# Scaffold Hopping through Virtual Screening Using 2D and 3D Similarity Descriptors: Ranking, Voting, and Consensus Scoring

Qiang Zhang and Ingo Muegge*

*Boehringer Ingelheim Pharmaceuticals Inc., 900 Ridgebury Road, P.O. Box 368, Ridgefield, Connecticut 06877-0368*

*Received May 18, 2005*

The ability to find novel bioactive scaffolds in compound similarity-based virtual screening experiments has been studied comparing Tanimoto-based, ranking-based, voting, and consensus scoring protocols. Ligand sets for seven well-known drug targets (CDK2, COX2, estrogen receptor, neuraminidase, HIV-1 protease, p38 MAP kinase, thrombin) have been assembled such that each ligand represents its own unique chemotype, thus ensuring that each similarity recognition event between ligands constitutes a scaffold hopping event. In a series of virtual screening studies involving 9969 MDDR compounds as negative controls it has been found that atom pair descriptors and 3D pharmacophore fingerprints combined with ranking, voting, and consensus scoring strategies perform well in finding novel bioactive scaffolds. In addition, often superior performance has been observed for similarity-based virtual screening compared to structure-based methods. This finding suggests that information about a target obtained from known bioactive ligands is as valuable as knowledge of the target structures for identifying novel bioactive scaffolds through virtual screening.

## Introduction

Finding bioactive compounds of novel chemotype using the information of already known bioactive small molecules and scaffolds is often referred to as scaffold hopping.[1] The ability of various descriptors and similarity measures to facilitate clustering of bioactive compounds and ultimately enable scaffold hopping has been studied for some time.[2−10] The seminal work of Brown and Martin suggests that 3D methods offer no advantages over topological searches when bioactive compounds are identified by similarity.[2,3] Recent work by Hert et al. involving 15 2D descriptors as well as 11 well-established drug targets suggests that 2D fingerprints are indeed powerful tools for similarity-based virtual screening.[11,12] Similar results have been obtained by others.[5,13] In contrast, recent work suggests also that 3D descriptors, if applied appropriately with respect to conformation generation and alignment, can provide similarity searching results that can compete with 2D descriptors.[14−16] For instance, Jenkins at al. has presented work showing that 3D pharmacophore feature type descriptors called FEPOPS can outperform standard 2D descriptors for five standard drug targets in a scaffold hopping-oriented virtual screening protocol.[15] Similarly, Good et al. have demonstrated for four drug targets applying a variety of 2D and 3D descriptors that 3D pharmacophore fingerprints are able to assist better than 2D descriptors in scaffold hopping, resulting in varying enrichments of novel chemotypes identified in virtual screens.[17] While 2D descriptors, especially topological descriptors, are expected to work better if the actives are topologically more related to each other than to the database of negative controls, an equally clear relationship cannot be established for 3D pharmacophore type descriptors. Therefore, this study is particularly concerned with the ability of 3D pharmacophore descriptors to facilitate scaffold hopping in virtual screening for two different cases: (i) The distribution of topological similarities among actives is comparable to that between actives and negative controls. (ii) The topological similarity between actives is significantly higher than the topological similarity between actives and negative controls.

* To whom correspondence should be addressed. Phone (203) 798-4334. FAX (203) 791−6072. imugge@rdg.boehringer-ingelheim.com.

3D descriptors used in the context of similarity searching have mostly been based on the idea of capturing pharmacophore features such as hydrophobic or aromatic moieties, hydrogen bond (HB) acceptors or donors, and negative and positive ionizable groups. Fingerprints have been introduced that store the information of pairs, triplets, or quartets of pharmacophore features for multiple conformations in the form of binned distance ranges.[18−25] 3D pharmacophore fingerprints are now being used more regularly as descriptors for QSAR as well as for the design of compound libraries.[24] Here we investigate the ability of 3D pharmacophore fingerprints to identify novel bioactive chemical classes through ligand-based virtual screening in comparison to atom pair descriptors as well as Daylight fingerprints as representative topological fingerprint[26] using the Tanimoto coefficient as similarity measure.[27] While recent work has mainly focused on exploring a broad range of 2D and 3D descriptors, we focus here on evaluating different ranking strategies including voting and consensus scoring to improve virtual screening performance for scaffold hopping. In addition, the studies are complemented by structure-based virtual screening experiments using Glide2.5 that allow for comparing ligand-based and structure-based virtual screening strategies on equal footing.

## Methods

**Descriptor Calculation.** Daylight fingerprints (DF) have been calculated using a Daylight toolkit program. 3D 3-point pharmacophore fingerprints (PFP) have been calculated using an in-house implementation of the pharmacophore fingerprint descriptor as outlined in detail by McGregor and Muskal[23] using six distance ranges and seven pharmacophore features, resulting in 10 549 bins. Multiple compound conformations have been generated using OMEGA.[28] To smooth hard distance bin assignments for distances that fall within 0.2 Å of a distance boundary (e.g., 4.5 Å as border between distance bins 1 and 2), two resulting PFP pharmacophore triplets have been recorded using both distance bins individually. PFP have been generated for each molecule using either one Corina[29]-generated conformation only (PFP1) or using the union of fingerprints over 50 OMEGA-generated conformations (PFP50). The Atom Pair[30]

10.1021/jm050468i CCC: $33.50 
Published on Web 02/04/2006

(AP) and Ghose and Crippen[31] descriptors have been implemented in-house.

**Ranking Methods.** 1. Tanimoto Average (TA): For each compound $j$ the average Tanimoto similarity $TA_j$ has been calculated as

$$TA_j = \frac{1}{\sum_{i=1}^{N_{active}} (1 - \delta_{ij})} \sum_{i=1}^{N_{active}} \tau_{ij}(1 - \delta_{ij}) \quad (1)$$

where $\tau_{ij}$ is the Tanimoto similarity between an active compound $i$ and all database compounds (actives and negative controls) $j$. $N_{active}$ is the number of all active compounds in the virtual screening database. $\delta_{ij}$ is the Kronecker Delta function.

2. Rank Average (RA): For each compound $j$ the average rank $RA_j$ has been calculated as

$$RA_j = \frac{1}{\sum_{i=1}^{N_{active}} (1 - \delta_{ij})} \sum_{i=1}^{N_{active}} \text{rank}_{ij}(1 - \delta_{ij}) \quad (2)$$

where $\text{rank}_{ij}$ is defined as the rank compound $j$ assumes when compared to active $i$ according to its $\tau_{ij}$ when all $\tau_{ij}$ are sorted in descending order.

3. Voting (V2): For each compound $j$ the average rank $V2_j$ has been calculated as

$$V2_j = \frac{1}{\sum_{i=1}^{N_{active}} (1 - \delta_{ij})} \sum_{i=1}^{N_{active}} \text{vote}_{ij}(1 - \delta_{ij}) \quad (3)$$

where $\text{vote}_{ij} = \max(0, \text{int}(11 - \text{rank}_{ij}/0.02N_{DB}))$, and $N_{DB}$ is the number of compounds in the entire virtual screening database. This procedure gives 10 votes to the first 2% of compounds, 9 votes for the next 2%, and so on. Finally, for compounds in the 18−20% rank bracket, 1 vote is given. Compounds in the bottom 80% of the ranking list for each individual active receive no votes. The votes for each compound are then averaged over all actives in the dataset.

4. Voting with Tanimoto cutoff (VT): For each compound $j$ the average rank $VT_j$ has been calculated as in eq 3 with the additional condition that $\text{vote}_{ij} = 0$ if $\tau_{ij} < \tau_{threshold}$. For the current study $\tau_{threshold}$ has been set to 0.35 as it has appeared to be the optimal number throughout a variety of data sets studied.

5. Consensus ranking (CR): Although several ideas of consensus scoring have been explored in the context of structure-based virtual screening,[32,33] consensus ranking in the context of similarity ranking has not been a focus in the past. Nevertheless, it has been recognized by others as a desirable path forward.[17] Therefore we have devised a consensus method that works similarly to a clustering algorithm with predefined cluster variance. The idea is to find the largest cluster of similar ranks among all ranking methods applied. Given a maximum allowed rank distance among a set of ranks for a given compound, the largest cluster is being identified and the average rank is taken among the ranks in this cluster. If no cluster can be established (all ranks are singletons), the average over the ranks of all methods is taken for the given compound.

The consensus rank idea described above is implemented as follows. For each of the four methods described above, the respective ranks for method $k$ and compound $j$ are called $r\text{Method}_k^j$. All pairwise rank differences between the methods for a given compound $j$ are calculated as

$$\Delta_j(\text{Method}_k, \text{Method}_l) = \text{abs}(r\text{Method}_k^j - r\text{Method}_l^j) \quad (4)$$

A threshold distance $\Delta_{threshold}$ is introduced to define a binary function that reads

$$\delta_j(\text{Method}_k, \text{Method}_l) = \begin{cases} 1 & \text{if } \Delta_j(\text{Method}_k, \text{Method}_l) \le \Delta_{threshold} \\ 0 & \text{if } \Delta_j(\text{Method}_k, \text{Method}_l) < \Delta_{threshold} \end{cases} \quad (5)$$

In addition, we define $N_j^k$ for a method $k$ and a compound $j$ as

$$N_j^k = 1 + \sum_{l \neq k} \delta_j(\text{Method}_k, \text{Method}_l) \quad (6)$$

Intermediate consensus ranks for a compound $j$ centering on each of the methods $k$ individually are then defined as

$$CR_j^k = \frac{1}{N_j^k}(r\text{Method}_k^j + \sum_{l \neq k} \delta_j(\text{Method}_k, \text{Method}_l) r\text{Method}_l^j) \quad (7)$$

Examining all four intermediate $CR_j^k$ the following rules are applied to choose the final consensus rank $CR_j$ for a given compound $j$: The maximum of $N_j^k$ for a given compound $j$ and all methods $k$, $N_j^{max}$, is examined among the four $CR_j^k$. If $N_j^{max} = 1$, no consensus has been reached. Therefore, the average over all four methods is used to calculate $CR_j$. If $N_j^{max} > 1$ and only one $N_j^k$ assumes $N_j^{max}$ then $CR_j = CR_j^k$ for this method. If $N_j^k = N_j^{max} > 1$ for several methods $k$, then among those, the method with the smallest sum of rank difference, $S\Delta_j^k$, is chosen to calculate $CR_j$.

$$S\Delta_j^k = \sum_{l \neq k} \delta_j(\text{Method}_k, \text{Method}_l)\Delta_j(\text{Method}_k, \text{Method}_l) \quad (8)$$

Note that it is possible due to the threshold rules that the voting methods produce a 0 score with the result that the ranks of all respective compounds become identical. Therefore, methods for which the score becomes 0 for a given compound $j$ are excluded for its consensus calculations. Throughout this study $\Delta_{threshold}$ has been set to 80.

