A Discussion of Measures of Enrichment in Virtual Screening: Comparing the Information Content of Descriptors with Increasing Levels of Sophistication

Andreas Bender and Robert C. Glen*

Unilever Centre for Molecular Science Informatics, Chemistry Department, University of Cambridge, Cambridge CB2 1EW, United Kingdom

Received January 18, 2005

We have performed virtual screening using some very simple features, by employing the number of atoms per element as molecular descriptors but without regard to any structural information whatsoever. Surprisingly, these atom counts are able to outperform virtual-affinity-based fingerprints and Unity fingerprints in some activity classes. Although molecular weight and other biases were known in target-based virtual screening settings (docking), we report the effect of using very simple descriptors for ligand-based virtual screening, by using clearly defined biological targets and employing a large data set (>100 000 compounds) containing multiple (11) activity classes. Structure-unaware atom count vectors as descriptors in combination with the Euclidean distance measure are able to achieve "enrichment factors" over random selection of around 4 (depending on the particular class of active compounds), putting the enrichment factors reported for more sophisticated virtual screening methods in a different light. They are also able to retrieve active compounds with novel scaffolds instead of merely the expected structural analogues. The added value of many currently used virtual screening methods (calculated as enrichment factors) drops down to a factor of between 1 and 2, instead of often reported double-digit figures. The observed effect is much less profound for simple descriptors such as molecular weight and is only present in cases of atypical (larger) ligands. The current state of virtual screening is not as sophisticated as might be expected, which is due to descriptors still not being able to capture structural properties relevant to binding. This fact can partly be explained by highly nonlinear structure−activity relationships, which represent a severe limitation of the "similar property principle" in the context of bioactivity.

1. INTRODUCTION

Similarity searching of molecules,[1−4] used extensively for virtual screening, is an everyday task in the pharmaceutical industry for the identification of novel active compounds.[5] Although predictions made through the application of the "molecular similarity principle" still need to be validated by performing experiments, it is often possible to screen out many inactive compounds—precisely predicting the degree of activity still remains a major challenge. The prediction of the bioactivity of compounds is often more difficult than the prediction of other properties such as solubility or log P.[3,4] Even compounds that are very similar overall may show subtle geometrical or functional differences that may reduce (enhance) binding affinity and bioactivity dramatically, for example, because of steric repulsion (better complementarity to the binding site). The phenomenon that structurally similar compounds show very different bioactivity is discussed in a recent review,[6] while empirical evidence has also been reported.[7]

Conventionally, so-called "enrichment factors" are calculated in the literature to establish a baseline for assessing the quality of virtual screening methods. Enrichment factors describe the number of active compounds found by employing a certain virtual screening strategy, as opposed to the number of compounds hypothetically found if compounds were screened randomly. Enrichment factors are defined by the success of the virtual screening algorithms at ordering the library, with the most likely compounds to be active being suggested by the algorithms to be screened first. Enrichment factors can range from 1, where molecules are sorted randomly and virtually no "enrichment" is achieved by the algorithm, to >100, in which only a small percentage of the library needs to be screened to find a large number of active molecules.[8,9] Very high enrichment can only be observed for very small parts of the sorted library because of the way the performance measure is calculated (see the formula in Material and Methods section). Enrichments smaller than 1 are indicative of a preferred selection of inactive compounds, corresponding to an enrichment of active compounds at the bottom of the sorted library.

At first sight, this seems to be a suitable benchmark since it compares the "rational" drug discovery procedure to the "irrational", or random, method. As we demonstrate here, enrichment factors, which are often reported in the double-digit numbers, give a very high performance estimate of virtual screening algorithms by using a low benchmark (random hit rates) for comparison. Enrichment factors of 10 create the impression that the particular virtual screening method saves a great amount of time, expenditure, and experimental work. Although indeed, at least in a retrospective sense, enrichment factors of that size are correct in the abstract way they are calculated, they are only a realistic benchmark for added value over truly random selection. The

* Corresponding author phone: +44 (1223) 336 432; fax: +44 (1223) 763 076; e-mail: rcg28@cam.ac.uk.

10.1021/ci0500177 CCC: $30.25 © 2005 American Chemical Society
Published on Web 08/05/2005

main justification for employing this baseline probably lies in the fact that we assume that molecules in a given compound library are “random” (show every arrangement of molecular features with the same likelihood), and hence, we apply random selection methods to the library as a benchmark. However, there is no such thing as a “fully comprehensive random library” in which all possible arrangements of features are present in a uniform distribution, so that every library shows a bias with respect to one set of properties or another.

