Journal of
Medicinal
Chemistry

Volume 44, Number 12 June 7, 2001

Expedited Articles

Simple Selection Criteria for Drug-like Chemical Matter

Ingo Muegge,* Sarah L. Heald, and David Brittelli

Bayer Research Center, 400 Morgan Lane, West Haven, Connecticut 06516

Received January 23, 2001

A simple pharmacophore point filter has been developed that discriminates between drug-like and nondrug-like chemical matter. It is based on the observation that nondrugs are often underfunctionalized. Therefore, a minimum count of well-defined pharmacophore points is required to pass the filter. The application of the filter results in 66−69% of subsets of the MDDR database to be classified as drug-like. Furthermore, 61−68% of subsets of the CMC database are classified as drug-like. In contrast, only 36% of the ACD are found to be drug-like. While these results are not quite as good as those obtained with recently described neural net approaches, the method used here has clear advantages. In contrast to a neural net approach and also in contrast to decision tree methods described recently, the pharmacophore filter has been developed by using "chemical wisdom" that is unbiased from fitting the structural content of specific drug databases to prediction models. Similar to decision tree methods, the pharmacophore point filter provides a detailed structural reason for the classification of each molecule as drug or nondrug. The pharmacophore point filter results are compared to neural net filter results. A statistically significant overlap between compounds recognized as drug-like validates both approaches. The pharmacophore point filter complements neural net approaches as well as property profiling approaches used as drug-likeness filters in compound library analysis and design.

Introduction

High-throughput screening (HTS) and combinatorial chemistry have become cornerstones in drug discovery.[1−3] With an increasing choice of compounds that can be synthesized and screened, it becomes more important to evaluate and enhance their chances to serve as lead structures in a drug discovery program. Attempts have been made to design screening libraries, mainly by maximizing diversity, to improve the screening hit rate beyond that of random libraries.[4−7] However, the general success of these approaches has yet to be proven.[8] Independent of hit rate and diversity, the quality of the hits in an HTS screen is crucial to identify a lead compound for drug discovery. That is, a screening library should be designed such that the chance of an HTS hit to be followed up by medicinal chemistry as a lead compound is increased. A molecule with such characteristics is generally referred to as "drug-like". This term includes the synthetic accessibility of the compound and its analogues. In addition, drug-like compounds are expected to meet ADME (absorption, distribution, metabolism, excretion) and toxicology profiles.[9,10] Several efforts have been made to invent computational filter cascades that discard compounds in databases that are not drug-like.[11] A recent review by Walters and co-workers describes methods to recognize drug-likeness by simple counting methods, functional group filters, chemistry space evaluation methods, and neural networks.[12] Therefore, we mention here only briefly the most promising attempts to classify drugs and nondrugs. Lipinski and co-workers have described the perhaps most famous set of counting rules to filter out compounds likely to show poor absorption

* To whom correspondence should be addressed. Phone: (203) 812-5139. Fax: (203) 812-3505. E-mail: ingo.mugge.b@bayer.com.

10.1021/jm015507e CCC: $20.00
Published on Web 05/15/2001

properties.[13] Filters on molecular weight, logP, number of hydrogen bond donors, hydrogen bond acceptor, and similar descriptors used by others[14,15] such as number of rotatable bonds, number of rigid bonds, number of rings in a molecule, or charges per molecule mainly address possible absorption issues but cannot easily discriminate between drugs and nondrugs. A successful method to distinguish between drugs and nondrugs has been established in parallel by Ajay and co-workers[16] and Sadowski and Kubinyi[17] employing neural net approaches together with topological descriptors such as ISIS keys[18] and Ghose and Crippen atom types[19] to code the molecular structures. The neural net classification method has been found to discriminate between drug-like chemical matter represented by databases such as the CMC (Comprehensive Medicinal Chemistry),[20] the MDDR (MACCS-II Drug Data Report),[21] the WDI (World Drug Index),[22] and nondrug-like chemical matter represented by databases such as the ACD (Available Chemicals Directory).[23] Eighty to ninety percent of the compounds in those databases have been correctly classified according to their heritage in a drug or nondrug database. While this is a very encouraging result, the neural net approach has some drawbacks coming from its inability to provide discernible rules directly related to the chemical structure of the classified compounds and from its database bias that may limit its use further discussed below.

