Journal of Molecular Graphics and Modelling 29 (2010) 157–170

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods

Jianxin Duan[b,*], Steven L. Dixon[a], Jeffrey F. Lowrie[a], Woody Sherman[a]

[a] Schrödinger Inc., 120 West 45th Street, 17th Floor, New York, NY 10036-4041, USA
[b] Schrödinger GmbH, Dynamostr. 13, 68161 Mannheim, Germany

ARTICLE INFO

Article history:
Received 18 February 2010
Received in revised form 14 May 2010
Accepted 18 May 2010
Available online 25 May 2010

Keywords:
2D fingerprint
Similarity searching
Enrichment
Scaffold hopping
Ligand-based virtual screening
Virtual screening

ABSTRACT

Virtual screening is a widely used strategy in modern drug discovery and 2D fingerprint similarity is an important tool that has been successfully applied to retrieve active compounds from large datasets. However, it is not always straightforward to select an appropriate fingerprint method and associated settings for a given problem. Here, we applied eight different fingerprint methods, as implemented in the new cheminformatics package Canvas, on a well-validated dataset covering five targets. The fingerprint methods include Linear, Dendritic, Radial, MACCS, MOLPRINT2D, Pairwise, Triplet, and Torsion. We find that most fingerprints have similar retrieval rates on average; however, each has special characteristics that distinguish its performance on different query molecules and ligand sets. For example, some fingerprints exhibit a significant ligand size dependency whereas others are more robust with respect to variations in the query or active compounds. In cases where little information is known about the active ligands, MOLPRINT2D fingerprints produce the highest average retrieval actives. When multiple queries are available, we find that a fingerprint averaged over all query molecules is generally superior to fingerprints derived from single queries. Finally, a complementarity metric is proposed to determine which fingerprint methods can be combined to improve screening results.

1. Introduction

Virtual screening [1–3] is an established component of modern drug discovery. Numerous studies [4–9] have compared and contrasted various 3D, 2D, and 1D [7] representations for retrieving active compounds from a database. While docking [10–14] is frequently the method of choice when crystallographic data are available, ligand-based techniques, such as pharmacophore matching [15–18], shape-based screening [19–22], and 2D fingerprint similarity [5,6,23,24] may in fact be the only viable approach when the relevant crystal structures are not available. Ligand-based methods are also the most practical choice when the number of compounds to screen is large and a fast turnaround is needed. In practice, many virtual screening methods [25–29] may be available to a modeler and decisions [30,31] about which methods to use will depend on the problem at hand.

2D fingerprints are widely used in drug discovery and have been shown to be more successful at retrieving active compounds than 3D shape or docking methods [9,32], despite a complete neglect of explicit information about the target–ligand interactions. 2D fingerprints offer a wide variety of possibilities in terms of the fragmentation strategy, atom/bond typing schemes, bit scaling rules, and similarity indices. For example, in the cheminformatics package Canvas [33], there are more than 25,000 possible combinations of these four variables. Earlier systematic investigations of reasonable combinations of these variables [34–39] have provided practical guidelines for 2D fingerprint queries; however, these studies have focused primarily on enrichment factors and other global measures of performance. Both the sizes of the datasets and the complexity of the interrelations among variables have precluded a detailed analysis of the nature of the hits. Yet this sort of investigation is vital to understanding how to obtain hits that exhibit specific molecular properties, chemical diversity, and other characteristics beyond activity itself.

In order to understand the relationship between fingerprints and retrieved compounds, we have systematically investigated the behavior of eight common fingerprinting methods by performing a series similarity screens on small, well-validated sets of ligands for five targets [27,29,40–42]. The primary aim is not to exhaustively investigate how the different variables quantitatively influence enrichments, but rather to understand important differences between the fingerprinting methods and associated

* Corresponding author. Tel.: +49 621 438 55 172, fax: +49 621 438 55 555.
E-mail addresses: jianxin.duan@schrodinger.com (J. Duan), steven.dixon@schrodinger.com (S.L. Dixon), jeff.lowrie@schrodinger.com (J.F. Lowrie), woody.sherman@schrodinger.com (W. Sherman).

1093-3263/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2010.05.008

Table 1
Fingerprint methods and descriptions.

Type	Description
Linear	Linear fragments + ring closures.
Dendritic	Linear and branched fragments.
Radial	Fragments that grow radially from each atom. Also known as extended connectivity fingerprints (ECFPs [48]).
Pairwise	Pairs of atoms [51], differentiated by type, and the distance separating them: $Type_i Type_j d_{ij}$.
Triplet	Triplets of atoms, differentiated by type, and the three distances separating them: $Type_i$-d_{ij}–$Type_j$-d_{jk}–$Type_k$-d_{ki}.
Torsion	Four consecutively bonded atoms [52], differentiated by type: $Type_i$-$Type_j$-$Type_k$-$Type_l$.
MOLPRINT2D	A radial-like fingerprint that encodes atom environments using lists of atom types located at different topological distances [53,54].
MACCS	SMARTS-based implementation of the MACCS structural keys [43].

parameters in order to facilitate rational choices for achieving a desired set of results.

2. Methods

2.1. Fingerprints

The eight types of fingerprints implemented in Canvas are shown in Table 1 [33]. All of the fingerprints except MACCS [43] are encoded by hashing each chemical pattern into an *N*-bit address space, where *N* is typically 32 or 64, and storing only the "on" bits. A 32-bit fingerprint is therefore represented by a list of integers that lie on the interval $[1,2^{32}]$, while a 64-bit fingerprint is a list of integers on the interval $[1,2^{64}]$. The integer values represent the positions of the on bits, thereby producing a sparse encoding scheme. This is in contrast with some other fingerprint implementations [44–46], where hashing is done into a much smaller address space, such as 2^{10}, and explicit on/off values are stored. The advantage of using a 32-bit or 64-bit address space is that the likelihood of two different chemical features setting the same bit, also known as a *collision*, is exceedingly small. For 32-bit linear fingerprints, a collision typically occurs only once for every few thousand structures fingerprinted, whereas collisions are effectively eliminated altogether with 64-bit fingerprints. By comparison, use of a 10-bit address space usually results in numerous collisions, even when fingerprinting only a single structure. In some scenarios one may choose to allow collisions in order to introduce fuzziness, although in all cases a collision results in loss of information. For the purpose of introducing fuzziness, we allow hashing to a reduced address space. In this study we used a 32-bit address space for all calculations. The eight fingerprints used in this work are shown in Table 1 and a description is provided below. A more detailed description of these methods and the general application of 2D fingerprints can be found in a number of excellent sources [5,6,23,24,34,46,47].

2.1.1. Linear

All linear paths containing up to seven bonds are mapped to a structure, and a hashing operation is performed on a string-based description of each linear fragment to produce an integer bit address. To improve the description of rings without causing a massive proliferation in the number of fragments, the maximum path is expanded from seven to fourteen bonds for ring closure only. This effectively encodes information in and around most ring systems, while producing only a fraction of the fragments that one would get if all paths containing 0–14 bonds were enumerated.

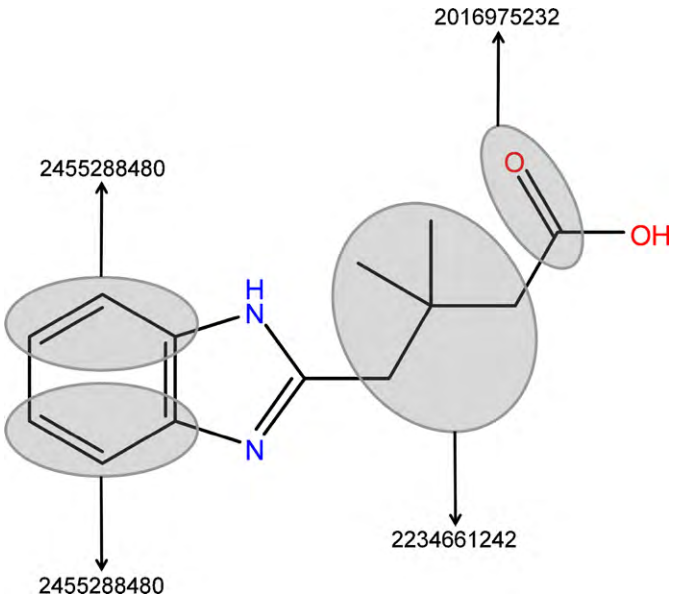

Fig. 1. Example fragments and hash codes for a 32-bit Dendritic fingerprint with Daylight invariant atom types.

2.1.2. Dendritic

To encode both linear and branched features, linear paths are augmented with intersections of linear paths, with a maximum of five bonds per path (Fig. 1). Dendritic fingerprints incorporate no special treatment of rings.

2.1.3. Radial

Also known as extended connectivity fingerprints (ECFPs) [48,49]. Radial fingerprints entail growing a set of fragments radially from each heavy atom over a series of iterations. Using a variant of the Morgan algorithm [50], each chemically unique fragment is mapped to a distinct integer value by hashing a description of the atoms and bonds within the fragment, and the bonds that connect it to the surrounding region

2.1.4. Pairwise

These fingerprints are based on the concept of an atom pair [51], which is simply two atom types and the topological distance separating them: $Type_i Type_j d_{ij}$. This representation is hashed, byte-by-byte, to an integer value. By default, all pairs of atoms in the molecule are considered, and distances are the number of bonds in the shortest path between each pair.

2.1.5. Triplet

An extension of the atom pair, a triplet consists of a set of three atoms and the topological distances separating them: $Type_i$-d_{ij}–$Type_j$-d_{jk}–$Type_k$-d_{ki}. Since there are six different ways to order the atoms in a triplet, a canonicalization is performed to ensure that there are no bits in the fingerprint that differ only in a permutation of the atoms.