6. Maximum Rank (MR) and Maximum Tanimoto (MT): Complementing the average Tanimoto (TA) and average rank (RA) scores, MR and MT have been calculated by taking the maximum rank or maximum Tanimoto similarity a given compound $j$ experiences when compared to all actives $i$.

$$MT_j = \max(\tau_{ij}(i = 1..N_{active})); \quad MR_j = \max(\text{rank}_{ij}(i = 1..N_{active})) \quad (9)$$

**Data Preparation.** The following data sets of 'actives' have been prepared from the literature for the present study: 12 CDK2 ligands,[34] 13 COX2 inhibitors,[35] 12 estrogen receptor (ER) ligands,[36] 37 11 HIV-1 protease inhibitors,[38] 4 neuraminidase inbibitors,[39] 6 p38 MAP kinase inhibitors,[40] and 8 thrombin inhibitors.[41] These compounds are depicted in Figures 1−7. Note that the compounds have been chosen with maximum structural diversity in mind. Each compound is meant to represent its own chemotype in order to recognize each virtual screening hit based on descriptor similarity as scaffold hopping event. Therefore,

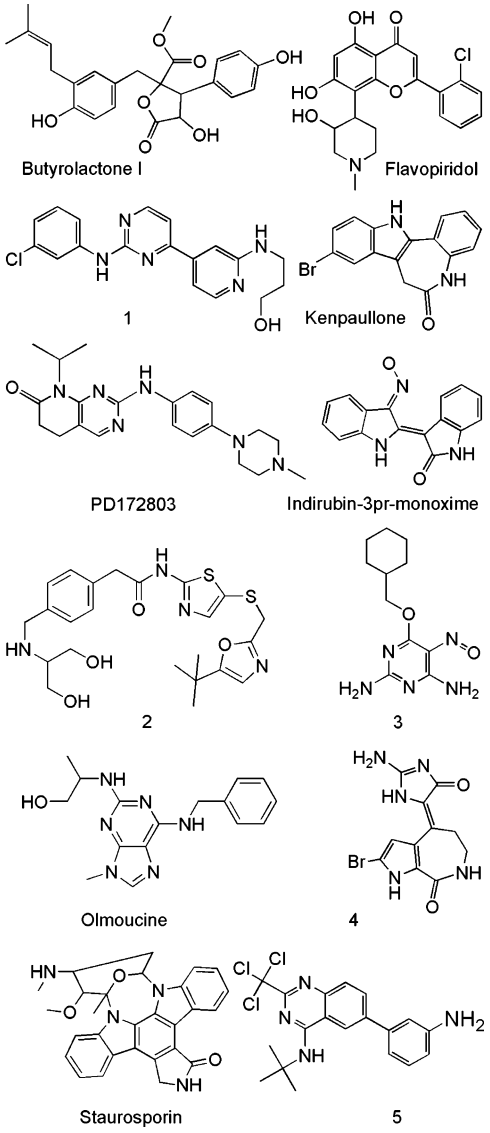


**Figure 1.** CDK2 ligands.

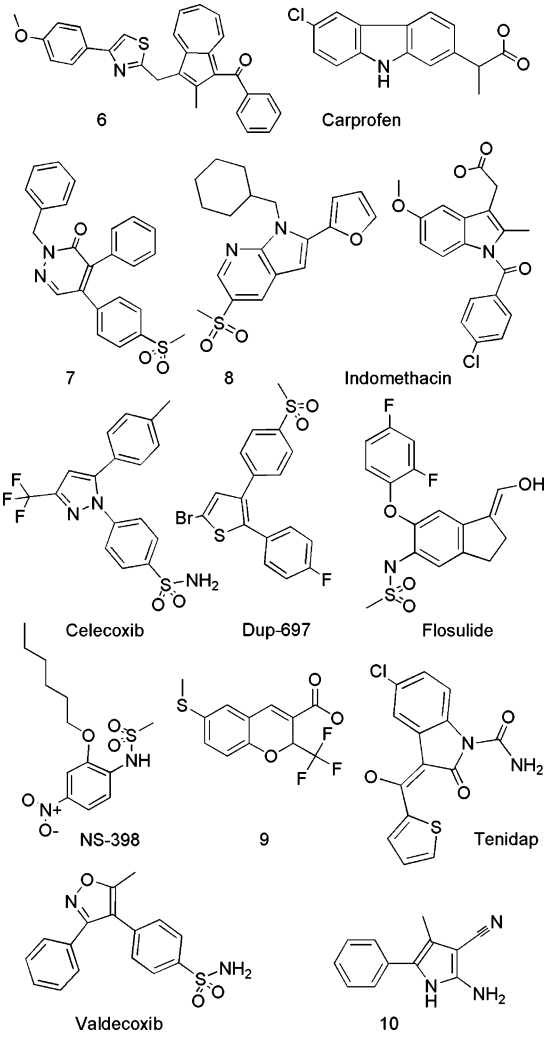


**Figure 2.** COX2 ligands.

the number of compounds in each dataset varies between 4 (neuraminidase) and 13 (COX2). These compounds serve as the only actives in the virtual screening studies. The actives are combined with 9969 drug-like compounds from the MDDR database as putative negative controls or 'inactives' (none of which are considered active). The MDDR compounds have been chosen following rules outlined elsewhere.[42] 3D structures of all compounds have been generated using Corina.[29] Conformations have been generated using OMEGA.[28] Daylight fingerprints (DF) of length 1024 have been generated for all compounds. The ionization states of the compounds have been determined using the Ionizer tool from Schrodinger Inc.[43]

Please note again that there are no other actives in the data set than those in the 'active' set for each target. For each individual similarity based virtual screen one of the actives against a given target is considered as template. All other actives are then mixed with the inactives from the MDDR compound set, and the similarity of the template to each compound in this mixed set is calculated. This procedure ensures that each recognition event between actives against the same target is guaranteed to be a scaffold hopping event. The various ranking methods described above are then applied as different techniques to combine the individual virtual screening results for all active compounds serving in succession as templates for a given target.

## Results and Discussion

**Scaffold Hopping Potential of Different Descriptors.** The ability of 2D and 3D descriptors to recognize different active scaffolds based on similarity is illustrated in Figure 8. The figure shows a set of 537 active kinase inhibitors (activities range between 5 and 500 nM $IC_{50}$ in a molecular assay for a kinase target) that belong to five different chemical classes: 315 carbolines, 150 indolinones, 44 diaminopyrimidines, 23 benzimidazoles, and 5 compounds with other scaffolds grouped together into one class. For each compound, the similarity against all the other 536 actives has been calculated using the Tanimoto coefficient and the respective descriptors. For each compound, the best rank for a compound belonging to a different compound class has been calculated and averaged over all 537 compounds. For each descriptor, this procedure has resulted in four points representing the average recognition of the first representative for each of the four other scaffolds. If one particularly focuses on the recognition of the first alternative scaffold, Figure 8 illustrates that the PFP fingerprint has the best potential of finding a novel chemotype. Topological descriptors such as Ghose and Crippen atom types,[31] atom pairs,[30] and Daylight fingerprints[44] are expectedly less able to find different scaffolds before exhausting same scaffold analogues. This is particularly apparent for the Daylight fingerprint descriptor where the average recognition of a different scaffold coincides with the theoretically worst case of scaffold hopping (all compounds of a particular scaffold are recognized first

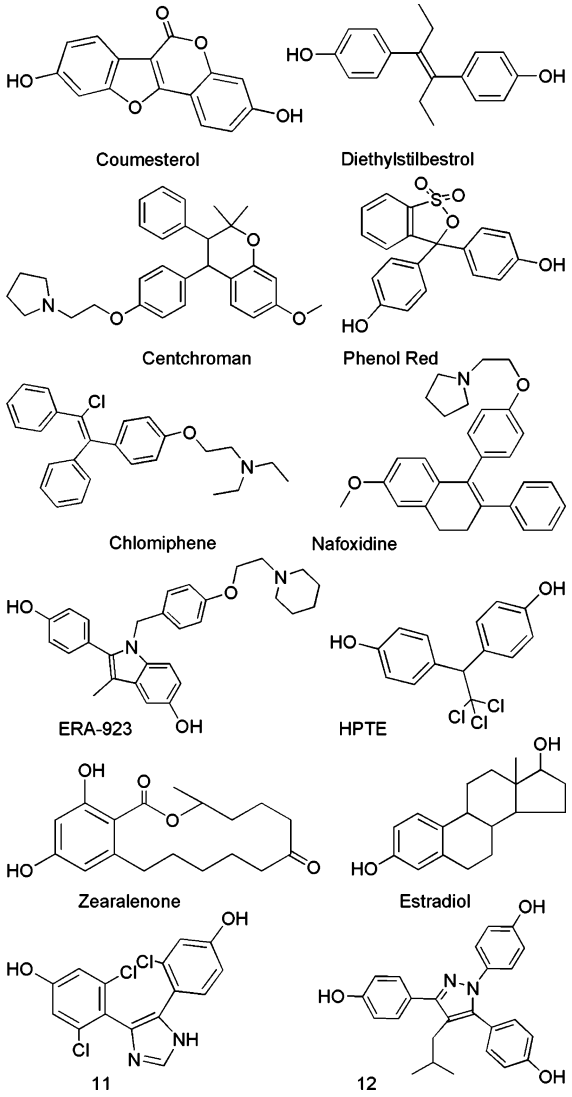


**Figure 3.** Estrogen receptor ligands.

before a compound of a different scaffold is chosen). This behavior has been observed before.[5,17] It is no shortcoming of the Daylight fingerprint but rather an expression of the strength of the Daylight fingerprints as a topological descriptor. As we expect, Daylight fingerprints first recognize the similarity to all compounds with the same scaffold before accepting others.