Gillet and co-workers have achieved good separation of drugs and nondrugs using a limited set of property descriptors and a genetic algorithm.[24] Ghose and coworkers have presented an iterative fitting study to quantitatively discriminate between drugs and nondrugs using the CMC and ACD databases. Ghose and Crippen atom types and additional physicochemical parameters have been used as descriptors.[25] Seventy five percent of the CMC and twenty five percent of the ACD have been classified as drug-like. Recursive partitioning methods have also been used with excellent results. Ninety-two percent of compounds of the WDI and 34% of the ACD databases have been classified as drug-like by decision tree methods.[26] Topological aspects of drug-like chemical matter have been analyzed recently by Bemis and co-workers,[27,28] Ghose and coworkers,[29] as well as Xu and Stevenson.[30] Their analyses help to understand which structural elements and functional groups are contained in drug-like molecules. However, topological approaches have not been used to discriminate between drug-like and nondrug-like chemical matter.

Neural net approaches, decision tree approaches, and Ghose's fitting approach are database-dependent. To complement these successful methods to classify drugs and nondrugs, it is desirable to identify a database-independent rule set that can classify drugs versus nondrugs. Such an expert system would give guidance to chemists for tasks such as compound acquisition and library design. The main advantage of an expert system over a neural net approach is the detailed understanding of the classification reasons for each molecule. In addition, rules can be easily customized for different tasks. Therefore, as a first step toward such an expert system, we have developed here a pharmacophore point filter that is able to separate drugs from nondrugs based on simple, database-independent rules with significant discrimination power—thereby complementing the neural net and property filter approaches.

Chart 1

Methods

Rule Base for Pharmacophore Point Filter. A set of simple rules is established to classify molecular structures. Four functional motifs are defined to be important in drug-like molecules (Chart 1). The occurrence of these functional motifs guarantees hydrogen-bonding capabilities that are essential for specific drug interactions with its targets. These functional groups can be combined to what we refer to here as pharmacophore points. These pharmacophore points include the following functional groups: amine, amide, alcohol, ketone, sulfone, sulfonamide, carboxylic acid, carbamate, guanidine, amidine, urea, and ester. Since we try to capture pharmacophore points that potentially provide key interactions with the target protein, other functional groups such as, e.g., nitro and imine are excluded. A nitro group, for instance, is rarely seen in a hydrogen-bonding functionality. More often it plays the role of a substituent that fine-tunes the physicochemical properties of drug molecules (nitro groups occur only in 2.4% (MDDR) to 2.7% (CMC) of drug candidates while they are much more frequent in reagent type databases (nitro groups occur in 7.9% of compounds in the ACD)). In principle, heterocyclic groups such as pyridyl or other functional groups such as nitrile could have been included in the pharmacophore points. It has been found, however, that those groups are somewhat indifferent toward distinguishing between drugs and nondrugs. Pharmacophore points are fused and counted as one when their heteroatoms are not separated by more than one carbon atom. The idea of a drug/nondrug classification is now based on the observation that nondrug molecules are often underfunctionalized. Therefore, a molecule with less than two pharmacophore points fails the filter. In addition, molecules with more than seven pharmacophore points fail also due to overfunctionalization.