2.1.6. Torsion

Also known as Topological Torsions [52], these are a special case of linear fingerprints, wherein every fragment consists of a linear path of four atoms that are differentiated by type: $Type_i$–$Type_j$–$Type_k$–$Type_l$. Because only fragments containing four atoms are enumerated, the number of on bits in a torsion fingerprint tends to be about an order of magnitude smaller than the number of on bits in a linear fingerprint.

Table 2
Atom-typing schemes. The shortened code is used throughout this work.

Code	Description
G	All atoms and bonds are equivalent (generic).
HB	Atoms are distinguished by whether they are hydrogen bond (HB) acceptors or donors; all bonds are equivalent.
Hybr	Atoms are distinguished by hybridization state; all bonds are equivalent.
Fn	Atoms are distinguished by functional type: {H}, {C}, {F,Cl}, {Br,I}, {N,O}, {S}, {other}. All bonds are equivalent.
Mol2	Sybyl Mol2 atom types; [69] all bonds are equivalent.
TXHB	Atoms are distinguished by whether they are terminal, halogen, HB acceptor/donor; all bonds are equivalent.
Elem	Atomic number and bond order.
RTXHB	Atoms are distinguished by ring size, aromaticity, whether terminal, whether halogen, HB acceptor/donor; bonds are distinguished by bond order.
Car	Carhart atom types (atom-pairs approach [51]); all bonds are equivalent.
Day	Daylight invariant atom types [56]; bonds are distinguished by bond order.
E	Estate [70] atom types. The EState coordinate is divided into bins of width 0.25, and each atom is typed according to the bin in which its EState value. falls. All bonds are equivalent.

2.1.7. MOLPRINT2D

Each heavy atom in a structure is characterized by an environment that consists of all other heavy atoms within a distance of two bonds [53,54]. A bit in the fingerprint is derived from a tabular data structure that stores the central heavy atom type, and lists of atom types found at distances of one and two bonds. Each member of the list is encoded into a string of the form *Type-freq*(*Type*)-*d*, where *freq*(*Type*) is the number of times a given atom type is found at a distance *d* from the central atom. These terms are sorted by distance, then by type, and finally hashed to an integer that sets a bit in the fingerprint

2.1.8. MACCS

These well-known structural keys [43] are encoded using a set of 155 SMARTS [55] patterns, which is 11 shy of the 166 public keys defined in the MDL documentation. The missing keys correspond to features that are not easily encoded using SMARTS, such as "isotope," or definitions that are ambiguous, for example, "(&.)," which is defined only as, "and other rare features."

2.2. Atom types

Hashed fingerprinting methods are sensitive to the atom-typing scheme used, because this determines which atoms are treated as chemically distinct, and therefore how many unique chemical features a structure possesses. Table 2 contains descriptions and shorthand codes for the 11 atom-typing schemes implemented. They range from purely generic (G), where each structure is essentially transformed to a single-bonded carbon skeleton, to highly specific treatments, such as Daylight invariant atom types (Day), which differentiate by atomic number, formal charge, valence, and the numbers of hydrogen and non-hydrogen connections [56]. In this study, Daylight invariant atom types (Day) are the default for all hashed fingerprints except MOLPRINT2D and Pairwise, for which the defaults are Sybyl Mol2 atom types (Mol2) and Carthart atom types (Car), respectively.

2.3. Similarity index

Canvas provides 24 different indices to measure the similarity or distance between a pair structures *A* and *B*. Here we investigated only the Tanimoto index, which is one of the most commonly used in the literature. The Tanimoto index is defined as $c/(a+b-c)$, where a is the number of bits set by structure *A*, b is the number of bits set by structure *B*, and c is the number of bits set jointly by *A* and *B*. Exploring the effect of different similarity indices is beyond the scope of this work and is the focus of a separate paper where we explore the effect on enrichment of varying the similarity index along with a number of other parameters [39].

Table 3
Number of compounds used for this study and earlier publications. Differences in numbers are a result of different versions of the MDL Drug Data Report (MDDR) at the time of each study.

Class	Rarey and Dixon	Briem and Lessel	This work
5HT	52	49	48
ACE	40	40	39
HMG	114	111	110
PAF	136	134	134
TXA	49	49	49
Random	579	574	572

2.4. Dataset

Briem and Lessel compiled a dataset from the MDL Drug Data Report (MDDR) [57] to measure the success of retrieving active compounds from similarity screens [41]. This dataset was chosen for the work presented here because it has been well-validated in previous publications and has been used to assess other methods, which facilitates head-to-head comparisons. The dataset contains active compounds for five different target classes: angiotensin-converting enzyme (ACE), thromboxane A2 receptor (TXA), 5-HT_3 receptor (5HT), HMG CoA reductase (HMG), and platelet activating factor receptor (PAF). Additionally, random compounds that do not belong to any of the above mentioned classes were also included. The MDDR reference codes of the dataset are available at http://cheminformatics.org/datasets. Using the MDDR [57] available in 2008, we were able to retrieve most of the compounds from the original dataset. One compound each from the 5HT, ACE and HMG activity classes was missing as well as two compounds from the random selection. This is due to updates in the MDDR since the time of the original publication in 2000. The final collection includes 48 5HT actives, 39 ACE actives, 110 HMG actives, 134 PAF actives, 49 TXA actives, and 572 random decoys, resulting in a total of 952 compounds (Table 3). The number of decoys is not particularly large compared to the number of actives in each class, which could potentially cause problems because the retrieval rate can be unrealistically high for all methods. However, this dataset has been used in a number of 2D and 3D similarity studies [27,29,40–42] and is therefore an excellent benchmark for validating the performance of Canvas fingerprints. With only a handful of missing compounds, the results should be comparable to other publications based on the same dataset.

2.5. Calculation details

For each target, a separate similarity screen was performed using each active compound as a query. Similarities were computed from seven different hashed fingerprinting methods (Dendritic, Linear, MOLPRINT2D, Pairwise, Radial, Torsion, and Triplets) as well as MACCS structural keys. Fingerprints were generated using default atom-typing schemes (Table 4) and no bit scaling. Hashed fingerprints were mapped into a 32-bit address space, which when combined with a highly specific scheme for encoding chemical fragments, results in a very low probability of collisions.

Retrieval rate is defined as the number of actives found within the first 1% of the screen, where compounds are scored in order of decreasing similarity to the query. The average retrieval rate for

Table 4
Default atom-typing schemes associated with each fingerprint. See Table 2 for atom-typing description.

Fingerprint	Atom-typing scheme
Linear	Daylight
Dendritic	Daylight
Radial	Daylight
MOLPRINT2D	Mol2
Pairwise	Carhart
Triplet	Daylight
Torsion	Daylight
MACCS	<N/A>

each target was determined by averaging the individual retrievals from the different queries for that target.

In order to quantify the chemical orthogonality of the various fingerprinting methods, we defined *complementarity* as one minus the correlation coefficient between the Tanimoto similarity scores for a pair of fingerprint methods across the full set of compounds. This is a useful concept, as combining fingerprints with high complementarity by applying a logical OR to the bit sets may lead to enhanced retrieval of active compounds. Although retrieving an orthogonal hit list is desirable when combining screening results from different fingerprints, it is also important to obtain enrichment in actives over decoys. To account for this, the complementarity is multiplied by the average retrieval rates of both fingerprints to obtain a final metric for assessing the performance of combined fingerprints.

Fingerprints for multiple queries can be compressed to a single fingerprint by averaging each bit and normalizing by the frequency. Such a procedure leads to a modal fingerprint that contains information from each of the query molecules. Additionally, it speeds up a virtual screening because one single screening is sufficient for several query molecules. In this study, we randomly chose sets of 3, 6 or 9 queries and used the modal fingerprint to screen the database. The retrieval rates were then compared to the average retrieval rate from performing separate screens with each of the randomly chosen queries in the set. We performed the exercise 50 times for each of the targets and fingerprint methods and the statistical significance is calculated based on paired Student's *t*-test.

We have also included the published performance of other popular ligand-based methods in order to provide a frame of reference for comparison. The first is a graph-based method called Feature Trees (FTrees) [29], which describes a molecule as a set of connected nodes, with each node originating from a hydrophobic fragment or functional group in the structure. Comparisons are made by matching parts of trees from different molecules. Two more traditional fingerprint methods, Tripos Hologram fingerprints [41] and Daylight fingerprints [29], are also included. The final method compared is called Lingo [27], which uses SMILES strings as the basis for comparing structures. These methods have proven to be successful in previous virtual screening studies [58–61] and have been applied to the Briem and Lessel dataset, which makes them good references for comparison.