In this article, we show that enrichment factors of around 4 (averaged over several classes of active compounds and two different data sets) can be achieved using very simple, nonstructural features. Atom counts distinguished by elements are used as “dumb” descriptors, and the sum of absolute distances of the atom count “fingerprint” is used as a similarity (or, rather, distance) measure. Enrichment factors obtained via this method on a data set derived from the MDL Drug Data Report (MDDR) database[10] are higher than, for example, those achieved via a virtual affinity fingerprint-based method (DOCKSIM). This is particularly surprising in the view that docking-based methods have a large amount of information about the binding site at hand (and are computationally very demanding).

On a different data set that is also derived from the MDDR, “virtual screening” using atom count vectors outperforms Unity fingerprints on some activity classes. Although, overall, Unity fingerprints are still superior, their added value (enrichment rate with respect to these very simple features) drops down to a factor of around 2. This is particularly surprising since atom counts do not capture structural information at all, which is exactly the kind of information captured in more complex fingerprints.

On the other hand, other simple molecular features such as molecular weight are shown to give only minimal enrichment. The simple assumption that it is possible to identify close active analogues by weight-based virtual screening can, thus, be dismissed. Interestingly, as shown below, count vectors of atoms as descriptors for molecules do not only retrieve a high number of active compounds, they are also able to identify actives from different structural classes.

When the simplicity principle is followed (Occam’s razor: one should not make more assumptions than the minimum needed), it can be concluded that the performance of virtual screening methods should be seen in relation to their complexity, and more sophisticated methods do not lead to superior results in every case. While this work, on one hand, shows that the added value of virtual screening methods is not as large as might be expected from random-selection-based enrichment factors, it also shows that there are fundamental differences present between different drug activity classes that can be detected by simple atom counts. These descriptors (partially) implicitly capture overall properties such as size, lipophilicity, hydrogen bonding capabilities, and polarity, which seem to be suitable discriminants for some activity classes; in those cases, they are as discriminating as structure-based fingerprints. Also, privileged substructures which are encountered for example in certain GPCR ligand classes may be partly and implicitly encoded in atom counts.

In the recent literature, several publications appeared that are related to the work presented here. In his review on one-dimensional descriptors,[11] Livingstone discusses overall molecular parameters that are able to discriminate between compounds showing different physicochemical or biological behavior. For example, blood−brain barrier penetration is closely related to log P, and electron density on a nitrogen atom in the HOMO of a set of aniline mustards and tumor inhibition can be related in a simple linear fashion. The focus is different in the present work, where we apply simple molecular descriptors to ligand-based virtual screening, which involves multiple activity classes, to gauge how many active molecules can be found in retrospective virtual screening runs by their application. Results are compared to both random selection and structure-based descriptors, and this is performed in a quantitative fashion.

One of the simple descriptors employed by us is molecular weight, whose influence on target-based (instead of ligand-based) virtual screening has been investigated by Pan et al.[12] In this earlier work, it was found that heavier molecules are favored by docking algorithms because of the simple fact that, on average, more atom−atom interactions are present that contribute to the predicted binding energy. As a remedy, normalization of the binding energy with respect to the number of heavy atoms per molecule (or a root thereof, depending on whether drug-like or lead-like compounds were desired) was suggested.

Bioactivity profiles (BPs), which include the number of hydrogen bond donors and acceptors, molecular weight, a kappa shape index, and the number of rotatable bonds as well as the number of aromatic rings, were introduced by Gillet et al.[13] BPs found application in distinguishing molecules from the World Drug Index and those from the SPRESI database (which were assumed to be inactive). When single features such as the number of hydrogen bond donors alone were used, enrichments of up to 4.6 were found in identifying WDI molecules in a merged data set. Although the simple description of the structures resembles our approach, Gillet et al. do not employ atom count vectors, which perform favorably in the work presented here. In addition, we employ clearly defined targets instead of therapeutic classes in order to predict activity on a particular receptor or enzyme, which is the usual aim in drug discovery programs. Similar differences in scope are given to the work by Wilton et al.,[14] who employed substructural analysis and bioactivity profiles in combination with binary kernel discrimination (among other methods) for the identification of active compounds from the well-known NCI AIDS data set. The activity-determining factor of the NCI AIDS data set is growth inhibition, so neither a consistent molecular target nor multiple activity classes are employed.

Attempts to avoid “artificial enrichment”, defined as the identification of active compounds by differing simple molecular properties, were presented by Verdonk et al.[15] in the context of target-based virtual screening. Considering heavy atom counts alone on two hypothetical libraries of active compounds, which are either, on average, much heavier or much lighter than the whole library, was shown to give considerable enrichments. Several steps were taken in joining a library of true active binders and nonactive HTS compounds to give a similar distribution of features in both data sets in order to eliminate “artificial enrichment”. In our

Table 1. Activity Classes, MDDR Activity IDs, and Sizes of Active Data Sets Derived from the MDDR

activity name	MDDR activity ID	data set size
5HT3 antagonists	06233	752
5HT1A agonists	06235	827
5HT reuptake inhibitors	06245	359
D2 antagonists	07701	395
renin inhibitors	31420	1130
angiotensin II AT1 antagonists	31432	943
thrombin inhibitors	37110	803
substance P inhibitors	42731	1246
HIV protease inhibitors	71523	750
cyclooxygenase inhibitors	78331	636
protein kinase C inhibitors	78374	452

case, we expand the work of Verdonk, whose main focus was on the performance of docking-based virtual screening methods, to multiple simple molecular features as well as multiple activity classes. Interestingly, the consideration of molecular weight alone did not give major improvements in identifying active compounds over random selection in our ligand-based virtual screening setting. This may simply depend on the fact that most of the activity classes are not as different in molecular weight from the distribution of the whole library as was the case in the hypothetical example given by Verdonk.