Beyond these main rules, the following additional rules are applied:

1. Primary, secondary, and tertiary amines are considered pharmacophore points but not pyrrole, indole, thiazole, isoxazole, other azoles or diazines.
2. Compounds with more than one carboxylic acid are dismissed.
3. Compounds without a ring structure are dismissed.
4. Intracyclic amines that occur in the same ring are fused (e.g., piperazine), i.e., they count as only one pharmacophore point

A pharmacophore filter has been built by the above rules. It is called pharmacophore filter 1 (PF1) below. One problem with the requirement of two pharmacophore points is that small CNS-active drugs have only one pharmacophore point and therefore fail PF1. Thus we define here a pharmacophore filter 2 (PF2) that differs from PF1 in that it allows for compounds with only one pharmacophore point to pass the filter provided it is of type carboxylic acid, amine, guanidine, or amidine.

For a set of six compounds that all have only one pharmacophore point, Figure 1 illustrates some classification rules. Compound **1** contains an amide and a urea that are fused since the nitrogen atoms are separated by only one carbon. A thiourea is also recognized but not counted. Compound **2** has only an amine as a pharmacophore point since nitriles, pyridines, and imines do not count. Also a nitro group (compound **3**) does not count as a pharmacophore point. Compound **4** fuses a thiourea with a carbonyl oxygen to one pharmacophore. The morpholine and thiophene groups are not recognized as pharmacophore points. In compound **5**, a sul-

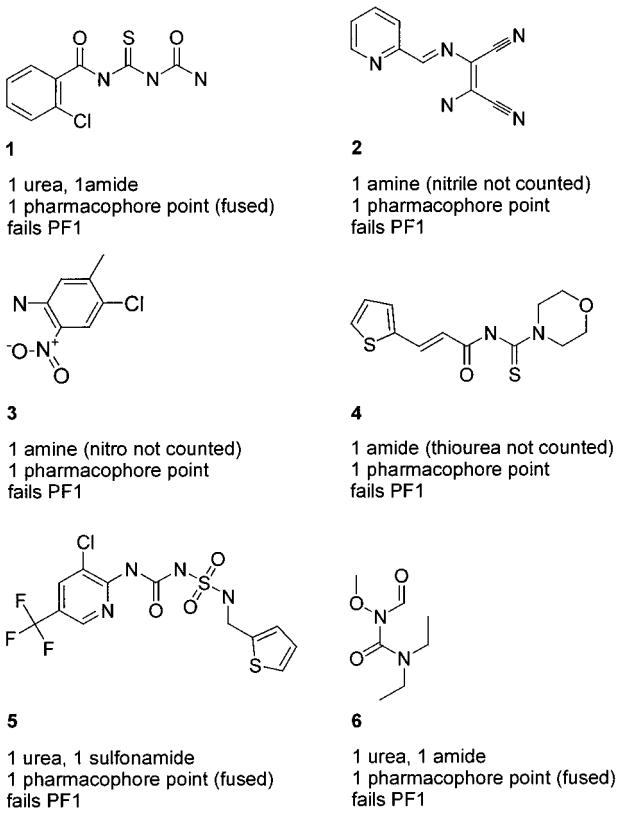

Figure 1. Examples of ACD compounds with one pharmacophore point count only.

fonamide is fused with a urea. In compound **6**, all heteroatoms are fused to one pharmacophore. In addition, compound **6** would be also dismissed due to a missing ring topology in the structure. Note that thioureas in compounds **1** and **4** are recognized but not counted. The implementation of the pharmacophore filter allows for the optional dismissal of thioureas, thioamides, and compounds with other structural motifs. However, these additional rules are not applied here because PF1 and PF2 are not meant to replace purge programs for reactive compounds.

Preparation of Databases. For test purposes of the pharmacophore point filter, we use the ACD as the nondrug database and the CMC and the MDDR as the drug databases. To further characterize the drug databases, subsets of the CMC and MDDR have been extracted according to the combined recommendations of Bemis and Murcko[27] and Lipinski and co-workers[13] and analyzed separately. All databases have undergone a filter cascade (Table 1) that excludes the following compounds from each database: compounds with missing or invalid structures; compounds with atoms other than C, N, O, S, H, P, Si, Cl, Br, F, I; reactive or otherwise not suited compounds. In addition, duplicates with CMC and MDDR have been removed in the ACD. Duplicates with MDDR have been removed in the CMC. To obtain single-record entries, counterions and solvent molecules have been removed from the structures. Self-duplicates have also been removed in each database.