3. Results and discussion

3.1. The dataset

Fig. 2 shows the active and decoy ligands characterized by a number of 1D descriptors, such as molecular weight, number of rotatable bonds, and counts of chemical features (donors, accep-

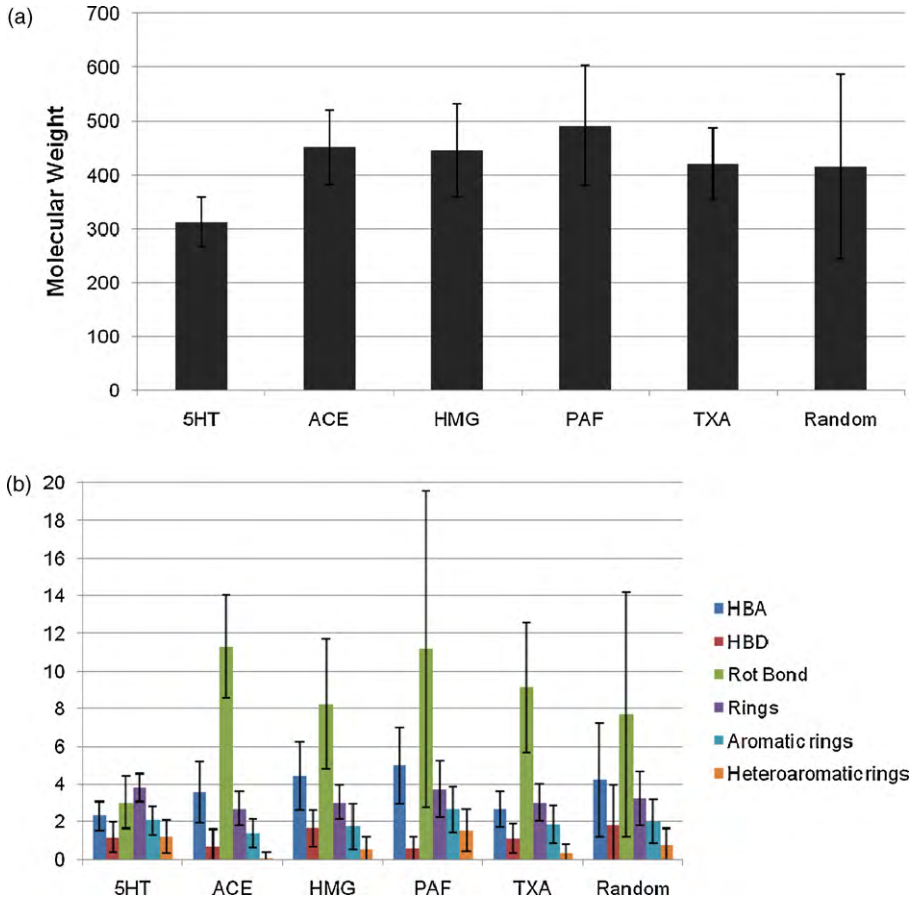

Fig. 2. (A) Average molecular weight for each activity class and the random selection (i.e. inactive compounds). (B) Average molecular descriptor values for each class.

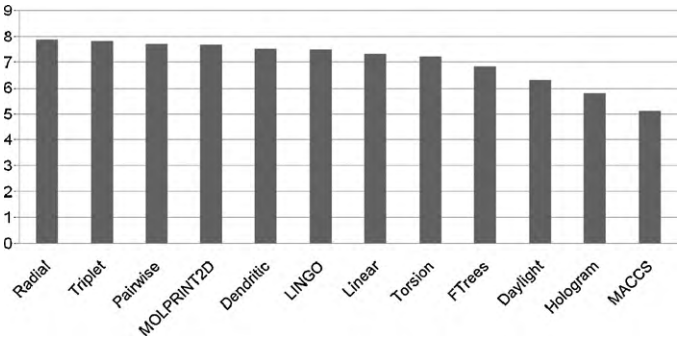

Fig. 3. Average retrieval rates across all targets, sorted by value. The data for LINGO, FTrees, Daylight fingerprints, and Tripos Holograms were taken from published benchmarks [27,29,41].

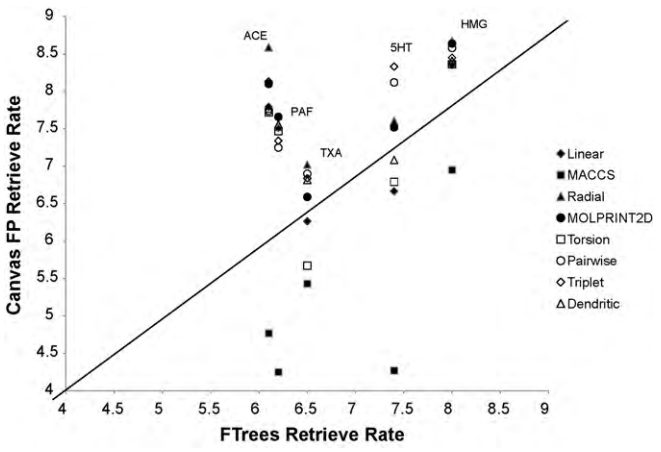

Fig. 4. Comparison of retrieval rates for Canvas fingerprints and FTrees. Points below the line correspond to cases where FTrees perform better.

tors, aromatic rings, etc.). 5HT actives have lower molecular weight and fewer rotatable bonds on average, but in other aspects do not significantly differ from the other actives. ACE and PAF actives have the largest numbers of rotatable bonds, and in the case of PAF, this property varies dramatically. A closer look at the PAF ligands reveals that there are essentially two families: extended structures with fatty acid-like aliphatic chains and structures with multiple rings connected by a few rotatable bonds. Interestingly, ACE inhibitors contain almost no heteroaromatic rings.

3.2. Retrieval rate

Before analyzing the hits from each query in detail, the general performance of Canvas fingerprints are compared to previous benchmarks obtained using the same dataset. Fig. 3 shows the retrieval rates for each Canvas fingerprinting method averaged over all targets along with earlier published results from several other methods (LINGO [27], Tripos Hologram fingerprints [41], FTrees [29], and Daylight fingerprints [29]). Overall, five Canvas fingerprints (Radial, Triplet, Pairwise, MOLPRINT2D, and Dendritic) performed equally well or better than all other methods. The good performance of Radial and MOLPTINR2D fingerprints agrees with earlier study of similar fingerprints, ECFP [36] and SEFP4 [54]. Of the eight Canvas fingerprints, seven were able to retrieve on average between 7.2 and 7.9 active compounds of the same class within the top 10, while MACCS keys are able to retrieve only 5.1 actives (Fig. 3). Since MACCS keys code only the presence or absence of functional groups and substructures without any connectivity information between the features, they are apparently not sufficiently discriminating to achieve high enrichment. Furthermore, far less information is encoded in the 155 SMARTS-based structural keys than is possible in a 32-bit fingerprint.

In addition to computing the retrieval rate, we also calculated the area under the receiver operating characteristic curve (ROC). The correlation between ROC and retrieval rate is very high ($R^2 = 0.97$), suggesting that using either one of these metrics is sufficient for comparing different methods. However, there is a change in the relative performance of the fingerprints, with MOLPRINT2D being the best performing fingerprint with an ROC score of 0.82 (ranked forth based on retrieval rates) and Triplet ranking fifth with an ROC score of 0.80 (ranked second based on retrieval rates). The enrichment of MACCS keys ranked last with both metrics (ROC score of 0.67). The relative ordering of the other fingerprint methods was unchanged. In the remaining part of this study we will use retrieval rate as the measure of enrichment.

Although the average retrieval rates in Fig. 3 provide a good overview of performance, it is useful to analyze how the different methods perform with respect to individual targets. In Fig. 4, average retrieval rates for each target and fingerprint are plotted against the same data for FTrees [29]. MACCS keys exhibit the lowest retrieval rates and underperform the hashed fingerprints in every case. All hashed fingerprint methods performed very similarly in retrieving HMG and PAF ligands, but a more significant spread in retrieval rates was observed for 5HT and TXA ligands. ACE inhibitor retrieval rates obtained using hashed fingerprints were spread over a smaller range than those of 5HT and TXA ligands.

TXA ligands proved to be particularly difficult to retrieve using FTrees and the various hashed fingerprinting methods, whereas inhibitors of HMG and ACE were among the easiest to identify. All Canvas hashed fingerprints performed consistently better than FTrees in retrieving ACE inhibitors and PAF ligands. Radial fingerprints were the most successful for all targets except 5HT, where Pairwise and Triplet fingerprints excelled. The superior performance of Pairwise and Triplet fingerprints may be related to the smaller size of the 5HT ligands, as the encoding of information from all pairs or triplets of distances makes these fingerprints particularly sensitive to molecular size. The overall success of Radial fingerprints suggests that discriminating power is increased by the highly restrictive requirement to match not just a fragment, but also the manner in which that fragment connects to its surroundings. However, whereas Radial fingerprints are highly effective for finding very similar compounds, they may not necessarily be optimal for early stage lead discovery where diverse hits are often sought.

3.3. Complementarity of fingerprints

Next, we examined the top 10 hits from pairs of fingerprints to ascertain their complementarity. If two fingerprints yield very different hits in the early part of the database it follows that a combination of the hit lists may increase the rate at which actives are found. However, high complementarity (i.e. different hit lists) does not necessarily imply high enrichment because it is possible that one or both of the methods does not do a good job at retrieving active compounds. To account for this, we devised an overall metric that is the product of the complementarity (one minus correlation coefficient between the Tanimoto similarity scores of the two screening lists) and the average retrieval rate from both fingerprints.

Table 5 summarizes this analysis for all pairs of Canvas hashed fingerprints. No highlighting indicates a high overlap between the fingerprints and/or a low enrichment, whereas darker gray signifies that the two fingerprints are simultaneously complementary

Table 5
Heat map of complementarity (Pearson correlation of all similarity scores) multiplied by retrieval rates of both fingerprints. Darker gray signifies that the two fingerprints are simultaneously complementary and have good enrichment. White indicates a high overlap (i.e. non-complementarity) between the fingerprints and/or low enrichments.

