANN
apomorphine	+	+ (0.57)	+ (0.75)
flupentixol	+	+ (0.71)	+ (0.88)
haloperidol	+	+ (0.65)	+ (0.89)
naltrexone	+	+ (0.64)	+ (0.90)
perphenazine	+	+ (0.75)	+ (0.88)
promethazine	+	+ (0.73)	+ (0.85)
thiopental	+	+ (0.51)	+ (0.53)
tamitinol	+	+ (0.51)	− (0.02)
cotinine	+	+ (0.58)	+ (1.01)
nicotine	+	+ (0.63)	+ (0.89)
nornicotine	+	+ (0.59)	+ (0.83)
astemizole	−	+ (0.57)	+ (0.61)
domperidone	−[a]	+ (0.64)	+ (0.55)
loperamide	−[a]	+ (0.59)	+ (0.73)
terfenamide	−[a]	+ (0.55)	+ (0.59)
atenolol	−	− (0.24)	− (−0.12)
mequitazine	−[b]	+ (0.71)	+ (0.92)
salbutamol	−	+ (0.53)	− (0.20)
furosemide	−	− (0.49)	− (0.20)
pirenzepine	−	+ (0.59)	+ (0.56)
ebastine	−	− (0.47)	− (0.49)
carebastine	−	− (0.44)	− (0.41)
carmoxirol	−	− (0.35)	− (−0.12)
delavirdine	−	− (0.23)	− (0.07)

[a] P-glycoprotein substrate drugs. [b] BBB penetration is dose limiting for the drug. [c] Reported BB penetration levels are taken from refs 26−28.

Twenty-four drugs for which BBB penetration rate were experimentally measured were used as a validation set for our derived models. The results are presented in Table 4. For the ANN calculations the result is the average predicted from all the nets in the cross-validation study. The standard deviation for the predictions is 0.02. For SUBSTRUCT the results are averaged over all fragment sets used in the cross-validation study. The standard deviation for the predictions is 0.02. The data set comprises 11 CNS+ compounds (log BB > -1) and 9 CNS− compounds (log BB < -1). Domperidone, loperamide, and terfenamide pass; however, they are transported back to the blood by P-glycoprotein transporters. Mequitazine can pass the barrier when applied in high concentration. All the 11 compounds with high CNS availability are predicted correctly with SUBSTRUCT, while the ANN predicts 10 of them correctly. For the compounds with low CNS availability, ANN predicts 7 out of 9 correctly, while SUBSTRUCT predicts six out of nine correctly, resulting in an overall correct prediction of 17 out of 20 compounds (85%). SUBSTRUCT and ANN predicted the three compounds, which are actively transported back, and mequitazine as CNS+. A probable reason may be common features of most CNS active drugs facilitating passive diffusion through the BBB. Thus the developed models

mainly classify molecules according to their ability of passive diffusion through the BBB. An analysis with SUBSTRUCT reveals that the molecules include both protonated tertiary nitrogen and several aromatic rings; both features are more common among CNS+ drugs.

A new feature of our substructure model is the inclusion of the protonation state in the model. To test the importance of this feature, we recalculated our models with the molecules in a neutral state. At a limit of 0.5, 79% of the CNS+ molecules were correctly classified compared with 83% with the protonation state included in the model. For the CNS− molecules 69% were correctly classified compared to 73% when the protonation state is included. The same trend is seen for the molecules with experimentally measured BBB penetration. Furosemide was predicted wrongly, when the protonation state was not taken into account. For all other compounds the prediction were identical. Thus inclusion of the protonation state improves the classification significantly.