To summarize, the novelty of the work presented here is that it employs simple features for ligand-based virtual screening on a large data set (>100 000 compounds), which, due to its size and comprehensiveness, can be seen as one of the best datasets currently present in the literature. The data set comprises multiple (11) activity classes for clearly defined molecular targets. When simple features are used on this data set, considerable enrichments can be found that are sometimes as good as those obtained using structural fingerprints. We believe that the performance of more complex descriptors employed for similarity searching has, thus, to be gauged in relation to their sophistication.

2. MATERIALS AND METHODS

Two different data sets were examined that were previously subjected to retrospective virtual screening using a variety of methods. The first data set was published by Briem and Lessel,[16] and it contains 957 ligands extracted from the MDDR database. The set contains 49 5HT3 receptor antagonists (5HT3), 40 angiotensin converting enzyme inhibitors (ACE), 111 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase inhibitors (HMG), 134 platelet activating factor antagonists (PAF), and 49 thromboxane A2 antagonists (TXA2). An additional 574 compounds were selected randomly that did not belong to any of these activity classes. The second and larger data set was presented recently by Hert et al.[17,18] Eleven sets of active structures were defined, ranging in size from 349−1236 structures. (Full details of the data set sizes are given in Table 1.) This data set spans a variety of targets as well as a very large number of compounds, which provides a useful benchmark for a similarity searching method. One has to aware of the shortcomings of all current drug database-derived data sets, which is the occurrence of close analogues, which favors 2D methods, and the fact that the MDDR does not include explicit information about the inactivity of compounds (which means that inactive compounds identified as false positives may well be active and, thus, true positives). Nonetheless, relative values on retrospectively analyzed data sets can be used to judge the relative performance of different molecular similarity searching methods.

For all compounds of both data sets, simple atom count vectors were calculated using MOE,[19] namely, the total number of atoms; the number of heavy atoms; and the numbers of boron, bromine, carbon, chlorine, fluorine, iodine, nitrogen, oxygen, phosphorus, and sulfur atoms. Thus, no structural descriptors at all were contained in this "fingerprint" representation, which, besides the compound ID, contains just 12 integer numbers describing the frequency of different elements in the molecule.

Single queries were selected from both of the data sets. On the first, smaller data set, each "active" compound was selected once and the remaining structures were sorted according to their similarity to the query. Hit rates among the 10 nearest neighbors were reported, following the earlier protocol by Briem and Lessel,[16] which compared feature trees;[20] ISIS MOLSKEYS;[21] Daylight fingerprints;[22] SYBYL hologram QSAR fingerprints;[23] and FLEXSIM-X,[16] FLEXSIM-S,[24] and DOCKSIM[25] virtual affinity fingerprints. From the second, larger data set, single queries were selected randomly 10 times and the number of active compounds in the top 5% of the sorted library was recorded, allowing comparison to Unity fingerprints[17] and circular fingerprints (MOLPRINT 2D).[26,27]

For both data sets, the fraction/number of active compounds found was compared between the screening runs using "dumb" atom count vector descriptors and (supposedly) more information-rich 2D and 3D descriptors. Enrichment factors (E_f) after x% of the focused library were calculated according to the following formula ($N_{\text{experimental}}$ = number of experimentally found active structures in the top x% of the sorted database, N_{expected} = number of expected active structures, N_{active} = total number of active structures in database).

$$E_{\text{f}} = \frac{{N_{\text{experimental}}}^{x\%}}{{N_{\text{expected}}}^{x\%}} = \frac{{N_{\text{experimental}}}^{x\%}}{N_{\text{active}} \cdot x\%}$$

3. RESULTS AND DISCUSSION

The average hit rate using "dumb" atom count descriptors, compared to a variety of 2D and 3D similarity searching methods, is shown in Figure 1 for the first, smaller data set. Atom count descriptors achieve an enrichment of about 4-fold (average hit rate of 34%), compared to a hit rate from random sampling of 7.9%. Hit rates vary through the five data sets of active compounds between 16% (thromboxane A2 antagonists) and 51% (HMG-CoA reductase inhibitors). 2D-based fingerprints (MOLPRINT 2D, feature trees, ISIS MOLSKEYS, Daylight fingerprints, SYBYL hologram QSAR fingerprints; results taken from Briem and Lessel[28] and Bender et al.[27]) are found, overall, to be superior to virtual affinity fingerprints (FLEXSIM-X, FLEXSIM-S, and DOCKSIM). Elemental atom counts achieve an average hit rate of 34% and are, thus, ranked worse than FLEXSIM-X and FLEXSIM-S but better than DOCKSIM.