	Dendritic	Linear	MOLPRINT2D	Pairwise	Radial	Torsion	Triplet
Dendritic	0.00						
Linear	2.04	0.00					
MOLPRINT2D	7.13	8.57	0.00				
Pairwise	19.35	22.82	20.81	0.00			
Radial	6.94	4.27	12.38	21.35	0.00		
Torsion	1.52	3.11	7.77	21.01	9.48	0.00	
Triplet	6.89	6.98	9.75	11.16	8.75	8.00	0.00

and have good enrichment. It should be readily apparent that the heat map is consistent with commonalities in the underlying fingerprinting methodologies. For example, the three fragment-based fingerprints (Linear, Dendritic and Torsion) exhibit a high overlap, as do the distance-encoding Pairwise and Triplet fingerprints. It is also seen that Pairwise and Triplet fingerprints retrieve compounds that are highly distinct from all other methods, particularly the fragment-based ones. MACCS keys also retrieve distinct hits, but since they exhibit low enrichments, they are not a desirable combination fingerprint partner. Based on the data in Table 5, a good combination is Pairwise or Triplet fingerprints teamed with any of the three fragment-based fingerprints.

To test this hypothesis, the Pairwise fingerprint was combined using bitset "OR" logic with all other fingerprints except MACCS, and average retrieval rates were computed and averaged over queries for each class. We included the Pairwise/Triplet combination as a negative control because we expect no significant change in retrieval rate. In Table 6, we show the difference between average retrieval rates achieved by the combined fingerprints and each component fingerprint. The improvement in retrieval rate was in many cases only marginal, and in 6 out of 30 cases, the combined fingerprint performed worse than one of the component fingerprints. It is worth noting that in no case did combined fingerprint perform worse than both component fingerprints.

In order to understand the averaged retrieval rate, we analyzed the data for 5HT class in detail, where the degradation of retrieval rate is the largest. We found no significant change ($P < 0.05$) in retrieval rate between Pairwise fingerprints and the combined fingerprints, except in the case of the Pairwise/Linear combination, which yielded a retrieval rate that was 0.93 lower than Pairwise fingerprints alone. However, combined fingerprints showed significant ($P < 0.01$) improvements in retrieval rates over the non-Pairwise partner in all cases except Triplet, where performance was nearly the same whether used by itself or in combination with Pairwise fingerprints. The most exceptional case was the Pairwise/Torsion combination, which retrieved additional hits compared to Torsion alone in 27 of 48 queries, with improvement margins of five or more actives in four cases. Generally speaking, the combined fingerprint is almost always better than the underperforming member of the pair, and often nearly as good as the over performing member. When the choice of the best fingerprint is not possible, a combination of orthogonal fingerprint such as Pairwise/Torsion or Pairwise/Radial is a prudent alternative.

3.4. The generic fingerprint

The general conclusion from our analysis of eight different 2D fingerprints is that there does not appear to be a method that consistently outperforms the others. A similar conclusion was reached by Sheridan and Kearsley [62]. Ideally, the choice of fingerprint should be made on a case-by-case basis, after thorough validation for each target. Although each of the fingerprints has similar average performance, it is useful to identify a method that has the most consistent performance across a range of queries and targets. With a given fingerprinting method, it is possible that some of the queries will retrieve many actives, whereas others will retrieve only a few, which indicates significant sensitivity to the query. In general, it is more desirable for all queries to retrieve a satisfactory number of actives, with relatively minor variation in that number from query to query. Furthermore, a fingerprinting method should perform well across a wide range of targets. A method that satisfies such criteria would be ideal as a *generic* way to scan the chemical space in an early stage of discovery.

In order to investigate this hypothesis, we counted for each fingerprint the percentage of queries that produced one or zero actives after 1% screened (i.e. $\leq$ the random success rate), and correlated that number with the average retrieval rate (Fig. 5). MACCS keys were excluded because the average retrieval rate is already quite low compared to the other methods. All of the fingerprints performed very well in this regard, with over 95% of the 380 queries producing better than random enrichment. MOLPRINT2D in particular stands out, with more than 98% of its queries outperforming random screening, followed closely by Radial, at 97%. Furthermore, the average retrieval rate of MOLPRINT2D is only slightly lower than that of Radial, and ranks third overall in retrieval rate. Therefore, in the absence of prior knowledge about the effectiveness of each fingerprint for the problem at hand, MOLPRINT2D appears to be a sound choice. Also, MOLPRINT2D has the lowest variation in retrieval (variance across all targets = 5.06), followed by Radial and Triplet, with retrieval variance of 5.82 and 6.31, respectively. The largest variation in retrieval came from Torsion (7.44).

Table 6
Difference in retrieval rate comparing combined fingerprint to single component fingerprint. Each pair of columns contains the difference between the combined fingerprint specified in the top row and the individual fingerprints specified in the second row. Positive values mean the combined fingerprint performs better than the individual fingerprint.

Target	Torsion/pairwise		Linear/pairwise		Dendritic/pairwise		MOLPRINT2D/pairwise		Radial/pairwise		Triplet/pairwise	
	Torsion	Pairwise	Linear	Pairwise	Dendritic	Pairwise	MOLPRINT2D	Pairwise	Radial	Pairwise	Triplet	Pairwise
5HT	1.40	0.07	0.52	−0.93	0.67	−0.37	0.83	0.23	0.67	0.15	0.05	0.26
ACE	0.56	0.54	0.42	0.47	0.54	0.57	−0.05	0.31	−0.23	0.62	0.05	0.44
HMG	0.36	0.14	0.29	0.06	0.34	0.17	0.02	0.08	0.06	0.15	0.02	−0.11
PAF	0.22	0.44	0.31	0.57	0.19	0.50	−0.10	0.31	0.05	0.32	0.03	0.12
TXA	1.60	0.37	0.79	0.16	0.53	0.45	0.55	0.24	0.10	0.22	0.06	0.00
Average	0.83	0.31	0.46	0.07	0.45	0.26	0.25	0.23	0.13	0.29	0.04	0.14

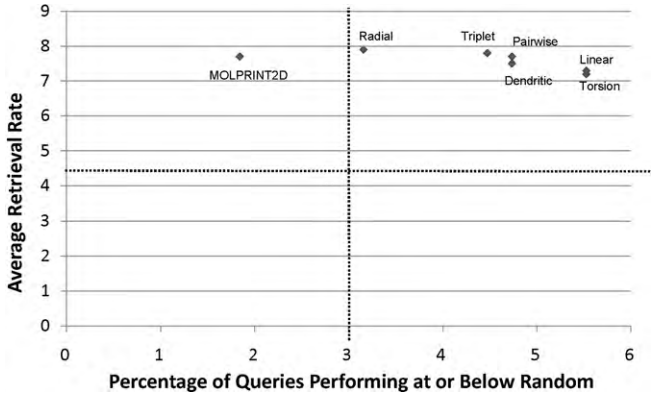

Fig. 5. Relationship between average retrieval rate and percentage of queries that produced fewer than two actives (random selection).

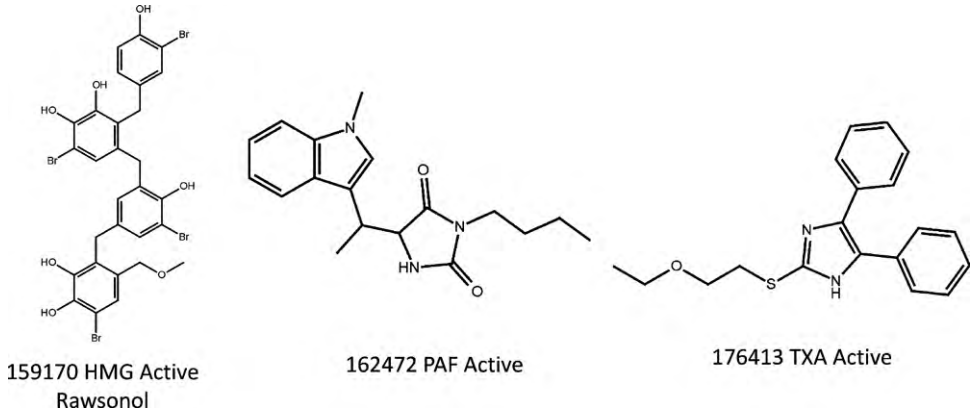

Fig. 6. Example query molecules where none of the fingerprints produced any enrichment better than random.

3.5. Understanding the failures

It is important to understand why certain fingerprints failed to return active compounds in the top 1% for some queries. Fig. 6 contains the structures of three actives that, when used as queries, yielded enrichments no better than random for all Canvas fingerprinting methods. In addition, the HMG inhibitor Tocotrienol (MDDR ID 171633) produced better than random results only with MACCS keys. With the exception of the non-statin HMG inhibitor 159170 (discussed further below), a cursory examination of the structures reveals no obvious reasons why these molecules were particularly ineffective as queries. However, the similarities tell a different story. The average Linear fingerprint similarity of TXA compound 176413 to other TXA actives is 0.024, whereas the average similarity between all pairs of TXA actives is 0.074. According to Student's paired *t*-test, all except one TXA actives have significantly higher similarities ($P < 0.05$) to other TXA actives. An analogous result was found for the PAF inhibitor 162472, where the similarities are 0.027 and 0.051 and 122 of 134 PAF inhibitors have significantly higher similarities ($P < 0.05$) to other PAF actives. These two actives are indeed less similar to the other actives in their respective classes. While the Tanimoto similarity values appear to be relatively low compared to values reported in the literature, this is due to the fingerprints being hashed into 32-bit address space by default. By decreasing to 10-bit address space (1024 total bits), the average Linear fingerprint similarity between all TXA actives is 0.254.