Implementation of the Pharmacophore Filter. The pharmacophore filter has been implemented in a simple FORTRAN program. It uses a multi-structure SD file as input. The program makes use of topology recognition routines that have been developed earlier in a different context.[31] This results in very fast data processing and guarantees that the filter can be used in high throughput. About 100 000 structures can be processed in 1 min of CPU time on an SGI R10000 processor.

Neural Network Preparation. For comparison purposes we have trained a neural network to recognize drug-like and nondrug-like chemical matter following closely the work of Sadowski and Kubinyi.[17] We have constructed a feedforward neural net with 91 input neurons, five hidden neurons, and one output neuron. All layers have been connected to each other. The input neuron has been generated for each molecule by converting its structural topology into 91 Ghose and Crippen atom types found to occur at least 20 times in a training set of 10 000 molecules.[19] Note that this setup differs slightly from the 92 Ghose and Crippen atom types Sadowski and Kubinyi used in their approach. The modification is probably due to slightly different statistics of atom type occurrences in our training set compared to that used by Sadowski and Kubinyi. Also note that we introduce an additional atom type for carboxylic acid oxygen atoms. One should keep in mind that the Ghose and Crippen atom types were created and optimized as an atom-based parametrization for logP calculations. Used as topological descriptors here without any specific parametrization in mind, there is no particular reason to keep them unchanged. Moreover, we feel that the carboxylic acid as privileged substructure to bind to a wide variety of proteins[32] deserves its own oxygen atom type in the current application. At the same time we expect this slight modification to have negligible consequences for the neural network classification. The output unit provides a score between 0.1 (nondrug-like) and 0.9 (drug-like). The neural net has been trained using the back-propagation with momentum scheme as implemented in the SNNS program.[33] The training set consists of 5000 randomly chosen ACD compounds and 5000 randomly chosen MDDR compounds, representing nondrug-like and drug-like chemical matter, respectively. Training of the neural net has been performed for 2000 cycles with a learning rate of 0.2 and a momentum term of 0.1. The training data set has been shuffled before each cycle.

Table 1. Database Preparation

	number of entries				
filter	ACD	MDDR	MDDR subset[a]	CMC	CMC subset
initial	280093	101338	1492	7183	3672
remove salt, compounds with atoms other than C, N, O, S, H, P, Si, Cl, Br, F, I, false entries, entries without structure	205049	97114	1492	6675	3672
remove reactive and unsuited compounds[b]	171192	81688	1322	5784	3672
remove self-duplicates[c]	157280	78018	1322	5693	3672
remove duplicated with MDDR	156213	N/A	N/A	4807	2617
remove duplicates with CMC	155402	N/A	N/A	N/A	N/A

[a] The MDDR and CMC subsets have been prepared according to the recommendations of Bemis and Murcko[27] and Lipinski and co-workers.[13] [b] Reactive compounds and unsuited lead structures were removed using a Daylight toolkit program and SMARTS for reactive and unsuited lead compounds provided by Hann and coworkers.[35] In addition, thioureas and thioamides are dismissed. [c] Duplicates were removed using SMILES matching.

Results and Discussion

We have applied the pharmacophore filters PF1 and PF2 to the databases prepared in Table 1. Table 2 summarizes the results. Roughly one-third of the prepared ACD compounds pass PF1. In contrast, about two-thirds of the MDDR and CMC compounds pass PF1. The drug database subsets contain fewer compounds that fail the filter indicating that they contain fewer underfunctionalized compounds. This is a statistically significant discrimination between drug-like and nondrug-like compounds based on simple pharmacophore rules. The MDDR and CMC subsets that contain supposedly even more drug-like compounds than the entire MDDR and CMC, respectively, show higher rates of compounds