Examining the recognition of compounds within a set of actives provides a feel for the ability of a descriptor to recognize similarities across chemotypes. However, the above analysis does not answer the question of how these descriptors are able to identify bioactive compounds with novel scaffolds in a virtual screening scenario where putative bioactive compounds are mixed with large numbers of inactives. This point can easily be understood if one considers the limiting case where molecules are ordered at random. In that case, one can easily find novel scaffolds with high ranks, but the enrichment will be near 1.0 because the random case makes no distinction between active and inactive molecules. Hence the bulk of this study is concerned with the behavior of 3D and 2D descriptors in a virtual screening setting.

**Scaffold Hopping through Compound Similarity-Based Virtual Screening.** Seven potential drug discovery targets have been chosen to investigate the scaffold hopping capabilities of PFP, atom pair, and Daylight fingerprints as representatives of 3D and 2D descriptors: CDK2, estrogen receptor (ER), COX-2, neuraminidase, HIV-1 protease, p38 MAP kinase, and thrombin. Figures 1−7 illustrate the active compounds used in the study. Each compound has been chosen to represent its own chemotype. Depending on the published ligand structures this has resulted in different numbers of compounds being used for each of the targets. In the virtual screening studies presented here the ability to recognize known ligands with higher similarity than putative inactives represented by a set of 9969 compounds taken from the MDDR database is considered a scaffold hopping event. Since we cannot guarantee that the MDDR compounds are not active against one of our targets (since not every molecule has been tested against every activity), there may be a small amount of noise in this study. Therefore, the enrichments in our tables and figures maybe somewhat underestimated relative to an ideal list of compounds that contains no false inactives.

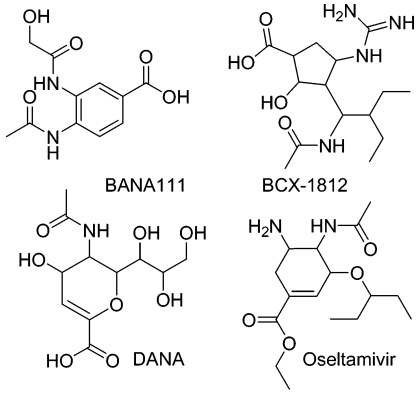


**Figure 4.** Neuraminidase ligands.

As suggested by others[26] for this study, we have decided to make the 2D topological fingerprint a standard for comparison also. The design of the study includes two different scenarios.

(I) The distribution of pairwise Tanimoto similarities of Daylight fingerprints among the active compounds for a given target is very similar to the distribution comparing the actives against the MDDR negative control compounds. This is true for the targets CDK2, COX2, and to a lesser extent for ER (Figure 9). For these cases the topological fingerprints are obviously less likely to be able to identify actives of other structural classes among the negative control compounds from the MDDR. In this case of low 2D descriptor competitiveness we expect to see the PFP descriptors perform significantly better than the topological fingerprint descriptor.

(II) The distribution of pairwise Tanimoto similarities of Daylight fingerprints among the active compounds for a given target is shifted toward higher similarities compared to the same distribution between the actives and the MDDR negative controls. This is true for the targets neuraminidase, HIV-1 protease, p38 MAP kinase, and thrombin (Figure 10). For these cases the topological fingerprints are much better suited for picking up actives in upper ranks of the virtual screen. In these cases of high 2D descriptor competitiveness we hope to see the 3D descriptor perform at least as good as the 2D descriptor.

Recent work has mainly focused on finding optimal 3D descriptors for compound similarity-based virtual screening and scaffold hopping.[15,17] Here we limit ourselves to only one type of 3D descriptor (PFP) and focus instead on optimizing the scoring (ranking) method for virtual screening and scaffold hopping. As outlined in the Methods section we have employed a series of different ranking methods to determine an optimum virtual screening protocol for scaffold hopping: Tanimoto averaging (TA), rank averaging (RA), voting (V2), Tanimoto