To better understand the differences in screening performance, we studied cases where certain fingerprints produced superior enrichments while others did not. Fig. 7 shows two HMG inhibitors that were especially challenging for Pairwise and Triplet fingerprints. When compound 153711 was used as a query, neither

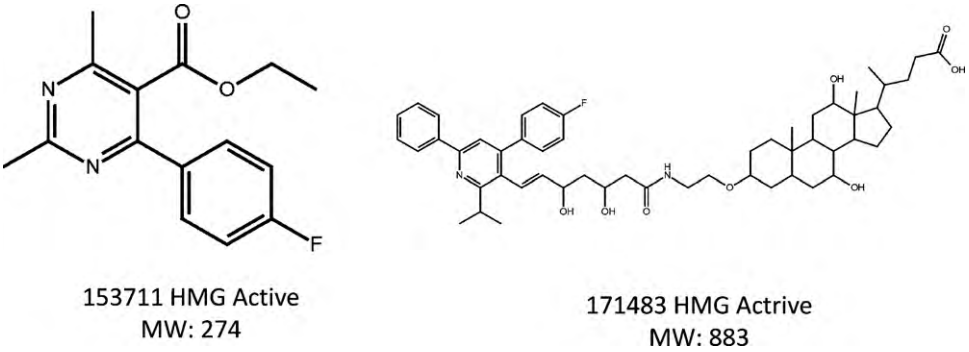

Fig. 7. Example query molecules where Pairwise and Triplet fingerprints performed poorly.

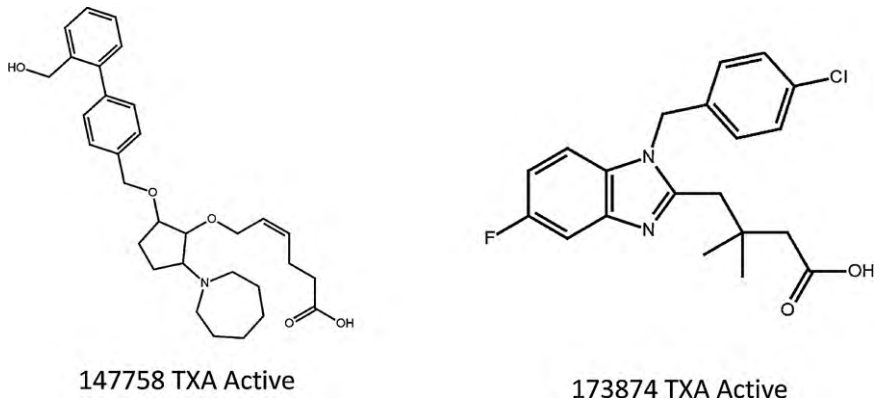

Fig. 8. Example query molecules where path-dependent fingerprints performed poorly. Other TXA active compounds are shown in Fig. 9.

method identified any hits in the first 1%. When screening with 171483 as a query, Triplet fingerprints retrieved only three actives. Notably, the molecular weights of 153711 and 171483 are 274 and 883, respectively, which differ significantly on the two sides of the extreme from the average HMG molecular weight of 445. This suggests that Pairwise and Triplet fingerprints may be especially sensitive to query molecule size, which is consistent with their superior ability to identify the smaller 5HT ligands (Fig. 4). Unlike fragment-based methods, Pairwise and Triplet fingerprints describe the molecule globally by encoding all possible pairs and triplets of atoms, so the number of features increases with structure size in a quadratic and cubic fashion, respectively. Thus if one structure is twice as large as another, the Pairwise and Triplet similarities are bounded above by roughly 0.25 and 0.125, respectively, regardless of how much local similarity the two structures possess. These observations suggest that Pairwise and Triplet fingerprints may be comparatively less successful in a primary screen that involves actives of widely varying size.

Two examples where Linear, Dendritic, and Torsion fingerprints failed to produce any enrichment above random are shown in Fig. 8. Although it is not obvious why all three methods failed on the TXA ligand 173874, it is important to note that Pairwise and Triplet fingerprints also performed poorly with this query, which may again point to an issue of molecular size. Molecule 173874 also contains a fused ring system as core, and although there are several TXA ligands with a similar core, they also contain a sulfonamide one to three bonds from the core (Fig. 9). The absence of such a feature in 173874 could be the explanation for the failure, and by reducing the information in the atom types, it should be possible to pick up actives. We scanned all atom-typing schemes for Linear, Dendritic, and Torsion fingerprints and found that when using RTXHB atom typing (atoms are distinguished by ring size, aromaticity, whether terminal, whether halogen, hydrogen bond acceptor/donor; bonds

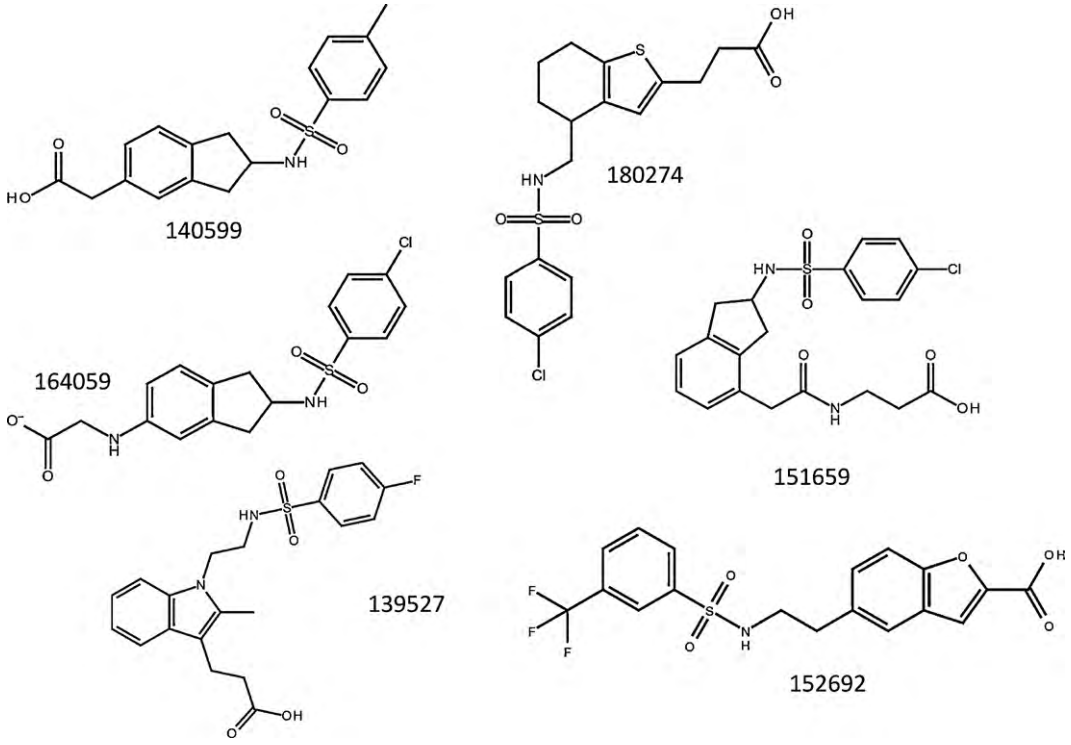

Fig. 9. Examples of TXA ligands with a sulfonamide group close to a fused 5,6-membered ring core similar to the TXA ligand 173874 with a benzimidazole core shown in Fig. 8.

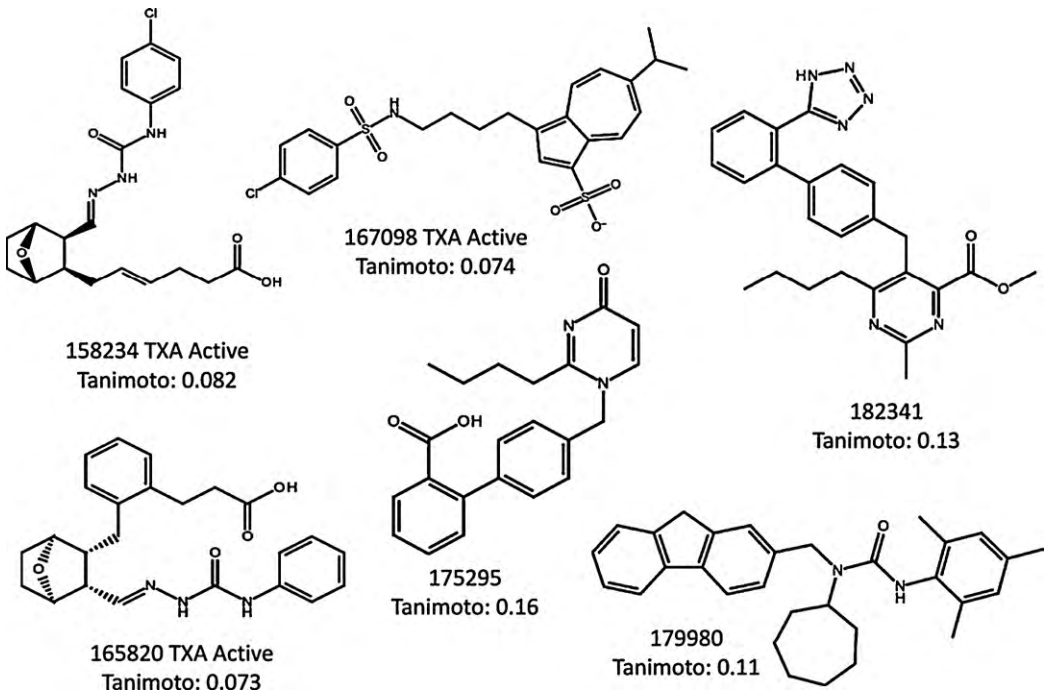

Fig. 10. The three most similar TXA ligands and three most similar decoy molecules compared to TXA active 147758 (shown in Fig. 8) with associated Tanimoto coefficients. Linear fingerprints were used in this comparison.

are distinguished by bond order), one of the TXA actives in Fig. 9 (ligand 152692) was scored highly by all three methods. Additional TXA actives with different scaffolds were picked if the HB atom-typing scheme (atoms are distinguished by whether they are hydrogen bond (HB) acceptors or donors; all bonds are equivalent) was applied. It is also worth noting that Radial fingerprints, which incorporate information up to four bonds from each atomic center, identified only three actives in the top 1%, whereas MOLPRINT2D fingerprints, with a neighborhood of only two bonds, retrieved seven actives. A highly localized description appears to be beneficial for finding actives when using 173874 as a query.