On this first data set, computationally much more demanding virtual affinity fingerprints (the DOCKSIM method) are outperformed by simple atom counts. Although docking-

Table 2. Activity Class, Hit Rate among the Top 5% of the Sorted Database, and Hypothetical Enrichment for the Different Sets of Active Compounds of the Large Test Set[a]

activity class	06233	06235	06245	07701	31420	31432	37110	42731	71523	78331	78374	average
hit rate atom counts	23.78	14.59	11.59	18.58	66.53	21.89	23.42	15.69	27.89	8.69	12.48	22.28
enrichment	4.76	2.92	2.32	3.72	13.31	4.38	4.68	3.14	5.58	1.74	2.50	4.46
hit rate Unity	21.15	18.43	24.02	17.53	80.54	48.04	33.51	26.87	37.60	9.39	19.42	30.59
Unity/atom counts	0.89	1.26	2.07	0.94	1.21	2.19	1.43	1.71	1.35	1.08	1.56	1.43
hit rate MOLPRINT 2D	25.40	27.73	22.75	23.24	95.04	68.01	34.79	31.03	49.56	13.16	21.13	37.44
MOLPRINT 2D/atom counts	1.07	1.90	1.96	1.25	1.43	3.11	1.49	1.98	1.78	1.51	1.69	1.68

[a] Using simple atom count descriptors, up to more than 10-fold enrichment can be observed, which is close to the results achieved using Unity fingerprints on the same data set.

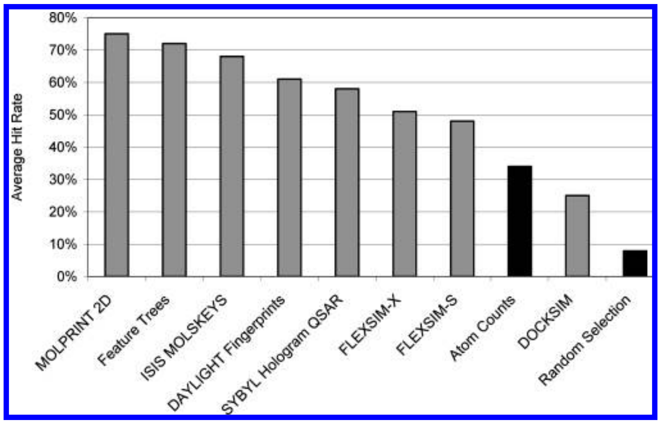

Figure 1. Average hit rate using “dumb” atom count descriptors, compared to a variety of 2D and 3D similarity searching methods. Even atom count descriptors achieve an enrichment of about 4-fold, which is already superior to one of the virtual affinity fingerprint methods, DOCKSIM, and around half the enrichment achieved by other methods employed.

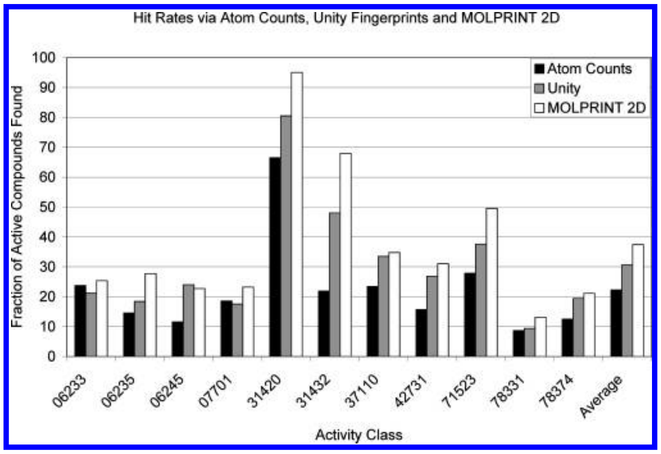

Figure 2. Fraction of active compounds found using simple atom counts, in comparison to Unity fingerprints and the MOLPRINT 2D method. Although Unity fingerprints outperform atom counts, overall, this margin is smaller than one might expect, given the fact that atom counts do not contain any structural information whatsoever, whereas Unity fingerprints have that information available.