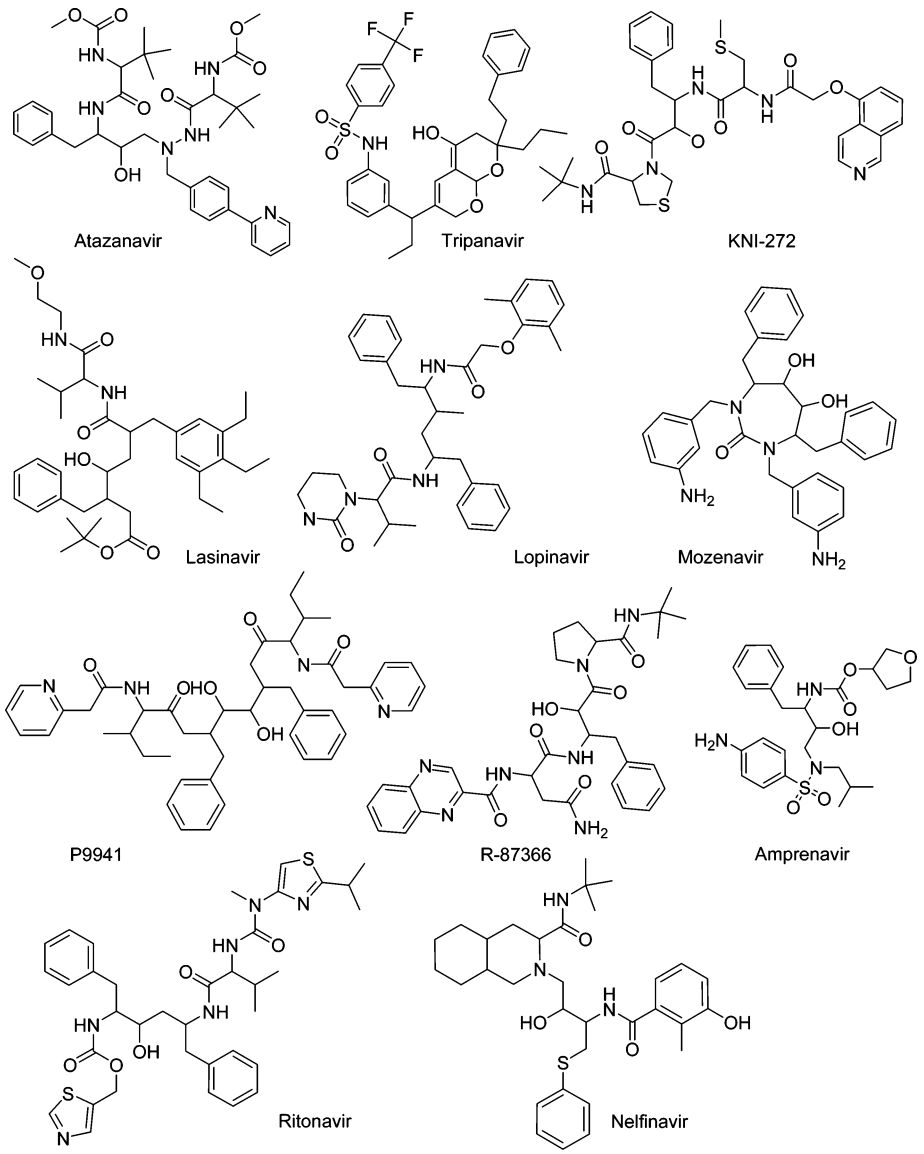


**Figure 5.** HIV-1 protease ligands.

**Table 1.** Effect of Consensus Scoring on Enrichment Factors at 2% of Database Sampled[a]

| | PFP1 | | | | | | | PFP50 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| screen | TA | RA | V2 | VT | MT | MR | CR | TA | RA | V2 | VT | MT | MR | CR |
| CDK2 | 0.0 | 12.5 | 12.5 | 0.0 | 0.0 | 0.0 | 12.5 | 4.2 | 4.2 | 4.2 | 0.0 | 0.0 | 4.2 | 4.2 |
| COX2 | 3.9 | 3.9 | 3.9 | 7.7 | 7.7 | 7.7 | 3.9 | 11.5 | 3.9 | 7.7 | 7.7 | 7.7 | 7.7 | 11.5 |
| ER | 16.7 | 0.0 | 4.2 | 16.7 | 16.7 | 33.3 | 8.3 | 8.3 | 0.0 | 8.3 | 12.5 | 8.3 | 29.2 | 8.3 |
| NEUR | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 37.6 | 50.0 | 50.0 | 25.0 | 12.5 | 25.0 | 25.0 | 37.5 | 37.5 |
| HIV1 | 31.8 | 27.3 | 22.7 | 31.8 | 31.8 | 27.3 | 22.7 | 36.4 | 27.3 | 27.3 | 36.4 | 36.4 | 31.8 | 36.4 |
| P38 | 8.3 | 8.3 | 25.0 | 0.0 | 0.0 | 0.0 | 8.3 | 0.0 | 0.0 | 16.7 | 16.7 | 0.0 | 0.0 | 25.0 |
| THROM | 0.0 | 0.0 | 0.0 | 12.5 | 12.5 | 0.0 | 0.0 | 6.3 | 6.3 | 6.3 | 6.3 | 0.0 | 6.3 | 6.3 |
| | DF | | | | | | | AP | | | | | | |
| screen | TA | RA | V2 | VT | MT | MR | CR | TA | RA | V2 | VT | MT | MR | CR |
| CDK2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 8.3 | 8.3 | 0.0 |
| COX2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 11.5 | 7.7 | 11.5 | 11.5 | 7.7 | 11.5 | 11.5 |
| ER | 8.3 | 0.0 | 0.0 | 16.7 | 0.0 | 16.7 | 4.2 | 8.3 | 0.0 | 4.2 | 12.5 | 16.7 | 12.5 | 4.2 |
| NEUR | 12.5 | 0.0 | 0.0 | 25.0 | 25.0 | 25.0 | 0.0 | 37.5 | 12.5 | 37.5 | 37.5 | 25.0 | 37.5 | 37.5 |
| HIV1 | 9.1 | 18.2 | 18.2 | 9.1 | 0.0 | 0.0 | 18.2 | 27.3 | 22.7 | 22.7 | 27.3 | 36.4 | 27.3 | 22.7 |
| P38 | 8.3 | 8.3 | 8.3 | 0.0 | 0.0 | 0.0 | 8.3 | 25.0 | 8.3 | 25.0 | 25.0 | 25.0 | 25.0 | 25.0 |
| THROM | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 18.8 | 0.0 | 18.8 | 18.8 | 0.0 | 25.0 | 12.5 |

[a] The maximum enrichment factor that can be achieved is 50 for 2% sampling.

threshold voting (VT), maximum Tanimoto (MT), maximum rank (MR), and a consensus ranking approach among the first four methods (CR). In addition, minimum rank and minimum Tanimoto similarity methods have been tested; however, because of their inferior performance, results are not reported here in detail. Figures 11−17 examine the performance of the four

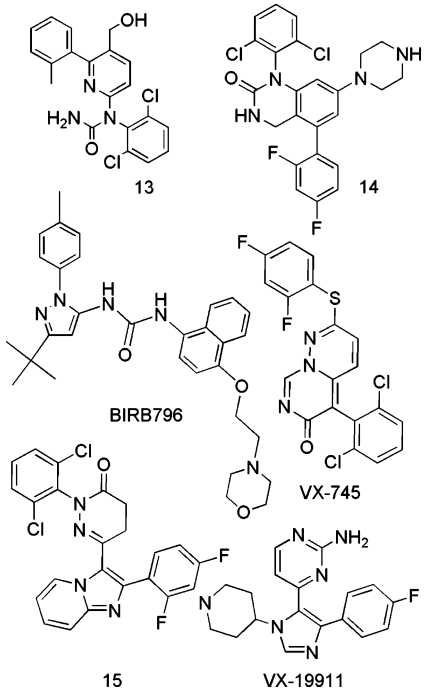


**Figure 6.** p38 MAP kinase ligands.

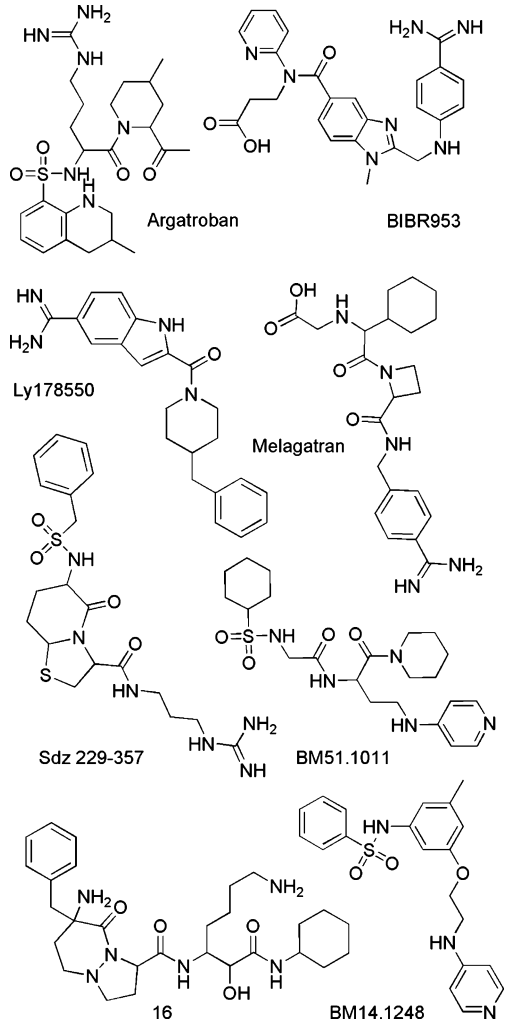


**Figure 7.** Thrombin ligands.

primary ranking techniques using Daylight fingerprints, atom pairs, and 3D pharmacophore fingerprints as descriptors for the

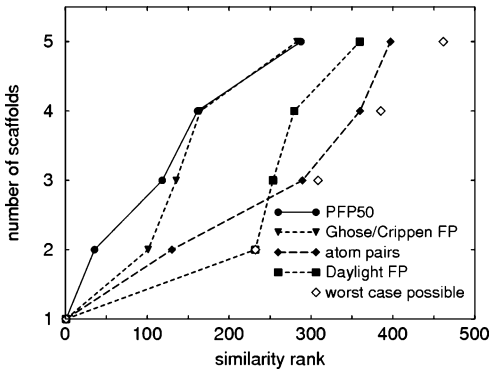


**Figure 8.** Scaffold hopping among 537 actives for a particular kinase belonging to 5 scaffold classes studied for different descriptors.

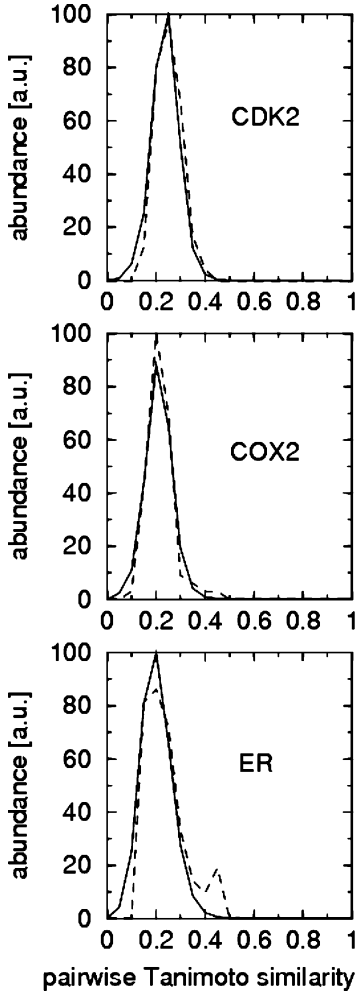


**Figure 9.** Distribution of pairwise Tanimoto similarities measured using Daylight fingerprints for all active ligand-active ligand pairs (dashed line) and all active ligand-MDDR negative control compound pairs (solid line) for the indicated targets. The distributions are normalized by the area under the curve.

seven targets introduced above. For the 3D pharmacophore fingerpints the figures contain curves for a single 3D conformation generated by using Corina,[29] (PFP1) as well as a fingerprint that has combined the pharmacophore fingerprint information from 50 OMEGA[28] generated conformations (PFP50).

Table 1 contains the enrichment factors for the individual ranking methods and the consensus ranking after the top-ranking 2% of the database have been retrieved. The enrichment factor is defined as the quotient between percent actives retrieved and

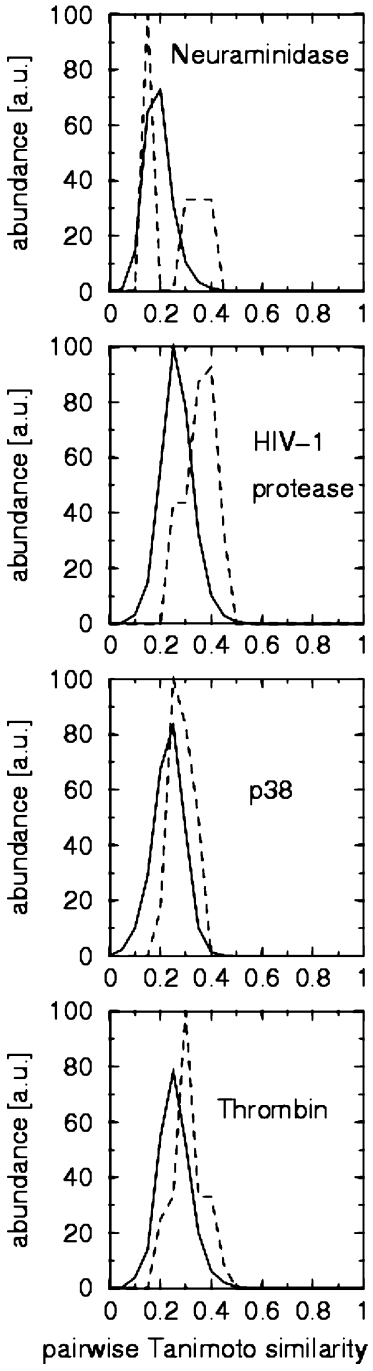


**Figure 10.** Distribution of pairwise Tanimoto similarities measured using Daylight fingerprints for all active ligands−active ligand pairs (dashed line) and all active ligand-MDDR negative control compound pairs (solid line) for the indicated targets. The distributions are normalized by the area under the curve.