By contrast, molecule 147758 exhibits more obvious structural differences from other actives in its class. The extended chain of closely spaced rings differs from the more compact structures of most of the remaining TXA ligands. This elongated form, combined with a regular occurrence of rings, probably gives rise to a rather unique pattern of bits in fragment-based fingerprints. This hypothesis is supported by the three most similar molecules to 147758 in the entire dataset based on Linear fingerprint, which all have recurring rings (Fig. 10).

3.6. Modal fingerprints

Frequently, multiple known active compounds are available as queries for similarity screening. As seen in the previous section, the choice of the query molecule can have a significant impact on the retrieval of actives. One solution is to perform a separate screen for each of the query molecules; however, the strategy for the final compound selection from the combined hit list is not obvious. Although various consensus scoring schemes [63,64] can be helpful in such selections, the choice of the component scoring functions and appropriate schemes is not straightforward and requires a sufficient number of known actives to validate [65,66]. An alternative that we take here is to condense the information from multiple queries into one single fingerprint by averaging the bits. We call this a "modal fingerprint", which is different from methods where either a bit is set only if it is found in more than a user-defined threshold percentage [67] or a bit is weighted by the number of actives with that bit set [68]. The first method requires user-defined parameter and the second method depends on exact number of actives that have that bit set.

By choosing a random subset of query molecules many times, we can compare the modal fingerprint retrieval with the average retrieval rate of the individual queries. In this case, we made 50 independent random selections of 3, 6, or 9 query molecules. As shown in Fig. 11, screens with modal fingerprints in general retrieved more actives than the individual screens. Additionally, the optimal performance occurs when the modal fingerprint is averaged over 6 known active molecules. One exception is Pairwise fingerprints, where retrieval actually decreased when averaging was done for 6 or 9 query molecules. A detailed look in Table 7 summarizes the difference in average retrieval rates as a function of fingerprinting method and target. Since we are looking for improvement in retrieval rates, we required that the modal fingerprint retrieve at least one active within top 10 hits. An earlier study compared the variants of modal fingerprints described by Shemetulskis et al. [67] and Hert et al. [68] as well as consensus scoring schemes and found that Shemetulskis' implementation increased the performance by 30% in average compared to single query similarity screens [68]. The relative improvements observed in the modal fingerprints described here are less, although that likely has to do with the dataset and the fact that that the individual screens already produces retrieval rates of close to 8, which leave little room to improve.

In most cases, screening using a modal fingerprint significantly ($P < 0.05$) improves retrieval rates. However, it is striking that Pairwise modal fingerprints perform poorly for all targets except PAF. This is especially surprising because Pairwise fingerprints performed particularly well for 5HT when using a single query. The poor retrieval using Pairwise modal fingerprints suggests that the 5HT size bias was not enhanced by averaging but rather smeared out and obscured. In contrast, both Pairwise and Triplet modal fingerprints improved the retrieval of PAF actives by two or more hits, which is better than other fingerprints. Unlike the 5HT actives,

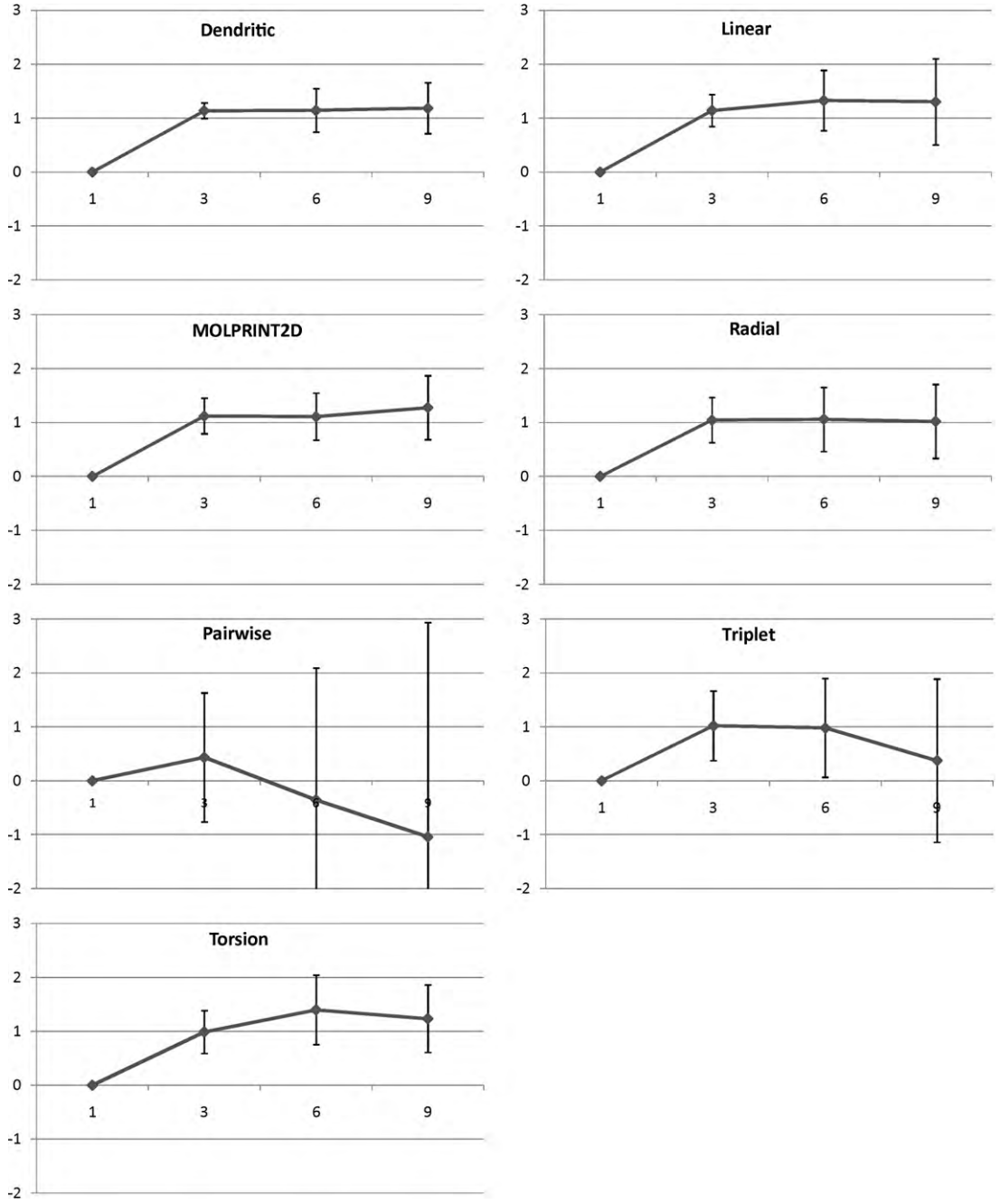

Fig. 11. The changes in average retrieval rate (*Y*-axis) for screening the modal fingerprints compared to screening with individual query molecules. The modal fingerprints were generated by averaging over fingerprints for 3, 6 or 9 randomly selected query molecules (*X*-axis) and repeating 50 times to get error estimates.

Table 7
Difference between average retrieval rate for modal fingerprint and average retrieval rate from individual screens. The numbers in light gray are not statistically significant and the numbers in dark gray are statistically significant with *P* values between 0.01 and 0.05. All others numbers are statistically significant with *P* values less than 0.01. NA indicates screening using the modal fingerprint failed to retrieve actives.

Target	5HT			ACE			HMG			PAF			TXA		
# of queries	3	6	9	3	6	9	3	6	9	3	6	9	3	6	9
Dendritic	1.26	1.51	1.72	1.05	1.20	0.91	0.93	0.65	0.83	1.27	1.54	1.68	1.17	0.81	0.78
Linear	1.19	2.09	1.97	1.13	0.81	0.35	0.68	0.75	0.55	1.19	1.58	1.60	1.51	1.40	2.04
MOLPRINT2D	0.94	0.84	1.04	0.99	0.85	0.92	0.75	0.69	0.63	1.36	1.59	2.04	1.56	1.57	1.74
Pairwise	−0.83	NA	NA	0.63	−0.52	−3.85	0.67	NA	NA	2.21	2.17	1.77	−0.51	−2.71	NA
Radial	0.85	0.58	0.71	0.77	0.83	0.80	0.62	0.54	0.36	1.54	1.91	2.16	1.45	1.43	1.07
Torsion	1.35	1.85	1.87	0.82	0.85	0.58	0.37	0.61	0.54	1.24	1.56	1.53	1.15	2.11	1.64
Triplet	0.27	0.29	−0.05	0.89	0.30	−1.50	0.87	0.72	−0.21	2.05	2.51	2.51	1.02	1.08	1.12

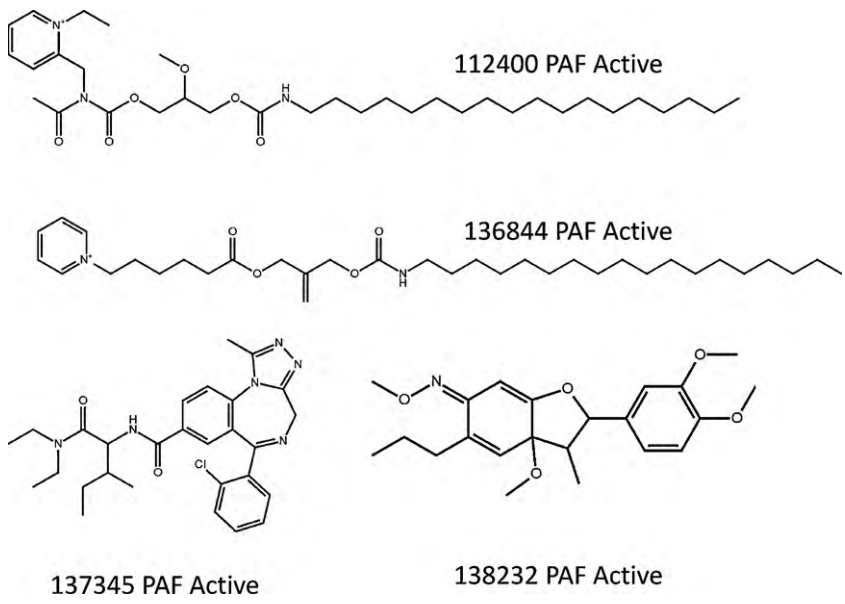

Fig. 12. Examples showing the diversity of PAF actives.