Table 2. Compounds Surviving Drug-likeness Filters

		compounds surviving filter [%]				
no.	filter	ACD 155408	MDDR 78028	MDDR subset 1322	CMC 4708	CMC subset 2627[a]
1	pharmacophore point filter 1[b]	36.5	65.9	68.6	60.6	67.7
1a	pharmacophore point filter 2[c]	48.3	78.4	83.5	76.4	85.3
2	neural net filter	25.3	83.7	78.4	66.2	67.9
3	1 and 2	14.9	42.0	58.9	46.4	51.9
	(1 and 2 overlap [%])	(59.0)	(74.8)	(85.9)	(76.7)	(76.7)
	{random overlap [%]}	{9.2}	{55.2}	{53.8}	{40.1}	{46.0}

[a] Number of compounds in each database (see Table 1). [b] Compounds with less than two pharmacophore points are dismissed. [c] For compounds with a carboxylic acid, amine, amidine, or guanidine pharmacophore point, the survival threshold is lowered to one.

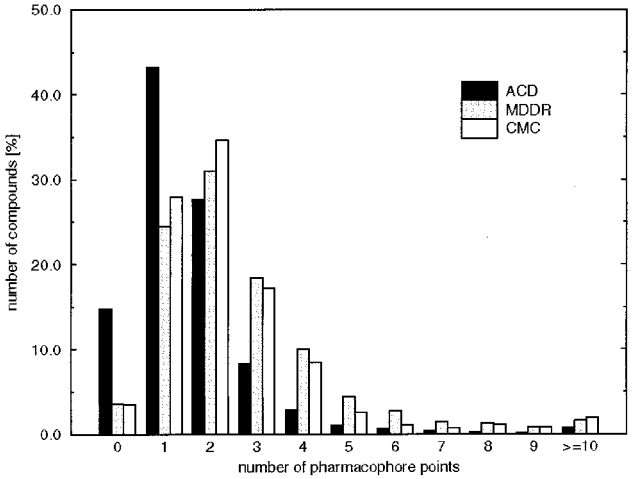

Figure 2. Distribution of pharmacophore points for drug and nondrug databases.

that survive PF1. Most notably, the number of PF1 survivors increases for the CMC subset over the entire CMC by 10%. Figure 2 shows the distribution of compounds according to their pharmacophore count for drug and nondrug databases. Although very different in size and without overlap of compounds, the drug databases have very similar pharmacophore count profiles. For both databases (MDDR and CMC), most compounds (more than 30%) have a pharmacophore count of two. In contrast, the nondrug database (ACD) has more than 40% of its compounds with only one pharmacophore point count in addition to 15% without any. These results support the finding that the pharmacophore point count is a suitable descriptor that is able to classify between drugs and nondrugs. One obvious shortcoming of PF1 is that many known drugs that are active in the central nervous system have only one pharmacophore. Therefore, a second pharmacophore filter (PF2) has been used that allows for certain compounds with only one pharmacophore point to pass the filter. Table 2 shows that the rate of compounds from drug databases that pass PF2 increases to 76−85%. However, the number of compounds from the nondrug database that pass the filter also increases. PF2 does not result in an increased discrimination between drugs and nondrugs. It merely illustrates that filter rules are adjustable to specific needs.