percent of total database compounds (actives + negative controls) retrieved. Examining Table 1 reveals the following performance trends. Single scoring methods vary significantly in performance. MR performs best among all methods for the atom pair descriptor. VT is best with DF, CR is best with PFP50 (followed by TA as individual method), and V2 is best with PFP1. The consensus score is calculated using four individual methods (TA, RA, V2, VT) throughout this article. Although including MR and MT increases the consensus scores, in some cases it does not do so consistently (data not shown). Consensus scoring, with one exception (p38, PFP50), never leads to results better than the best individual scoring method used to calculate CR. In 12 cases the consensus method mirrors the best individual method used to derive CR, in 7 cases consensus scoring has an averaging effect, and only in 5 cases consensus scoring mirrors the worst individual method. In 5 cases all methods yield the same enrichment. This observation suggests that consensus scoring leads to more robust results compared to the individual methods. For the consensus score, PFP1 and PFP50 perform slightly better overall than AP does. However, studying the best descriptors for each ranking methods reveals that AP descriptors perform best for VT, MR, and MT. For V2, AP and PFP1 perform equally well. For TA, PFP50 performs best; for RA, PFP1 shows the best results, and for CR, PFP1 and PFP50 perform equally well. It is also noticeable that in most cases, particularly for methods that rely on ranking, voting improves the performance.

Figures 11−13 examine virtual screening data for scenario I (no topological similarity bias exists among actives compared to actives vs inactives, see Figure 9). Figure 11 shows a comparison of PFP1 (green), PFP50 (red), AP (blue), and DF (black) percent CDK2 actives retrieved as a function of the virtual screening rank according to four different ranking schemes. As expected for a scenario I case, the enrichment of actives is much better for the 3D fingerprint than it is for the topological fingerprint, which is close to the result that would have been obtained randomly. Interestingly, while voting improves the performance of DF, rank averaging (RA) performs best for the PFP method very closely followed by V2, albeit V2 delivers recognition of the first actives at a slightly better rank. The Tanimoto-based approaches, TA and VT, perform significantly worse. Somewhat surprisingly it has been found that involving multiple conformations does not improve the virtual screening performance for CDK2. However, as observed for other targets also, PFP50 offers early recognition of actives at ranks higher than those for the best PFP1 retrieved actives. Consensus ranking at the 2% database retrieval point is able to provide enrichment that is as good as that of the best single method (Table 1).

Figure 12 shows the virtual screening results for COX-2. As illustrated in Figure 9, COX2 belongs to scenario I also. Accordingly, DF performs similarly to CDK2 delivering little enrichment that is close to random. The PFP methods lead to higher enrichment. Interestingly, PFP1 and PFP50 perform more similarly albeit PFP50 performs better at early recognition of actives. More importantly, Table 1 shows that the consensus approach is able to carry the favorable PFP50 performance using TA over to CR resulting in a significantly better consensus-based performance of PFP50 over PFP1 at the 2% mark. The AP descriptor appears particularly strong for V2. At the 2% mark AP and PFP50 perform equally well using consensus scoring.

For estrogen receptor (Figure 13) all descriptors perform similarly using different methods, with the exception of RA for which DF and AP perform better than the PFP descriptors. However, all descriptors perform significantly worse using the RA method compared to the other methods. The two voting methods perform significantly better than their respective counterparts (the Tanimoto-similarity-emphasizing VT improves compared to TA and the rank-based V2 improves compared to RA). The PFP1 descriptor using the VT ranking method performs best. Particularly, PFP1 exhibits a very good enrichment at low ranks. The consensus score acts as averaging function between the methods (Table 1).

Figures 14−17 examine virtual screening data for scenario II (there is a topological similarity bias among actives compared

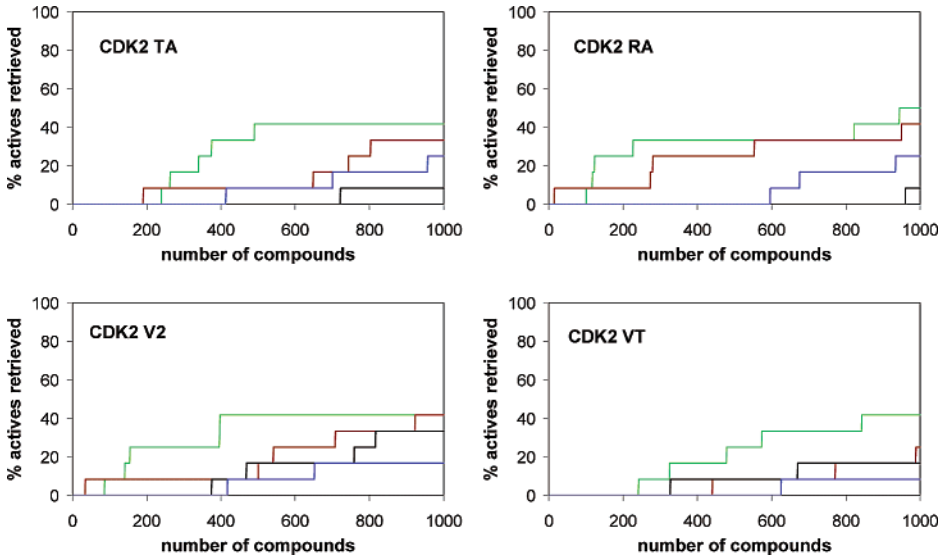


**Figure 11.** CDK2: Percentage of active ligands retrieved as function of virtual screening rank using Daylight fingerprints (black), atom pairs (blue) single confirmation pharmacophore fingerprints (PFP1) (green) and pharmacophore fingerprint descriptor with 50 conformations (PFP50) (red) for different ranking methods.

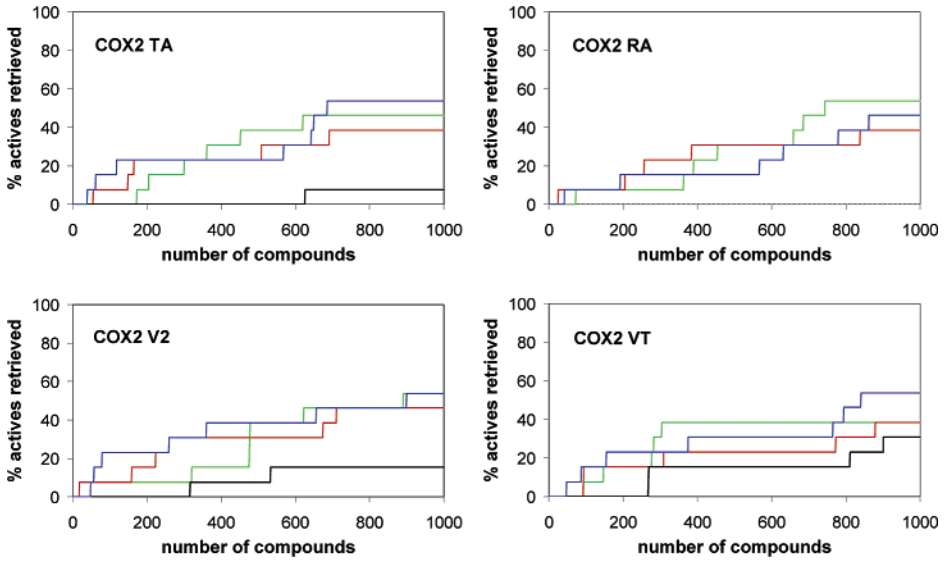


**Figure 12.** COX2: Percentage of active ligands retrieved as function of virtual screening rank using Daylight fingerprints (black), atom pairs (blue) single confirmation pharmacophore fingerprints (PFP1) (green) and pharmacophore fingerprint descriptor with 50 conformations (PFP50) (red) for different ranking methods.

to actives vs inactives, see Figure 10). Neuraminidase exhibits the best virtual screening performance among the selected targets (Figure 14). This has to be expected because of the prevalence of comparably rare ionizable pharmacophore features being present in the actives. Because these prominent pharmacophore features are better represented in the PFP descriptors, it is not surprising that they perform better here than DF, albeit DF shows a very good performance too as has to be expected for scenario II. It is interesting to note that for these high performance cases, voting does not seem to be beneficial for the PFP methods. VT performs particularly badly for PFP1. Nevertheless, the consensus method is able to retrieve an optimal enrichment (Table 1). The AP descriptor performs strongly with the Tanimoto based approaches outranking the PFP methods particularly at low ranks.

Figure 15 illustrates the performance of PFP, AP, and DF descriptors for HIV-1 protease. The PFP and AP descriptors perform particularly well with PFP50 being best at the rank based approaches. For the nonvoting methods using multiple conformations seems beneficial here. Interestingly, the PFP and AP descriptors consistently outperform DF for all methods despite a clear topological preference existing among the HIV-1 ligands (Figure 15). Consensus scoring, again, is able to preserve the best enrichment among all methods (Tables 1 and 2) for PFP50, AP, and DF (worst for PFP1).

Figure 16 shows for p38 MAP kinase that rank-based methods prevail over Tanimoto-based methods. PFP descriptors perform better than DF. AP performs best for the Tanimoto based approaches. Voting improves the enrichment, particularly for the ranking methods. Although AP outperforms the PFP descriptors for most methods, PFP50 manages to improve his consensus score (the only case observed) over all individual methods to deliver an enrichment which is identical to that obtained with AP (Table 1).