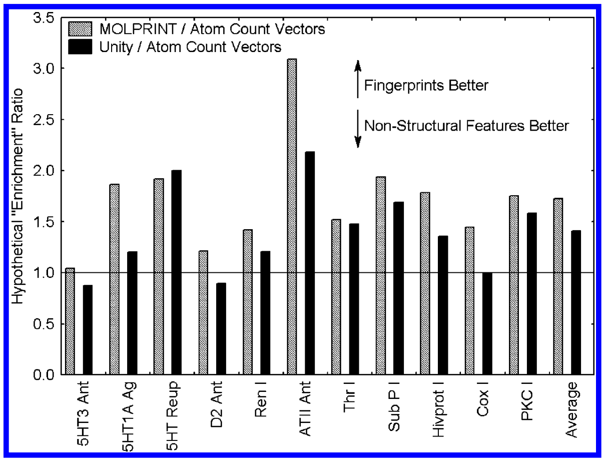

Figure 3. Enrichment factors of circular fingerprints (MOLPRINT 2D) and Unity fingerprints, put in relation to enrichment factors obtained from nonstructural fingerprints (atom count vectors). Overall, MOLPRINT 2D descriptors achieve hit rates that are, on average, about 70% larger than those achieved by atom count vectors, whereas enrichments from Unity fingerprints are, on average, 40% larger than atom count vector enrichments. On some classes (5HT3 antagonists and D2 antagonists), both descriptor types perform comparably.

based methods, in principle, are able to exploit a wealth of information of the binding site, this confirms—in agreement with other recent research[29]—that current scoring functions are not able to predict the binding affinity of a ligand to a receptor reliably. 2D descriptors are generally able to add value to similarity searching protocols, but not as much as a random screening baseline suggests.

The average fraction of active compounds retrieved on the second, large data set is given in Table 2 and Figure 2. Between 1.7-fold and 13-fold enrichment can be observed for simple atom counts within the top 5% of the sorted database. The lowest enrichment was observed for cyclooxygenase inhibitors (enrichment of 1.7) and 5HT reuptake inhibitors (enrichment of 2.3), the highest enrichment for renin inhibitors (13.3-fold enrichment) and HIV protease inhibitors (enrichment of 5.6). Compared to virtual screening results using Unity fingerprints, simple atom counts are able to outperform Unity fingerprints in two instances, namely, on the 5HT3 and dopamine D2 antagonist data sets, while in other cases, Unity fingerprints are superior to simple atom counts by a factor of up to 2. Although circular fingerprints (atom-centered fingerprints) retrieve, overall, more active compounds, in cases such as the 5HT3 antagonist data set, simple atom counts are not outperformed by a large margin. All three methods retrieve a remarkably similar number of compounds.

To gauge the relative performance of structural and nonstructural descriptors for retrieving active compounds of the different activity classes, enrichments factors of structural descriptors (Unity fingerprints, MOLPRINT 2D fingerprints) were divided by enrichment factors obtained by employing purely nonstructural descriptors (atom count vectors). Results are shown in Figure 3. An enrichment ratio of both methods of 1 corresponds to equal performance of both methods, with respect to the number of active compounds retrieved from this particular activity class. Enrichment ratios larger than 1 indicate superior performance of structural descriptors, while ratios smaller than 1 show superior performance of atom count vectors. Overall, MOLPRINT 2D descriptors achieve hit rates that are, on average, about 70% larger than those achieved by atom count vectors, while enrichments from Unity fingerprints are, on average, 40% larger than atom count vector enrichments. On some classes (5HT3 antagonists and D2 antagonists), both descriptor types perform comparably. Enrichments should, thus, be seen in relation to the sophistication and computational expense of the descriptor.

In this second data set, Unity fingerprints are superior to simple atom counts in the majority of the activity classes. Still, put another way, Unity fingerprints do not capture more information relevant to activity in some data sets than simple atom counts do. Indeed, they perform marginally worse than atom counts in those cases. In cases where the performance of Unity fingerprints is superior, their added value amounts to a factor between 1 and 2, with circular fingerprints performing only slightly better. This result is particularly remarkable since nonstructural information is commonly not expected to give nearly as good results for screening as when utilizing structural information. Although Unity fingerprints were chosen in this work as a reference method, it should be noted that they were employed simply because of their availability and ubiquity, not because of their particularly good or bad performance. In practice, Unity fingerprints, circular fingerprints, or other fingerprint definitions will probably still be preferred, but it should be kept in mind that their information content is, at least on the data set employed here and with respect to virtual screening, not as large as one might naively expect.

We also employed other simple molecular descriptors on this data set to investigate whether the good result of atom count vectors is simply due to differences in size, which is already known to have a profound impact on target-based (instead of ligand-based) virtual screening (docking). It is well-known[12] that docking favors larger molecules simply because of the larger number of interactions present between the ligand and target. In Figure 4, enrichment factors for the first 5% of the sorted library are given for MOLPRINT 2D structural fingerprints, Unity fingerprints, atom count vectors, the total number of atoms, and molecular weights for the 11 classes of active compounds. Enrichment factors using molecular weight and the number of atoms alone show similar performance in that they only give meaningful enrichment in the cases of two data sets, renin inhibitors and HIV protease inhibitors, which are, on average, much heavier than the rest of the database. For other compound classes, molecular weight and the total number of atoms are not meaningful discriminants between activity classes. Thus, the number of active compounds found by atom count vectors cannot be reduced to their ability to capture differences in size and molecular weight.