In comparison to the pharmacophore filter, the neural network is able to reach a better classification result between drug-like and nondrug-like compounds similar in performance to neural nets reported before.[16,17] While the neural net in most cases performs better than the pharmacophore filter, there are problems attached to the neural net approach. Although tested on large sets of the databases that were not used for training, the neural net is trained to distinguish between "drug-database-like" (e.g., CMC-like) and "nondrug-database-like" (e.g., ACD-like) compounds rather than directly addressing properties of drugs. The approach is only conceptually "true" if the drug-like (nondrug-like) database contains all the drug-like (nondrug-like) chemical matter and no nondrug-like (drug-like) chemical matter. One strength of the neural net is its ability to cope with contaminated data (a small percentage of drug-like compounds present in nondrug databases and vice versa). However, it cannot compensate very well for the absence of chemical classes in databases used for the training of the neural net. To illustrate this point, we have trained our neural net on MDDR and ACD data only, using the CMC database that has no structural overlap with the MDDR as test set. While 83% of the MDDR database compounds are classified as being drug-like by the neural net, only 66% of the CMC compounds are identified as drug-like (Table 2). This is a significant shortfall in performance of our neural net. As a matter of fact, the performance of this CMC-database-unbiased neural net is now similar to that of the pharmacophore filter. This experiment illustrates the database bias of the neural net approach. Of course, in this particular case the performance of our net could have been improved by using representatives of the CMC in the training set. (A neural net trained on ACD and both MDDR and CMC, but still not fully optimized for recognizing CMC compounds as drug-like, is able to identify more than 70% of the CMC as drug-like.) However, our goal here is not to optimize the neural net but rather to point out its possible pitfalls.

It may be interesting to ask how far the differences in distributions of simple physicochemical parameters of compounds in databases attributed with drug-likeness or nondrug-likeness may bias the interpretation of the neural net or pharmacophore filter classification results. It may be argued, for example, that since the ACD (155 402 compounds, Table 1) has an average molecular weight of only 312 Da compared to the MDDR (78,018 compounds, Table 1) with 428 Da, molecular weight as a simple physiochemical descriptor may dominate the classification decisions. To eliminate molecular weight as a factor in the classification of compounds, we choose a subset of the ACD that shows the same molecular weight distribution as the MDDR. About 22% of compounds in the MDDR have a molecular weight above 500 Da; only 4.5% of the compounds in the ACD have a molecular weight above 500 Da. These percentages determine the largest subset of the ACD that can possibly exhibit the same distribution in molecular weight as the MDDR does. This ACD subset contains 37 029 compounds. It holds all ACD compounds with molecular weight above 500 (corresponding to 22% of the subset) as well as random selections of ACD compounds in other molecular weight brackets. Applying PF1 to this ACD subset yields an increase in drug-like compounds to 45% compared to 36.5% for the entire

ACD. The neural net applied to the ACD subset yields 36% drug-like compounds for the ACD subset compared to 25% for the entire ACD. The results suggest that both the neural net as well as the pharmacophore filter depend on the molecular weight of the compounds. For the pharmacophore filter, this observation may not be surprising since larger compounds will on average bear more functionality. Discrimination against high molecular weight compounds in the ACD subset lowers the percentage of drug-like compounds to 42%. Introducing a molecular weight cutoff at 500 Da for drug-like compounds ('rule of 5') somewhat remedies the obvious molecular weight dependence of PF1 (60.5% of compounds with molecular weight above 500 Da are found to be drug-like by PF1). The neural net performs similarly. Applying the same molecular weight cutoff of 500 Da, 58.5% of the compounds with high molecular weight in the ACD are recognized as drug-like. The increase in drug-likeness for the neural net also underlines a point made earlier that it is the database-likeness that is evaluated by the neural net rather than the drug-likeness. Obviously, the molecular weight-adjusted ACD subset becomes more drug-like than the entire ACD. There is also no doubt that the ACD contains drug-like compounds. So one could also argue that the increased drug-likeness in the ACD subset as seen by both the pharmacophore filter and the neural net are due to the enrichment of drug-like compounds in the ACD subset. At the same time this finding raises the question, Is the ACD a good representative of nondrugs in the first place? An anonymous referee pointed out that a more desirable data set may consist of active compounds labeled drug-like and a set of similar but inactive compounds labeled nondrug-like. While this seems to be a good idea, there is no rich selection of data available to assemble such a data set of tens of thousands of compounds necessary to train a neural net. So it seems, at least for now, that further investigations of drug-likeness have to continue using the somewhat imperfect databases available to us today.