PAF actives compounds consist of a both long fatty acid-like compounds with a large variation in rotatable bonds, and more compact drug-like molecules (Fig. 12). The results suggest that Pairwise and Triplet modal fingerprints benefit from size diverse query molecules. Unlike the other fingerprint methods, where structures are encoded by a series of fragments, Pairwise fingerprints describe the topological shape and size of the molecule by enumerating all pairs of atoms and the shortest path separating them. Although Triplet fingerprints result from a straightforward extension of Pairwise fingerprints, the difference in performance between their modal fingerprints seems to indicate that the greater information content of Triplet fingerprints may be important for averaging. Our findings suggest that if multiple known active molecules are available, it is advisable to use the modal fingerprint. However, Pairwise fingerprints should be avoided for averaging unless the known actives vary significantly in size.

3.7. Dependency on atom typing

The different fingerprint implementations incorporate different levels of fuzziness in the description of molecular detail. For example, Radial fingerprints are the most restrictive, with requirements on the atom types, the internal connectivity of a fragment, and the manner in which a fragment connects to the surrounding region.

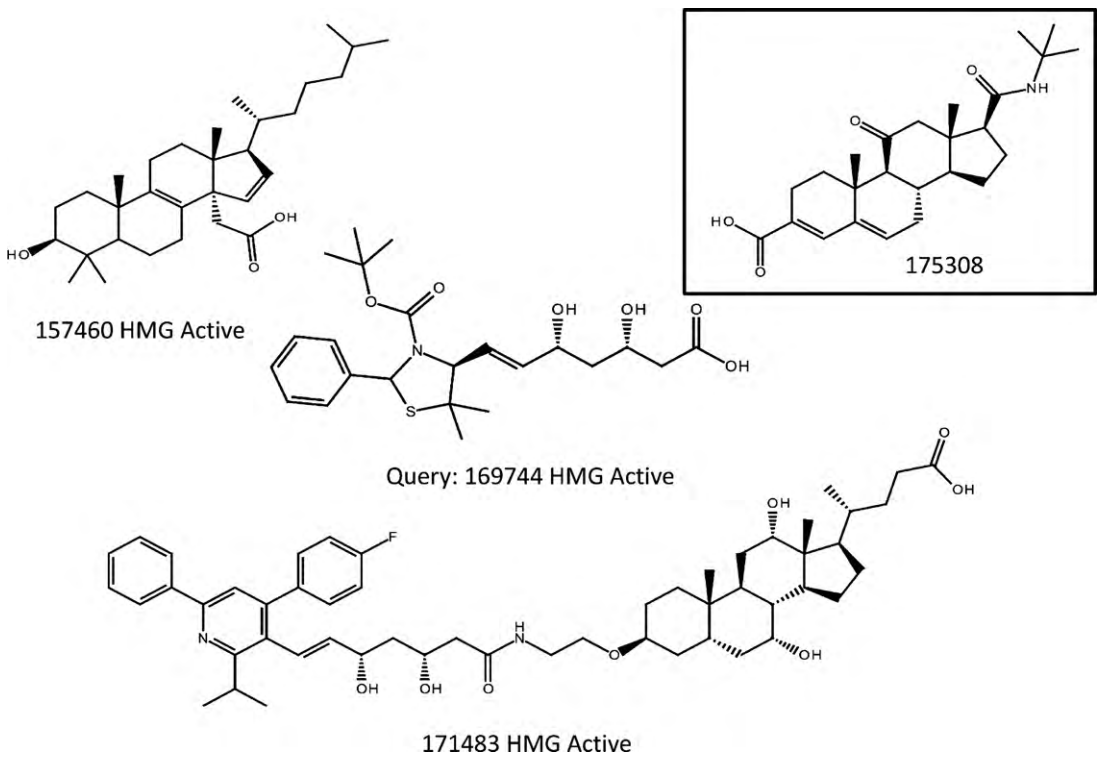

Fig. 13. Example of a high-scoring inactive molecule that looks similar to known HMG actives. The Triplet-Hybr screen with molecule 169744 (center) as a query retrieved no HMG actives in the top 1% of the hits; however, 175308, highlighted in the box, shares the same steroid-like scaffold as two other HMG actives (157460 and 171483).

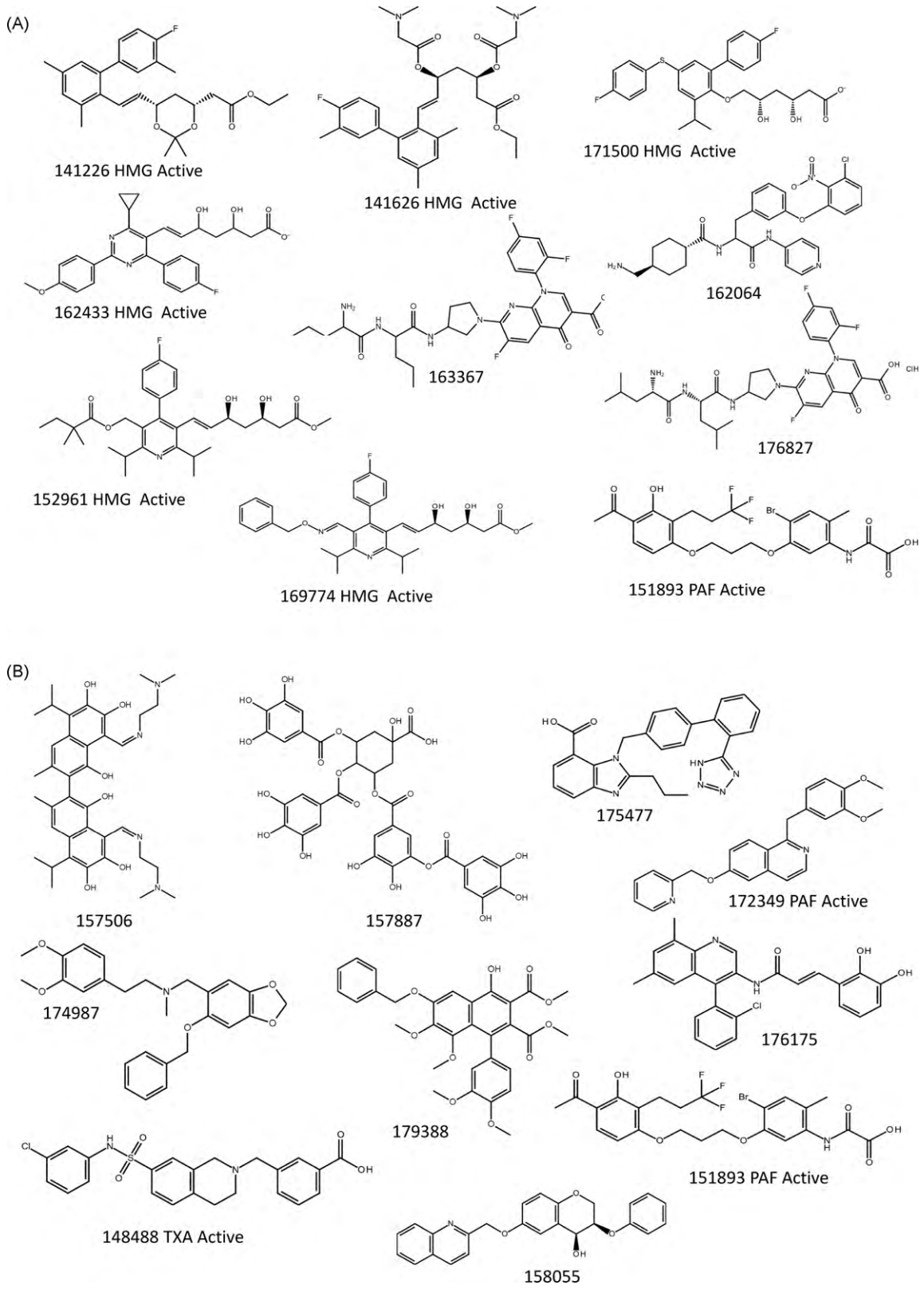

Fig. 14. Top 10 molecules retrieved by Triplet-Hybr (A) and Triplet-Day (B) using the HMG active Rawsonol (shown in Fig. 6) as the query. Triplet-Hybr retrieved 6 HMG actives whereas Triplet-Day retrieved none.