The overlap of active compounds retrieved using different descriptors is shown in Table 3, together with the highest theoretically achievable overlap, given in parentheses (which is smaller than 100% because of the different hit list sizes). Overall, an overlap between 21% and 54% between the hit lists can be observed. Interestingly, hit lists from circular fingerprints and atom count vectors overlap by less than 50% of the theoretically possible value, indicating partly orthogonal behavior of those descriptors.

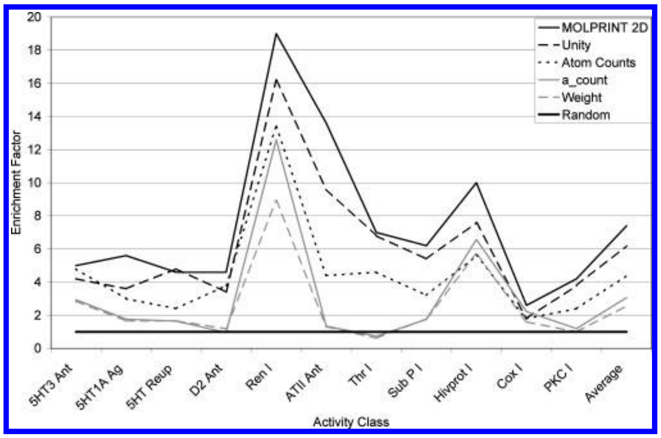

Figure 4. Enrichment factors obtained within the top 5% of the sorted library, depending on the activity class, for the descriptors MOLPRINT 2D, Unity fingerprints, atom count vectors, total number of atoms, molecular weight and for random selection. Atom count vectors perform surprisingly well with average enrichments of >4, and they give, in two cases, results comparable to those of Unity fingerprints. Molecular weight and the total number of atoms perform, on average, worse with enrichments of slightly more than 2, but they are still able to achieve considerable enrichment in the case of renin inhibitors and HIV protease inhibitors, which are much larger than the average library compound.

Table 3. Overlap of the Active Compound Set identified by Employing Structural Features (Circular Fingerprints, MOLPRINT 2D), Atom Count Vectors, the Total Number of Atoms, and the Molecular Weight[a]

	MOLPRINT 2D	count vector	# atoms	mol. weight
MOLPRINT 2D	100%			
count vector	35% (81%)	100%		
# atoms	25% (40%)	54% (57%)	100%	
mol. weight	21% (35%)	35% (55%)	38% (88%)	100%

[a] The highest possible overlap if the smaller set is completely contained in the larger set is in parentheses. An overall medium overlap can be observed, with the two best performing methods (circular fingerprints and count vectors) showing less than a 50% overlap of the retrieved active compounds.

To investigate this route further, retrieved active compounds were inspected manually, an example of which is given in Figure 5. Shown is the 5HT3 antagonist used to perform virtual screening on the database, at the top of the figure, as well as two hit lists of active compounds, depicted below. On the left-hand side, active compounds are shown that are retrieved using circular fingerprints and the Tanimoto coefficient in the first 20 positions of the sorted list. On the right-hand side, active compounds are given that are retrieved using atom count vectors within the first 50 positions of the sorted list. While the absolute number of active compounds is larger in the case of structural fingerprints, it is interesting to see that atom count vectors retrieve not only close analogue compounds. Instead, a variety of scaffolds are found using this approach. This contradicts the simple assumption that atom count vectors are only able to identify close analogues, which, given their similar overall structure, might also be assumed to show similar atom count vectors.

4. CONCLUSIONS

"Dumb" molecular features such as atom count vectors, which do not contain any structural information, are able to