CONCLUSIONS

In this article CNS activity of small molecules have been modeled by Artificial Neural Networks (ANN) and by the developed program SUBSTRUCT based on substructure analysis. CNS active and nonactive compounds were extracted from WDI. The models were able to separate CNS active and nonactive compounds with approximately 80% accuracy. The ANN model is faster in classifying unknown models; however, the classifications are difficult to interpret. ANN also needs a time-consuming training step. The SUBSTRUCT classification tool has complementary properties. The results are easy to interpret, and there is no need for time-consuming training. It is very easy to include new experimental data, in contrast to ANN where the whole training procedure has to be repeated. Substructural analysis shows that the fragment distribution of CNS active and nonactive compounds is different. Protonated nitrogens, aromatic rings, chlorine, and fluorine are more frequent among CNS active compounds, while the opposite is true for oxygen containing compounds.

ACKNOWLEDGMENT

The authors thank Gisbert Schneider and Holger Kalkhof for helpful discussions and encouragement.

APPENDIX

The following guidelines have been used to create fragments. An atom type is defined by its atom number, the number of bonds to heavy (non-hydrogen) atoms, the number of electrons from the atom participating in these bonds, the number of attached hydrogen atoms and its protonation state. The fragments of a molecule of size 1 are made up of all unique atom types in the molecule. For instance, 1-propanol has three fragments of size 1, methyl, methylene and hydroxyl, three fragments of size 2, methyl-methylene, methylene-methylene, and methylene-hydroxyl. And two fragments of size 3, methyl-methylene-methylene and methylene-methylene-hydroxyl. The procedure is illustrated in Figure 3. With this scheme all molecules can be fragmented up to any fragment size.

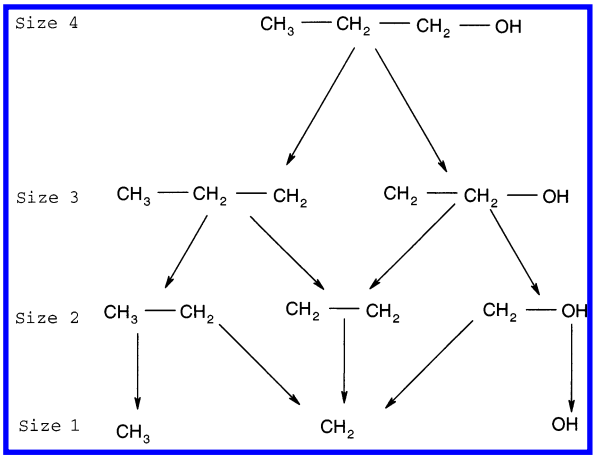

Figure 3. SUBSTRUCT fragmentation scheme of 1-propanol.