For thrombin, Figure 17 shows that voting significantly improves the scaffold hopping performance for both the rank-based and Tanimoto-based methods, particularly for the PFP descriptors. AP performs significantly better than all other

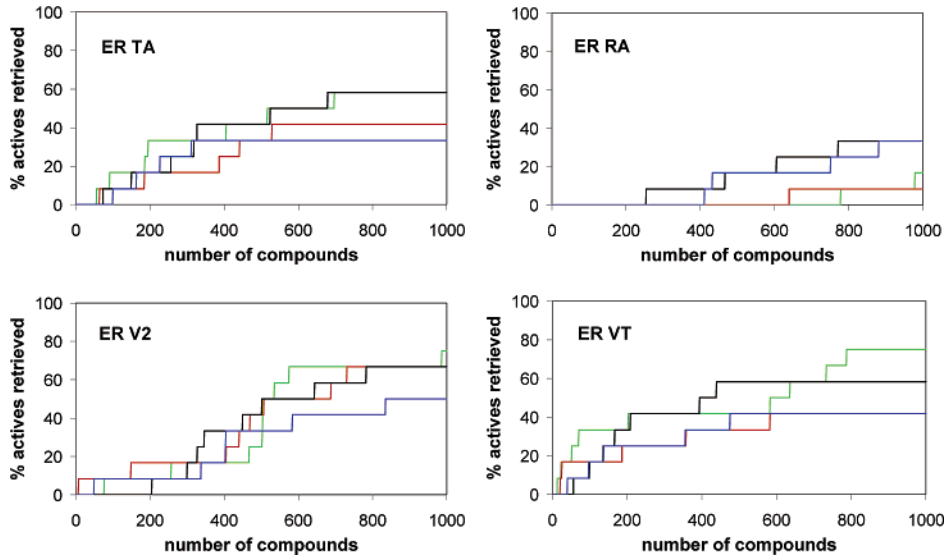


**Figure 13.** ER: Percentage of active ligands retrieved as function of virtual screening rank using Daylight fingerprints (black), atom pairs (blue) single confirmation pharmacophore fingerprints (PFP1) (green) and pharmacophore fingerprint descriptor with 50 conformations (PFP50) (red) for different ranking methods.

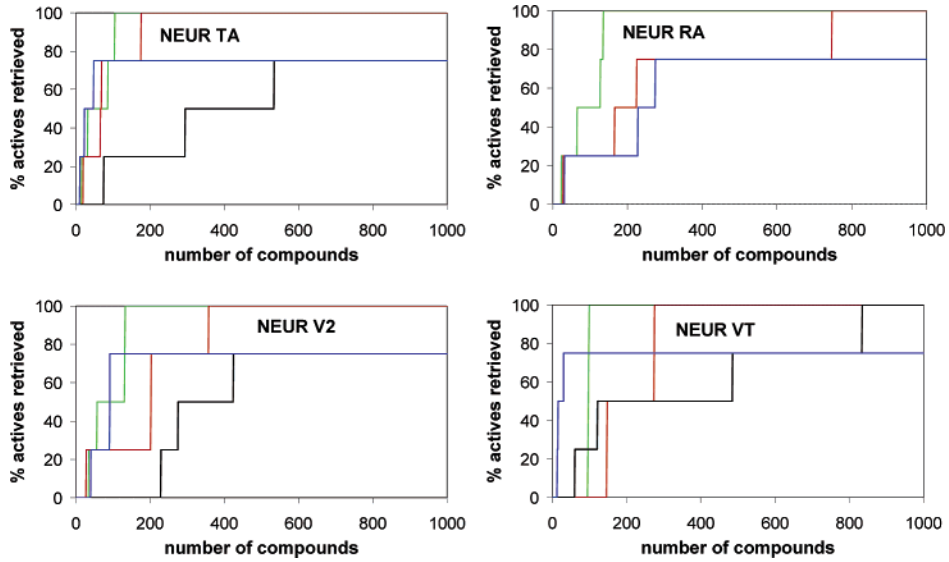


**Figure 14.** Neuramindiase: Percentage of active ligands retrieved as function of virtual screening rank using Daylight fingerprints (black), atom pairs (blue) single confirmation pharmacophore fingerprints (PFP1) (green) and pharmacophore fingerprint descriptor with 50 conformations (PFP50) (red) for different ranking methods.

descriptors for thrombin. Unfortunately, for thrombin the consensus scoring method has not been able to overcome the bad performance of three of the four methods at the 2% mark for PFP1. The good performance of VT has not resulted in a better CR. PFP1 has performed weaker for thrombin than for other targets while AP and DF have performed better. A possible explanation for this week performance could be the high flexibility of the thrombin ligands used in the study. To test this hypothesis, PFP1 results have been obtained using cocrystal structures of thrombin ligands when available instead of Corina generated structures. Unfortunately, this modified PFP1 descriptor has not improved performance (data not shown).

Generally it is observed that Daylight fingerprints do not perform well for scaffold hopping. This observation is not surprising and has been noted before.[5,17] Also, the good performance of atom pair descriptors has been noted before. Encouraging is the finding that in case of no topological bias (e.g., for CDK2) 3D pharmacophore descriptors perform better than 2D descriptors.

**Comparison with Structure-Based Virtual Screening.** As discussed above, Table 1 summarizes the enrichment factors achieved in the ligand-based virtual screens against the seven target proteins studied. For the same targets and compound sets, Table 2 compares the achieved enrichment factors from the similarity virtual screens with those obtained using Glide2.5 docking against the same target crystal structures as used by others for virtual screening studies of the same targets.[46,47] In addition to the enrichment factors provided in Table 2, Figure 18 shows the percent actives retrieved at 2, 5, and 10% of the highest ranks of the entire database for the Glide2.5 virtual screen and the similarity based virtual screening methods. Comparing the similarity-based virtual screening results with Glide2.5 enrichments, the docking method provides the highest enrichment only in case of COX2 followed closely by the AP descriptor. For all other targets, several similarity methods perform better than docking. PFP1, PFP50, and AP perform best in different cases. Glide2.5 performs significantly weaker than all similarity methods for the CDK2 and neuramindase

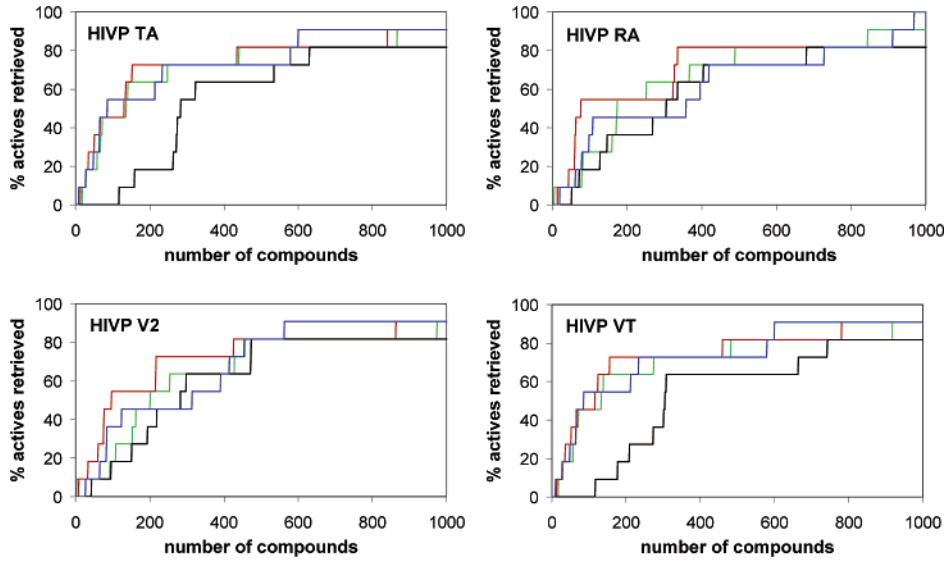


**Figure 15.** HIV-1 protease: Percentage of active ligands retrieved as function of virtual screening rank using Daylight fingerprints (black), atom pairs (blue) single confirmation pharmacophore fingerprints (PFP1) (green) and pharmacophore fingerprint descriptor with 50 conformations (PFP50) (red) for different ranking methods.