Another shortcoming of the neural net approach is its "black box" character. One may trace the statistics of the descriptors used in the input neurons and get some rough idea about whether a certain compound has a chance to be classified as drug or nondrug. However, there is no rigorous way to derive rules due to the nonlinear character of the network architecture.[16,34] This may limit the design power of the neural net approach, for instance, for combinatorial libraries. Since the neural net cannot be applied with confidence to the building blocks alone, there are no direct rules to guide the design of combinatorial libraries. Virtual libraries can only be filtered after enumeration; this greatly complicates the optimization process. This example illustrates why it is highly desirable to have a database-independent approach to classify drug-likeness that also provides a detailed understanding of why a molecule is classified as drug or nondrug. In light of the above discussion, we therefore consider the pharmacophore point filter as a very useful tool even though its performance may not be as good as that of the neural net.

It is instructive to further compare the pharmacophore filter and neural net results by analyzing their overlap. Table 2 shows that 59% of the ACD compounds identified by the neural net are also recognized by the pharmacophore filter as drug-like. This is a significant enrichment over a random overlap of 9%. The overlap of drug-like chemical matter identified by the neural net and the pharmacophore filter are even higher in the case of the drug databases (75−86%). However, these results are less impressive because the random overlays are also significantly higher (40−54%). Nevertheless, the significant overlap between both approaches shows that similar characteristics of molecular topology are recognized here by different means, thereby cross-confirming the validity of both approaches.

The characterization of drugs by structural motifs is not new. Bemis and Murcko[27,28] as well as Ghose and co-workers[29] have systematically studied the topology of molecules in drug databases. While their studies clearly help to understand the topology of drug-like molecules, it is hard to use their data directly for drug/nondrug discrimination purposes. This point is illustrated best by looking at benzene − the most frequently found structural motif in drugs. While this, as much as other hydrophobic ring structures, is a very important feature of many drugs, benzene also occurs in more than 60% of compounds in the ACD. Therefore, benzene as structural motif cannot help to discriminate between drugs and nondrugs. Attempts have been made by us to introduce criteria to the pharmacophore filter based on ring counts in a molecule. Although Oprea showed recently that the distribution of the number of ring counts significantly differs between ACD and MDDR,[15] we could not improve the pharmacophore filter discrimination on those databases by defining additional ring count rules.

Conclusion

A pharmacophore point filter has been developed that discriminates significantly between drug-like and nondrug-like chemical matter based on simple structural rules. While its performance is weaker than that of a neural network approach, it is free of the main drawbacks of the neural net: its "black box" character and its database bias. Strong overlaps between survivors of the pharmacophore point filter and survivors of the neural network filter validate the drug-likeness criteria of both approaches. In addition to property filters, the complimentary use of both approaches can greatly enhance our ability to characterize compounds in vendor, combinatorial, and/or virtual databases.

References

(1) Lutz, M. W.; Menius, J. A.; Choi, T. D.; Laskody, R. G.; Domanico, P. L.; Goetz, A. S.; Saussy, D. L. Experimental-design for High Throughput Screening. *Drug Discovery Today* **1996**, *1*, 277−286.

(2) Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, A. M. Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. *J. Med. Chem.* **1994**, *37*, 1233−1251.

(3) Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. Applications of combinatorial technologies to drug discovery 2. Combinatorial organic synthesis, library screening strategies, and future directions. *J. Med. Chem.* **1994**, *37*, 1385−1401.

(4) Martin, E. J.; Blaney, J. M.; Siani, M. A.; Spellmeyer, D. C.; Wong, A. K.; Moos, W. H. Measuring Diversity; Experimental Design of Combinatorial Libraries for Drug Discovery. *J. Med. Chem.* **1995**, *38*, 1431−1436.

(5) Warr, W. A. Combinatorial Chemistry and Molecular Diversity. An Overview. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 134−140.