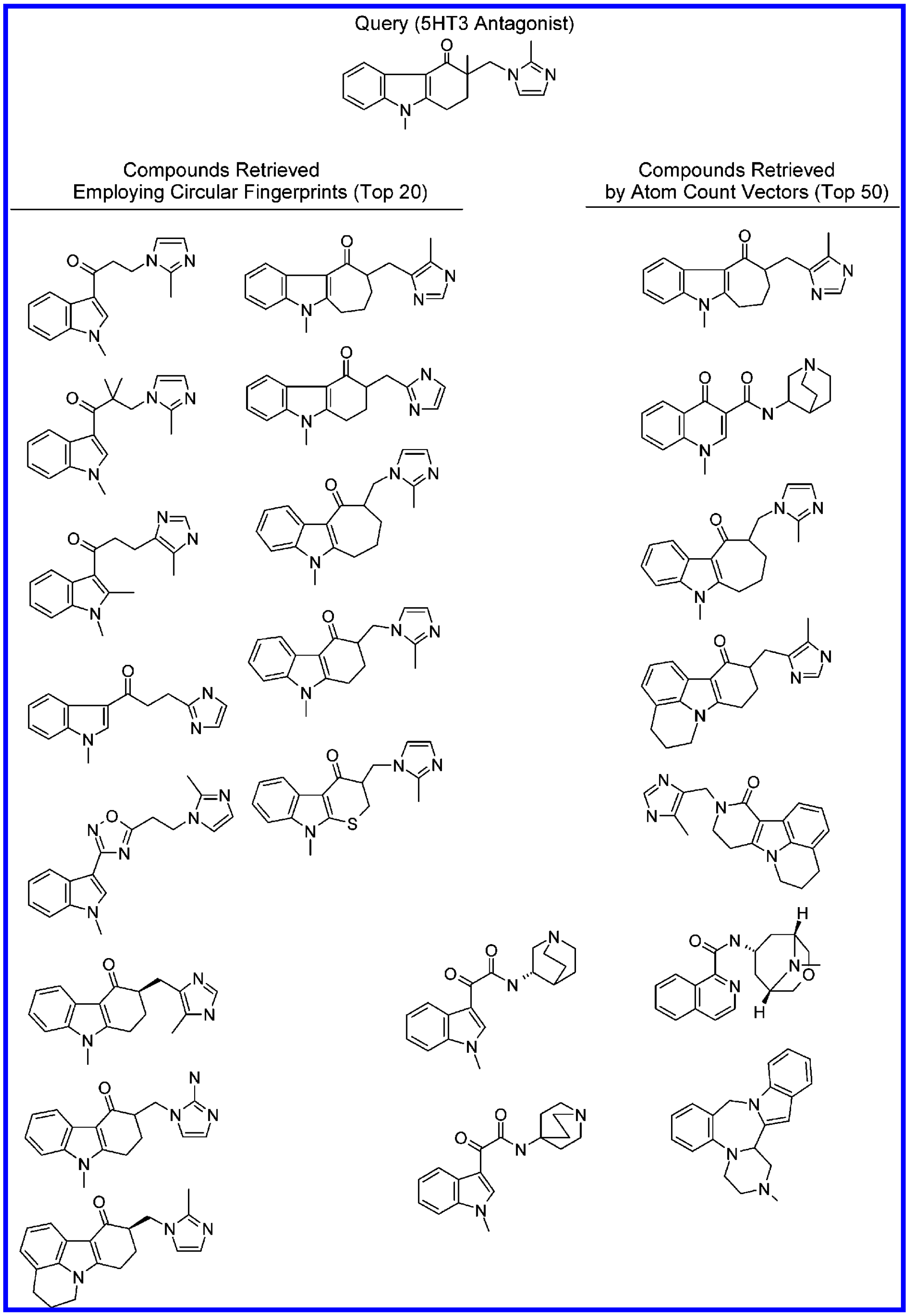
Query (5HT3 Antagonist)
Compounds Retrieved Employing Circular Fingerprints (Top 20)
Compounds Retrieved by Atom Count Vectors (Top 50)

Figure 5. Compounds retrieved using the query (5HT3 antagonist) at the top of the page and employing structural information (circular fingerprints; hits on the left) and simple atom counts (structures on the right). Although the number of active compounds found is larger in the case when structural fingerprints are employed, enrichment is considerable by using atom count vectors alone. Interestingly, not only close analogues of the query compounds are identified by the atom count “descriptor”.

achieve “enrichment factors” of around 4 in ligand-based virtual screenings, and in some cases, they even outperform Unity fingerprints. This cuts down the added value of virtual screening methods to a factor of between 1 and 2, putting previously reported double-digit figures for “enrichments” into perspective. It follows that performance measures reported for more sophisticated virtual screening methods should be seen in relation to their complexity. Performance, overall, increases with the complexity of the descriptor employed (molecular weight < atom count vectors < structural fingerprints), but this is not true in every case. In practice, Unity fingerprints, circular fingerprints, or other fingerprint definitions will probably still be preferred, but it should be kept in mind that their information content is, at least on the data set employed here and with respect to virtual screening, not as large as one might naively expect.

On the basis of these results, it would seem that virtual screening methods do not add as much value, compared to nonstructural features, as might be inferred from comparisons to hit rates achieved from random screening. On the other hand, simple atom counts seem to capture a lot of information about the difference between structures from different activity classes, implicitly encoding some of the global molecular parameters. This is seen only in two particular cases that we examined and is to a much lesser extent true for molecular weight and the total number of atoms, which are known to be important determinants of predicted activity in target-based virtual screenings. It follows that there may be physicochemical properties important for binding that are (partly) implicitly captured by atom count vectors, such as size, hydrogen bond capabilities, and polarity of the molecule, that are not contained in the total number of atoms or the molecular weight.