REFERENCES AND NOTES

(1) Smith, D. A.; van de Waterbeemd, H.; Walker, D. K. *Pharmacokinetics and Metabolism in Drug Design*; WILEY−VCH: 2001.
(2) *Pharmacokinetic Optimisation in Drug Research*; Testa, B., van de Waterbeemd, H., Folkers, G., Guy, R., Eds.; WILEY−VCH: 2000.
(3) Avdeef, A. Physicochemical profiling (Solubility, Permeability and Charge State). *Curr. Top. Med. Chem.* **2001**, *1*, 277−351.
(4) Podlogar, B. L.; Muegge, I. Holistic *in silico* methods to estimate the systemic and CNS bioavaliabilities of potential chemotherapeutic agents. *Curr. Top. Med. Chem.* **2001**, *1*, 257−275.
(5) Clark, D. E. Prediction of intestinal absorption and blood-brain barrier penetration by computational methods. *Comb. Chem., High Through. Screen.* **2001**, *4*, 477−496.
(6) Sadowski, J.; Kubinyi, H. A Scoring Scheme for Discriminating between drugs and nondrugs. *J. Med. Chem.* **1998**, *41*, 3325−3329.
(7) Ajay; Bemis, G. W.; Murcko, M. A. Can we learn to distinguish between "Drug-like" and "Nondrug-like" molecules? *J. Med. Chem.* **1998**, *41*, 3314−3324.
(8) Wagener, M.; van Geerestein, V. J. Potential drugs and nondrugs: prediction and identification of important structural features. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 280−292.
(9) Ajay; Bemis, G. W.; Murcko, M. A. Designing libraries with CNS activity *J. Med. Chem.* **1999**, *42*, 4942−4951.
(10) Cramer, R. D., III; Redl, G.; Berkoff, C. E. Substructural analysis. A novel approach to the problem of drug design. *J. Med. Chem.* **1974**, *17*, 533−535.
(11) Hodes, L.; Hazard, G. F.; Geran, R. I.; Richman, S. A statistical-heuristic method for automated selection of drugs for screening. *J. Med. Chem.* **1977**, *20*, 469−475.
(12) Ormerud, A.; Willet, P.; Bawden, D. Comparison of fragment weighting schemes for substructural analysis. *Quant. Struct-Act. Relat.* **1989**, *8*, 115−129.
(13) Ormerud, A.; Willet, P.; Bawden, D. Further comparative studies of fragment weighting schemes for substructural analysis. *Quant. Struct-Act. Relat.* **1990**, *9*, 302−312.
(14) Sheridan, R. P.; Nachbar, R. B.; Bush, B. L. Extending the trend vector: The trend matrix and sample-based partial least squares. *J. Comput.-Aid. Mol. Des.* **1994**, *8*, 323−340.
(15) WDI (Derwent World Drug Index), Version 2001/01 (Issue 18), Derwent Information Ltd., 2001. http://www.derwent.com/.
(16) The following rules for filter out unsuitable compounds were used: MW below 80 and above 800 were removed. The compound must contain at least one carbon and one nitrogen or oxygen or sulfur. The compound can have at most 6 Cl, Br, and I. Only molecules consisting of the organic subset are allowed, i.e., H, C, N, P, C, S, O, F, Cl, Br, and I. The following reactive groups are excluded: acylhalides, phosphanes, peroxides, isocyanates, anhydrides, acylhalides, acylcyanides, azides, cyanides, and diazonium compounds. The rules are similar to what have been applied for druglikeness (see ref 6).
(17) Bone, R. G. A.; Villar, H. O. Exhaustive enumeration of molecular substructures. *J. Comput. Chem.* **1997**, *18*, 86−107.
(18) Rücker, G.; Rücker, C. Substructure, subgraph, and walk counts as measures of the complexity of graphs and molecules. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1457−1462.
(19) A simple rule-based protonation state program was developed to predict the protonation state at physiological pH. The main rules are as follows: i) organic acids are deprotonated, ii) aliphatic nitrogen binding only to sp^3-hybridised carbons are protonated, and iii) amidine and guanidine groups are protonated.

(20) Visvandham, V. N.; Ghose, A. K.; Revankar, G. R.; Robbins, R. K. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure−activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163−172.

(21) SNNS (Stuttgart Neural Network Simulator), Version 4.2, University of Tübingen, 1998.

(22) 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.

(23) Matthews, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **1975**, *405*, 442−451.

(24) Schneider, G.; Lee, M. L.; Stahl, M.; Schneider, P. De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. *J. Comput.-Aid. Drug. Des.* **2000**, *14*, 487−494.

(25) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. "Scaffold Hopping" by topological pharmacophore search: A contribution to virtual screening. *Angew. Chem.* **1999**, *111*, 3068−3070; *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 2894−2896.

(26) Seelig, A.; Gottschlich, R.; Devant, R. M. A. A method to determine the ability for drugs to diffuse through the Blood-Brain Barrier. *Proc. Nat. Acad. Sci. U.S.A.* **1994**, *91*, 68−72.

(27) Chang, M.; Sood, V. K.; Wilson, G. J.; Kloosterman, D. A.; Sanders, P. E.; Hauer, M. J.; Zhang, W.; Branstetter, D. G. Metabolism of the HIV-1 Reverse Trancriptase inhibitor Delavirdine in mice. *Drug Metab. Dispos.* **1997**, *25*, 828−839.

(28) Crooks, P. A.; Li, M.; Dwoskin, L. P. Metabolites of nicotine in rat brain after peripheral nicotine administration. *Drug Metab. Dispos.* **1997**, *25*, 47−54.

CI0102721