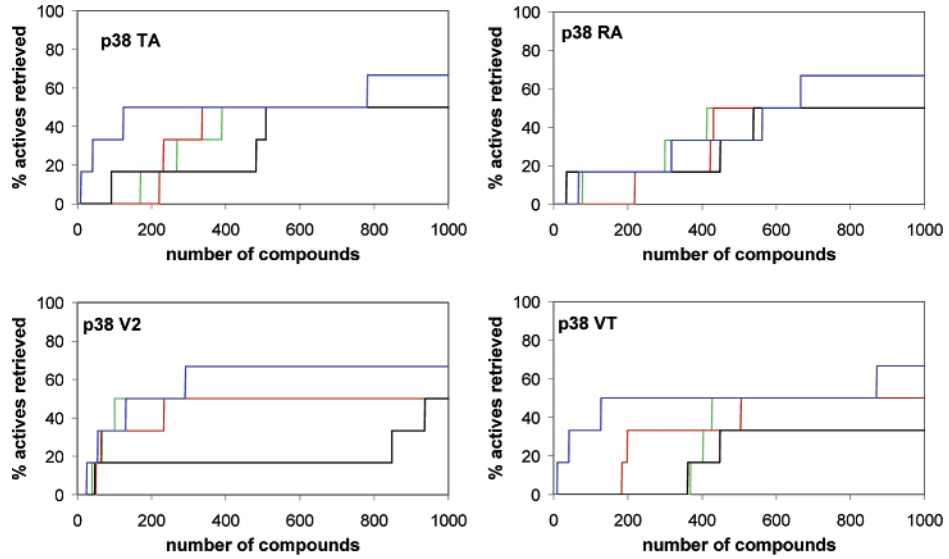


**Figure 16.** p38 MAP kinase: Percentage of active ligands retrieved as function of virtual screening rank using Daylight fingerprints (black), atom pairs (blue) single confirmation pharmacophore fingerprints (PFP1) (green) and pharmacophore fingerprint descriptor with 50 conformations (PFP50) (red) for different ranking methods.

targets. This finding is particularly noteworthy comparing the Glide2.5 virtual screening results with those reported in the literature for the same targets (using the same crystal structures of the targets) but using different sets of actives and inactives (Table 3).[45,46] As we have observed before, enrichment factors cannot always be compared on equal footing when different actives (and numbers of actives) and inactives are used.[47] Nevertheless, the large discrepancies in docking results presented in Table 3 are somewhat surprising. Scaffold hopping is often taken for granted in structure-based approaches because no information about ligands is taken into account. However, the actual ability of structure-based approaches to scaffold hop has rarely been quantified. While reports of enrichment factors in structure-based virtual screening experiments are abundant, no comprehensive study has been presented yet illustrating how much ligand structure similarity among actives contributes to their enrichment in structure-based virtual screens. While much care is taken in diversifying decoys[48] for structure-based virtual screening experiments, actives often come from the same class of compounds. The results presented in Table 3 suggest that investigating scaffold hopping abilities of docking approaches will be a worthwhile topic for future research.

As noted before by others[5,17] ligand similarity-based methods, particularly the 3D pharmacophore fingerprint methods, perform surprisingly well compared to the structure-based methods. Especially comparing the methods on equal footing (Glide2.5 results and similarity-based results have been obtained with the same actives/negative controls) suggests that the knowledge of active ligands may be as valuable as a crystal structure for the purpose of virtual screening, particularly when new scaffolds need to be identified. However, to draw more solid conclusions from the presented comparison between docking and compound similarity approaches more extensive studies involving more targets, more diverse ligands and different crystal structures for docking experiments are needed. The latter point may be particularly relevant in cases of known induced fit effects such as for Phe-in and Phe-out binders in kinases such as p38 used in this study.

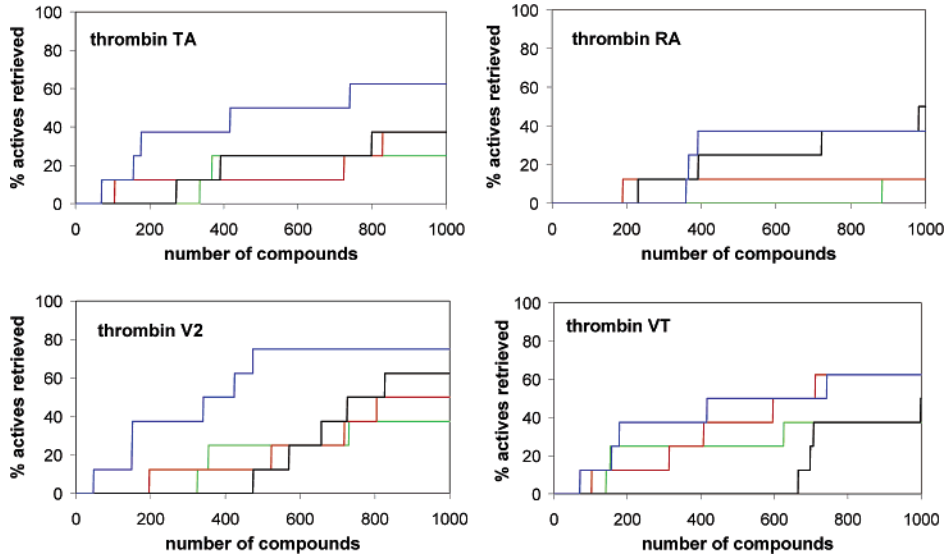


**Figure 17.** Thrombin: Percentage of active ligands retrieved as function of virtual screening rank using Daylight fingerprints (black), atom pairs (blue) single confirmation pharmacophore fingerprints (PFP1) (green) and pharmacophore fingerprint descriptor with 50 conformations (PFP50) (red) for different ranking methods.

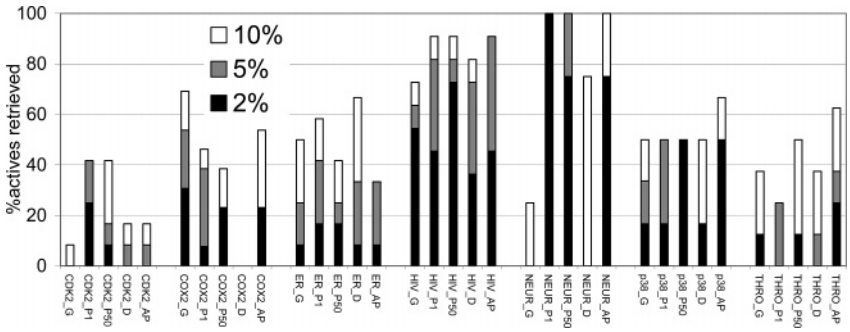


**Figure 18.** Comparison of percent actives retrieved using Glide2.5 docking and actives retrieved with similarity methods as described in this study for seven protein targets (CDK2, COX2, estrogen receptor, HIV-1 protease, neuramindase, p38 MAP kinase, and thrombin). The suffix _G refers to Glide2.5 docking results, _P1 refers to PFP1, _P50 refers to PFP50, _D refers to DF, and _AP refers to virtual screening results obtained with the AP descriptor. For the ligand-based descriptors the consensus ranking method (CR) has been used.

**Table 2.** Comparison of Enrichment Factors at 2% of Database Sampled[a]

| | Glide2.5 | PFP1 CR | PFP50 CR | DF CR | AP CR |
|---|---|---|---|---|---|
| CDK2 | 0.0 | 12.5 | 4.2 | 0.0 | 0.0 |
| COX2 | 15.4 | 3.9 | 11.5 | 0.0 | 11.5 |
| ER | 4.2 | 8.3 | 8.3 | 4.2 | 4.5 |
| NEUR | 0.0 | 50.0 | 37.5 | 0.0 | 37.5 |
| HIV1 | 27.3 | 22.7 | 36.4 | 18.2 | 22.7 |
| P38 | 8.3 | 8.3 | 25.0 | 8.3 | 25.0 |
| THROM | 6.3 | 0.0 | 6.3 | 0.0 | 12.5 |

[a] The maximum enrichment factor that can be achieved is 50 for 2% sampling. The in-house docking results reported for Glide2.5 are based on the same set of actives/negative controls used for obtaining the results for the similarity-based methods. The consensus scores (CR) are reported for the similarity-based methods.

**Table 3.** Comparison of Structure-based Virtual Screening Results with Literature Data[a]

| | ScreenScore Lit. | Glide2.5 Lit. | Glide2.5 in-house |
|---|---|---|---|
| CDK2 | - | 12.5 | 0.0 |
| COX2 | 10.0 | 16.6 | 15.4 |
| ER | 22.0 | 32.5 | 4.2 |
| NEUR | 11.0 | - | 0.0 |
| HIV1 | - | 36.7 | 27.3 |
| P38 | 12.0 | 7.8 | 8.3 |
| THROM | 26.0 | 25.0 | 6.3 |

[a] The maximum enrichment factor that can be achieved is 50 for 2% sampling. The GLIDE2.5 (column 2) and ScreenScore (column 1) data have been taken from the literature using sets of actives/negative controls different from those used in this study.[45,46] The in-house docking results reported for Glide2.5 (column 3) are based on the same set of actives/negative controls used for obtaining the results for the similarity-based methods (identical to first column of Table 2).

## Conclusions

Different ranking strategies for compound similarity-based virtual screening and scaffold hopping have been explored for seven well established drug targets. Significant differences between scoring methods have been observed. Rank-based and Tanimoto similarity-based methods show varying performance depending on the target at hand. Maximum ranking and maximum Tanimoto similarity against a set of know actives perform well in some cases. Voting methods have often but not consistently been observed to increase virtual screening performance. A simple consensus method has been introduced and applied. There has been only one example found where the consensus ranking has outperformed the best single ranking method. However, in the majority of cases the consensus score has been able to either display the highest rank achieved by any single method or to show an average between the methods used. For the PFP50 descriptor consensus scoring has performed better than any individual score considering the set of seven targets studied here. Therefore, it is recommended to use the consensus score because of the high volatility among ranking performances of the single functions used in this study.

Nevertheless it should be noted that the maximum rank performs also well particularly in combination with the atom pair descriptor yielding the overall best result comparing all scoring methods and descriptors averaged over all 7 targets. Especially the much better performance of atom pair descriptors for thrombin is noteworthy.

The design of the studies has allowed interpreting virtual screening hits as scaffold hopping events at the same time. Given this added challenge to a virtual screening setup the achieved enrichments of actives in upper ranks of the virtual screening hit lists, albeit varying greatly among targets, have generally been surprisingly high. Although the enrichment has been mostly higher in cases of greater topological similarities among the actives compared to negative controls, even in cases of low topological advantage the enrichment rates using pharmacophore fingerprint descriptors have been encouragingly high. In general, pharmacophore fingerprints have been shown to outperform topological descriptors represented by Daylight fingerprints even in cases of topological similarity bias among active ligands compared to negative controls. Interestingly, single conformation pharmacophore fingerprints have performed better than multi-conformational fingerprints in most cases. Applying known bioactive conformations for some of the thrombin inhibitors has not improved virtual screening results compared to Corina-generated conformations. Atom pair descriptors have performed well throughout the targets studies here. They have been particularly strong for thrombin. In other cases atom pair descriptors have performed comparably to pharmacophore fingerprint descriptors.