(6) Brown, R. D.; Martin, Y. C. Designing Combinatorial Library Mixtures Using a Genetic Algorithm. *J. Med. Chem.* **1997**, *40*, 2304−2313.
(7) Potter, T.; Matter, H. Random or Rational Design? Evaluation of Diverse Compound Subsets from Chemical Structure Databases. *J. Med. Chem.* **1998**, *41*, 478−488.
(8) Martin, Y. C. Does Virtual Pre-Screening Selection Increase the Observed Quality and Quantity of Hits from Vendor and Combinatorial Libraries? Lecture given at the International Workshop Virtual Screening, March 15−18, 1999, Schloss Rauischholzhausen, Germany.
(9) Clark, D. E.; Pickett, S. D. Computational methods for the prediction of 'drug-likeness'. *Drug Discovery Today* **2000**, *5*, 49−58.
(10) Podlogar, B. L.; Muegge, I.; Brice, L. J. Computational methods to estimate drug development parameters. *Curr. Opin. Drug Discovery Dev.* **2001**, *4*, 102−109.
(11) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening − an overview. *Drug Discovery Today* **1998**, *3*, 160−178.
(12) Walters, W. P.; Ajay; Murcko, M. A. Recognizing molecules with drug-like properties. *Curr. Opin. Chem. Biol.* **1999**, *3*, 384−387.
(13) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.
(14) Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T. The Design of Leadlike Combinatorial Libraries. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 3743−3748.
(15) Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 251−264.
(16) 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.
(17) Sadowski, J.; Kubinyi, H. A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* **1998**, *41*, 3325−3329.
(18) SSKEYS, MDL Information Systems Inc., San Leandro, CA.
(19) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, 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.
(20) Comprehensive Medicinal Chemistry is available from MDL Information Systems Inc., San Leandro, CA, 94577, and contains drugs already on the market.
(21) MACCS-II Drug Data Report is available from MDL Information Systems Inc., San Leandro, CA, 94577, and contains biologically active compounds in the early stages of drug development.
(22) World Drug Index is available from Derwent Information, London, U.K. Website: www.derwent.com.
(23) Available Chemicals Directory is available from MDL Information Systems Inc., San Leandro, CA, 94577, and contains specialty bulk chemicals from commercial sources. Website: www.mdli.com.
(24) Gillet, V. J.; Willett, P.; Bradshaw, J. Identification of Biological Activity Profiles Using Substructural Analysis and Genetic algorithms. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 165−179.
(25) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. Quantitative Structure and Physicochemical Property Based Scoring Scheme to Evaluate Druglikeness of Small Organic Compounds. Lecture given at the 219th National Meeting of the American Chemical Society, San Francisco, CA, March 26−30, 2000.
(26) Wagener, M.; vanGeerestein, V. J. Potential Drugs and Nondrugs: Prediction and Identification of Important Structural Features. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 280−292.
(27) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.
(28) Bemis, G. W.; Murcko, M. A. Properties of Known Drugs. 2. Side Chains. *J. Med. Chem.* **1999**, *42*, 5095−5099.
(29) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A Kowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem.* **1999**, *1*, 55−68.
(30) Xu, J.; Stevenson, J. Drug-like Index: A new approach to measure drug-like compounds and their diversity. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1177−1187.
(31) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein−ligand interactions: A simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791−804.
(32) Hajduk, P. J.; Bures, M.; Praestgaard, J.; Fesik, S. W. Privileged Molecules for Protein Binding Identified from NMR−Based Screening. *J. Med. Chem.* **2000**, *43*, 3443−3447.
(33) SNNS: Stuttgart Neural Net Simulator, Version 4.0; University of Stuttgart, 1995.
(34) Ajay; Bemis, G. W.; Murcko, M. A. Designing Libraries with CNS Activity. *J. Med. Chem.* **1999**, *42*, 4942−4951.
(35) Hann, M.; Hudson, B.; Lewell, X.; Lifely, R.; Miller, L.; Ramsden, N. Strategic Pooling of Compounds for High-Throughput Screening. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 897−902.

JM015507E