Put another way, the information content of two common structure-based descriptors for virtual screening purposes is, in some cases, not higher than the nonstructural information about the number of atoms per element in the structure. The extent of this finding depends on the class of active structures. At the same time, overlap between compounds retrieved on the basis of structural features (circular fingerprints) and atom count vectors is rather low, suggesting some orthogonality in the features they describe. Although “enrichment factors” obtained by atom count vectors are usually lower than those obtained by using structural descriptors, retrieved active compounds show a surprising variety of scaffolds as exemplified by a 5HT3 antagonist virtual screening run, contradicting the assumption that close analogues are mainly identified by atom count vectors.

As a bottom line, the limitations of the “molecular similarity principle” in the context of virtual screening should never be forgotten: small structural changes may, in the arena of bioactivity, give rise to large changes in activity space. This may partly explain the problem in finding suitable molecular descriptors for this task and, thus, also the relatively good performance of very simple approaches to the problem.

ACKNOWLEDGMENT

We thank Unilever, The Gates Cambridge Trust, and Tripos Inc. for support. Uta Lessel and Jérome Hêrt are thanked for providing us with their MDDR data set IDs and The Computational Chemistry Group for providing MOE.

REFERENCES AND NOTES

(1) Johnson, M. A.; Maggiora, G. M. *Concepts and Applications of Molecular Similarity*; John Wiley & Sons: New York, 1990.
(2) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.
(3) Bender, A.; Glen, R. C. Molecular similarity: a key technique in molecular informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204−3218.
(4) Nikolova, N.; Jaworska, J. Approaches to measure chemical similarity− A review. *QSAR Comb. Sci.* **2004**, *22*, 1006−1026.
(5) Schneider, G.; Bohm, H. J. Virtual screening and fast automated docking methods. *Drug Discov. Today* **2002**, *7*, 64−70.
(6) Kubinyi, H. Similarity and dissimilarity: A medicinal chemist’s view. *Perspect. Drug Discov. Design* **1998**, *9*−*11*, 225−252.
(7) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J. Med. Chem.* **2002**, *45*, 4350−4358.
(8) Feher, M.; Deretey, E.; Roy, S. BHB: a simple knowledge-based scoring function to improve the efficiency of database screening. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1316−1327.
(9) Jain, A. N. Ligand-based structural hypotheses for virtual screening. *J. Med. Chem.* **2004**, *47*, 947−961.
(10) MDL Drug Data Report; MDL ISIS/HOST software, MDL Information Systems, Inc.
(11) Livingstone, D. J. The characterization of chemical structures using molecular properties. A survey. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 195−209.
(12) Pan, Y. P.; Huang, N.; Cho, S.; MacKerell, A. D. Consideration of molecular weight during compound selection in virtual target-based database screening. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 267−272.
(13) 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.
(14) Wilton, D.; Willett, P.; Lawson, K.; Mullier, G. Comparison of ranking methods for virtual screening in lead-discovery programs. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 469−474.
(15) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; et al. Virtual screening using protein−ligand docking: Avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793−806.
(16) Lessel, U. F.; Briem, H. Flexsim-X: a method for the detection of molecules with similar biological activity. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 246−253.
(17) Hert, J.; Willett, P.; Wilton, D. J. Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1177−1185.
(18) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; et al. Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org. Biomol. Chem.* **2004**, *2*, 3256−3266.
(19) *MOE* (Molecular Operating Environment); Chemical Computing Group Inc.: Montreal, Quebec, Canada.
(20) Rarey, M.; Dixon, J. S. Feature trees: a new molecular similarity measure based on tree matching. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 471−490.
(21) *ISIS*, Version 2.1.4; Molecular Design Ltd.: San Leandro, U.S.A.
(22) *DAYLIGHT*, Version 4.62; DAYLIGHT Inc.: Mission Viejo, California, U.S.A.
(23) *SYBYL*, Version 6.5.3; HQSAR Module, Tripos Inc.: St. Louis, Minnesota, U.S.A.
(24) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: a method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, *41*, 4502−4520.
(25) Briem, H.; Kuntz, I. D. Molecular similarity based on DOCK-generated fingerprints. *J. Med. Chem.* **1996**, *39*, 3401−3408.
(26) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Molecular similarity searching using atom environments, information-based feature selection, and a naive bayesian classifier. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 170−178.
(27) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity searching of chemical databases using atom environment descriptors: evaluation of performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1708−1718.
(28) Briem, H.; Lessel, U. In vitro and in silico affinity fingerprints: Finding similarities beyond structural classes. *Perspectives in Drug Discovery and Design* **2000**, *20*, 231−244.
(29) Marsden, P. M.; Puvanendrampillai, D.; Mitchell, J. B. O. Predicting protein−ligand binding affinities: a low scoring game? *Org. Biomol. Chem.* **2004**, *2*, 3267−3273.

CI0500177