Compound similarity-based virtual screening and structure-based virtual screening results for the same targets and compound sets have been compared. Several of the ligand-based virtual screening methods have performed better than Glide2.5 docking for all but one target (COX2). Interestingly, the virtual screening enrichments obtained with Glide2.5 docking using the compound sets designed for this scaffold hopping study are mostly much lower than those reported in the literature for other data sets using the same targets and docking method. This observation illustrates the sensitivity of virtual screening outcomes to the composition of the data sets. More importantly, it also raises the question to what extent structure-based virtual screening results may depend on the diversity of the actives − a question that is worth exploring more thoroughly in the future.

Most encouraging is the finding that even in cases of no topological bias toward the pairwise similarities among the actives compared to the negative controls, particularly for CDK2, the pharmacophore fingerprint methods perform very strongly compared to the structure-based virtual screening method. This finding suggests that knowledge about active ligands for a drug target can be as valuable as a crystal structure for obtaining novel scaffolds from virtual screening.

**Acknowledgment.** The authors thank Dr. Miguel Teodoro for implementing the atom pair descriptor as well as for help in assembling the ligand data sets used in this study.

## References

(1) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894−2896.
(2) Brown, R. D.; Martin, Y. C. Use of Structure−Activity Data To Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572−584.
(3) Brown, R. D.; Martin, Y. C. The information content of 2D and 3D structural descriptors relevant to ligand−receptor binding. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1−9.
(4) Miller, M. D.; Sheridan, R. P.; Kearsley, S. K. SQ: a program for rapidly producing pharmacophorically relevent molecular superpositions. *J. Med. Chem.* **1999**, *42*, 1505−1514.
(5) Sheridan, R. P.; Kearsley, S. K. Why do we need so many chemical similarity search methods? *Drug Discovery Today* **2002**, *7*, 903−911.
(6) Pirard, B.; Brendel, J.; Peukert, S. The discovery of Kv1.5 blockers as a case study for the application of virtual screening approaches. *J. Chem. Inf. Model.* **2005**, *45*, 477−485.
(7) Cramer, R. D.; Jilek, R. J.; Guessregen, S.; Clark, S. J.; Wendt, B.; Clark, R. D. "Lead hopping". Validation of topomer similarity as a superior predictor of similar biological activities. *J. Med. Chem.* **2004**, *47*, 6777−6791.
(8) Rush, T. S., III; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein−protein interaction. *J. Med. Chem.* **2005**, *48*, 1489−1495.
(9) Jenkins, J. L.; Glick, M.; Davies, J. W. A 3D similarity method for scaffold hopping from known drugs or natural ligands to new chemotypes. *J. Med. Chem.* **2004**, *47*, 6144−6159.
(10) Lloyd, D. G.; Buenemann, C. L.; Todorov, N. P.; Manallack, D. T.; Dean, P. M. Scaffold hopping in de novo design. Ligand generation in the absence of receptor information. *J. Med. Chem.* **2004**, *47*, 493−496.
(11) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org. Biomol. Chem.* **2004**, *2*, 3256−3266.
(12) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1177−1185.
(13) Matter, H.; Schwab, W.; Barbier, D.; Billen, G.; Haase, B.; Neises, B.; Schudok, M.; Thorwart, W.; Schreuder, H.; Brachvogel, V.; Lonze, P.; Weithmann, K. U. Quantitative Structure−Activity Relationship of Human Neutrophil Collagenase (MMP-8) Inhibitors Using Comparative Molecular Field Analysis and X-ray Structure Analysis. *J. Med. Chem.* **1999**, *42*, 1908−1920.
(14) Andrews, K. M.; Cramer, R. D. Toward general methods of targeted library design: topomer shape similarity searching with diverse structures as queries. *J. Med. Chem.* **2000**, *43*, 1723−1740.
(15) Jenkins, J. L.; Glick, M.; Davies, J. W. A 3D similarity method for scaffold hopping from known drugs or natural ligands to new chemotypes. *J. Med. Chem.* **2004**, *47*, 6144−6159.
(16) Makara, G. M. Measuring molecular similarity and diversity: total pharmacophore diversity. *J. Med. Chem.* **2001**, *44*, 3563−3571.
(17) Good, A. C.; Hermsmeier, M. A.; Hindle, S. A. Measuring CAMD technique performance: a virtual screening case study in the design of validation experiments. *J. Comput.-Aided Mol. Des* **2004**, *18*, 529−536.
(18) Good, A. C.; Cho, S. J.; Mason, J. S. Descriptors you can count on? Normalized and filtered pharmacophore descriptors for virtual screening. *J. Comput.-Aided Mol. Des* **2004**, *18*, 523−527.
(19) Mason, J. S.; Cheney, D. L. Library design and virtual screening using multiple 4-point pharmacophore fingerprints. *Pac. Symp. Biocomput.* **2000**, 576−587.
(20) Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C.; Labaudiniere, R. F. New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem.* **1999**, *42*, 3251−3264.
(21) Mason, J. S.; Cheney, D. L. Ligand−receptor 3-D similarity studies using multiple 4-point pharmacophores. *Pac. Symp. Biocomput.* **1999**, 456−467.
(22) Deng, Z.; Chuaqui, C.; Singh, J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein−ligand binding interactions. *J. Med. Chem.* **2004**, *47*, 337−344.
(23) McGregor, M. J.; Muskal, S. M. Pharmacophore Fingerprinting. 1. Application to QSAR and Focused Library Design. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 569−574.
(24) McGregor, M. J.; Muskal, S. M. Pharmacophore Fingerprinting. 2. Application to Primary Library Design. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 117−125.
(25) Abrahamian, E.; Fox, P. C.; Naerum, L.; Christensen, I. T.; Thogersen, H.; Clark, R. D. Efficient Generation, Storage, and Manipulation of Fully Flexible Pharmacophore Multiplets and Their Use in 3-D Similarity Searching. *J. Chem. Inf. Comput. Sci.* **2005**, *43*, 458−468.
(26) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J. Med. Chem.* **2002**, *45*, 4350−4358.
(27) Willett, P. Similarity-based approaches to virtual screening. *Biochem. Soc. Trans.* **2003**, *31*, 603−606.

(28) OMEGA is a software product from OpenEye URL: http://www.eyesopen.com/fred.html. 2005.
(29) Gaisteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537−547.
(30) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom Pairs as Molecular Features in Structure−Activity Studies: Definition and Application. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64−73.
(31) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic Physicochemical Paramerts 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.
(32) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graph. Model.* **2002**, *20*, 281−295.
(33) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100−5109.
(34) Sielecki, T. M.; Boylan, J. F.; Benfield, P. A.; Trainor, G. L. Cyclin-dependent kinase inhibitors: useful targets in cell cycle regulation. *J. Med. Chem.* **2000**, *43*, 1−18.
(35) Flower, R. J. The development of COX2 inhibitors. *Nat. Rev. Drug. Discovery* **2003**, *2*, 179−191.
(36) Jordan, V. C. Antiestrogens and selective estrogen receptor modulators as multifunctional medicines. 2. Clinical considerations and new agents. *J. Med. Chem.* **2003**, *46*, 1081−1111.
(37) Jordan, V. C. Antiestrogens and selective estrogen receptor modulators as multifunctional medicines. 1. Receptor interactions. *J. Med. Chem.* **2003**, *46*, 883−908.
(38) Surleraux, D. L.; de Kock, H. A.; Verschueren, W. G.; Pille, G. M.; Maes, L. J.; Peeters, A.; Vendeville, S.; De, M. S.; Azijn, H.; Pauwels, R.; de Bethune, M. P.; King, N. M.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Wigerinck, P. B. Design of HIV-1 protease inhibitors active on multidrug-resistant virus. *J. Med. Chem.* **2005**, *48*, 1965−1973.
(39) von, I. M.; Dyason, J. C.; Oliver, S. W.; White, H. F.; Wu, W. Y.; Kok, G. B.; Pegg, M. S. A study of the active site of influenza virus sialidase: an approach to the rational design of novel anti-influenza drugs. *J. Med. Chem.* **1996**, *39*, 388−391.
(40) Cirillo, P. F.; Pargellis, C.; Regan, J. The non-diaryl heterocycle classes of p38 MAP kinase inhibitors. *Curr. Top. Med. Chem.* **2002**, *2*, 1021−1035.
(41) Steinmetzer T.; Sturzebecher J. Progress in the development of synthetic thrombin inhibitors as new orally active anticoagulants. *Curr. Med. Chem.* **2004**, *11*, 2297−2321.
(42) Muegge, I.; Brittelli, D.; Heald, S. L. Simple selection criteria for druglike chemical matter. *J. Med. Chem.* **2001**, *44*, 1841−1846.
(43) Ionizer is a software product of Schroedinger, Inc. URL: www.schroedinger.com. 2005.
(44) Daylight Chemical Information Systems, Inc. Mission Viejo, CA URL www.daylight.com.
(45) Stahl, M.; Rarey, M. Detailed Analysis of Scoring Functions for Virtual Screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.
(46) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.
(47) Muegge, I.; Enyedy, I. Virtual screening of kinase targets. *Curr. Med. Chem.* **2004**, *11*, 691−705.
(48) Graves, A. P.; Brenk, R.; Shoichet, B. K. Decoys for Docking. *J. Med. Chem.* **2005**, *11*, 3714−372.

JM050468I