By contrast, MOLPRINT2D fingerprints incorporate only counts of various atom types at different radial distances. However, a given fingerprint can be made more or less fuzzy by changing the way in which atoms and bonds are distinguished. Thus, in addition to the comparison of different fingerprinting methods, we have also explored how the various atom-typing schemes can affect results.

It is expected that when using an atom-typing scheme with low information content, it will be more difficult to selectively retrieve similar compounds from a database of decoys. On the other hand, there may be some benefit to applying a higher level of fuzzi-

ness in order to retrieve more diverse compounds. To investigate theses ideas, atom-typing schemes with different levels of information content were applied to the two best performing fingerprints, Radial and Triplet. The default scheme for both was "Day," which employs Daylight invariant atom types, and distinguishes bonds by bond order. This highly specific scheme was contrasted with two fuzzy approaches: functional atom types (H, C, [F, CL], [Br,I], [N,O], S, other), referred to as "Fn," and hybridization-based atom types, abbreviated as "Hybr." Unlike the default Day scheme, both of these fuzzy methods regard all bonds as equivalent. We will refer to the fingerprints in combination with these atom-typing schemes as <fingerprint>–<atom type>, where <fingerprint> is Triplet or Radial and <atom type> is Day, Fn, or Hybr.

In the case of Radial-Day fingerprints, the average retrieval rate across all targets decreased only slightly from the default of 7.9, to 7.8 for Radial-Fn and 7.3 for Radial-Hybr. By comparison, the effect on Triplet fingerprints was far more substantial, with the average retrieval rate dropping from 7.8 with Triplet-Day to 6.5 for Triplet-Fn and 6.0 for Triplet-Hybr. According to the paired Student's t-test, the degradation is significant, with a P value less than 0.01 for all fingerprint-atom type pairs except Radial-Day and Radial-Fn. This finding is perhaps not surprising, since Radial fingerprints are inherently specific and therefore not as sensitive to atom typing.

Next, we examined cases where certain fingerprints produced no enrichment above random. An example is found with the HMG query in Fig. 13, where Triplet-Hybr failed to retrieve any active compounds in the top 1% of the database. Interestingly, Triplet-Hybr ranks molecule 175308 highly, which shares the same steroid-like scaffold as two other actives in the HMG class. This suggests that the use of a fuzzier atom-typing scheme may in fact promote scaffold hopping capabilities. Surprisingly, we found two cases where Triplet-Day fingerprints identified no actives but Triplet-Hybr achieved a reasonable retrieval rate. One of these cases is shown in Fig. 14. As mentioned in earlier results, queries involving the HMG inhibitor Rawsonol (molecule 159170) yielded no actives in the top 1% when Daylight atom types were used. Visual inspection of the HMG inhibitors shows Rawsonol as the only non-statin, which reduces its ability to identify other inhibitors. Interestingly, Triplet fingerprints with low information content atom types (Hybr) find six actives. This result shows that a fuzzy atom-typing scheme may be a key factor in facilitating scaffold hopping with 2D fingerprints.

4. Conclusions

Similarity searching based on 2D fingerprints is frequently used in drug discovery, but it is rarely known which fingerprint is most suitable for the particular target and query. Consequently, it is important to validate different fingerprinting methods against a multitude of targets. In this work, most fingerprints performed comparably on average and there was no single best method for all target classes. However, for a given target and query there were significant differences between the methods. This re-emphasizes the need for a variety of 2D similarity methods, as Sheridan and Kearsley suggested [62]. On the other hand, we found that MOLPRINT2D provides a good baseline when there is no previous experience to suggest which fingerprinting method might work best.

We found that the size dependency of Pairwise and Triplet fingerprints can be a limitation when query molecules are significantly different in size from other actives. However, this same feature may be an asset when active molecules similar in size to the query are being sought. The usefulness of combining the results from two orthogonal fingerprints was also investigated. Although this approach did not lead to significantly higher average retrieval rates, it did suggest a viable means of minimizing the risk of choosing the wrong fingerprinting method for a given screen. Similarly, a single modal fingerprint derived from multiple queries was found in general to be superior to using individual fingerprints from different queries.

An investigation into the information content of fingerprints and atom-typing schemes revealed that no approach works best in all cases. However, if fingerprints are being used in a late stage lead optimization project, where compounds that are very similar to a known lead are desirable, a high-information content fingerprint and/or a high-information content atom-typing scheme might be more appropriate. Alternatively, if the objective is to identify actives that are chemically diverse, such as in early lead discovery, a fuzzier atom-typing scheme should be considered.

Acknowledgement

The authors thank Dr. Ken Dyall for helpful comments in the preparation of this manuscript.

References

[1] H.-J. Böhm, G. Schneider, Virtual Screening for Bioactive Molecules, Wiley-VCH, New York, 2000.
[2] B.K. Shoichet, J. Alvarex, Virtual Screening in Drug Discovery, CRC Press, Boca Raton, FL, 2005.
[3] W.P. Walters, M.T. Stahl, M.A. Murcko, Virtual screening—an overview, Drug Discov. Today 3 (1998) 160–178.
[4] A. Ajay, W.P. Walters, M.A. Murcko, Can we learn to distinguish between "drug-like" and "nondrug-like" molecules? J. Med. Chem. 41 (1998) 3314–3324.
[5] R.D. Brown, M.A. Martin, The information content of 2D and 3D structural descriptors relevant to ligand–receptor bond, J. Chem. Inf. Comput. Sci. 37 (1997) 1–9.
[6] R.D. Brown, Y.C. Martin, Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection, J. Chem. Inf. Comput. Sci. 36 (1996) 572–584.
[7] S.L. Dixon, K.M. Merz Jr., One-dimensional molecular representations and similarity calculations: methodology and validation, J. Med. Chem. 44 (2001) 3795–3809.
[8] H. Matter, Selecting optimally diverse compounds from structure databases: a validation study of two-dimensional and three-dimensional molecular descriptors, J. Med. Chem. 40 (1997) 1219–1229.
[9] G.B. McGaughey, R.P. Sheridan, C.I. Bayly, J.C. Culberson, C. Kreatsoulas, S. Lindsley, et al., Comparison of topological, shape, and docking methods in virtual screening, J. Chem. Inf. Model. 47 (2007) 1504–1519.
[10] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, et al., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, J. Med. Chem. 47 (2004) 1739–1749.
[11] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, A geometric approach to macromolecule-ligand interactions, J. Mol. Biol. 161 (1982) 269–288.
[12] M.R. McGann, H.R. Almond, A. Nicholls, J.A. Grant, F.K. Brown, Gaussian docking functions, Biopolymers 68 (2003) 76–90.
[13] M.D. Miller, S.K. Kearsley, D.J. Underwood, R.P. Sheridan, FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure, J. Comput. Aided Mol. Des. 8 (1994) 153–174.
[14] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, A fast flexible docking method using an incremental construction algorithm, J. Mol. Biol. 261 (1996) 470–489.
[15] D.E. Clark, G. Jones, P. Willett, Pharmacophoric pattern matching in files of three-dimensional chemical structures: comparison of conformational searching algorithms for flexible searching, J. Chem. Inf. Comput. Sci. 34 (1994) 197–206.
[16] J. Greene, S. Kahn, H. Savoj, P. Sprague, S. Teig, Chemical function queries for 3D database search, J. Chem. Inf. Comput. Sci. 34 (1994) 1297–1308.
[17] O.F. Guener, D.R. Henry, R.S. Pearlman, Use of flexible queries for searching conformationally flexible molecules in databases of three-dimensional structures, J. Chem. Inf. Comput. Sci. 32 (1992) 101–109.
[18] P. Gund, Three-dimensional pharmacophore pattern searching, in: F.E. Hahn (Ed.), Progress in Molecular and Subcellular Biology, vol. 5, Springer-Verlag, Berlin, 1977, pp. 117–143.
[19] ROCS, version 3.0, OpenEye Scientific Software, Sanra Fe, NM, 2009.
[20] P.J. Ballester, W.G. Richards, Ultrafast shape recognition to search compound databases for similar molecular shapes, J. Comput. Chem. 28 (2007) 1711–1723.
[21] A.Y. Meyer, W.G. Richards, Similarity of molecular shape, J. Comput. Aided Mol. Des. 5 (1991) 427–439.
[22] S. Putta, C. Lemmen, P. Beroza, J. Greene, A novel shape-feature based approach to virtual library screening, J. Chem. Inf. Comput. Sci. 42 (2002) 1230–1240.
[23] S.L. Dixon, R.T. Koehler, The hidden component of size in two-dimensional fragment descriptors: side effects on sampling in bioactive libraries, J. Med. Chem. 42 (1999) 2887–2900.
[24] D.R. Flower, On the properties of bit string-based measures of chemical similarity, J. Chem. Inf. Comput. Sci. 38 (1998) 379–386.

[25] S.J. Cato, Exploring pharmacophores with Chem-X, in: O.F. Guener (Ed.), Pharmacophore Perception, Development and Use in Drug Design, International University Line, La Jolla, CA, 2000, pp. 110–125.
[26] R.D. Cramer, M.A. Poss, M.A. Hermsmeier, T.J. Caulfield, M.C. Kowala, M.T. Valentine, Prospective identification of biologically active structures by topomer shape similarity searching, J. Med. Chem. 42 (1999) 3919–3933.
[27] J.A. Grant, J.A. Haigh, B.T. Pickup, A. Nicholls, R.A. Sayle, Lingos, finite state machines, and fast similarity searching, J. Chem. Inf. Model. 46 (2006) 1912–1918.
[28] M.J. McGregor, S.M. Muskal, Pharmacophore fingerprinting. 1. Application to QSAR and focused library design, J. Chem. Inf. Comput. Sci. 39 (1999) 569–574.
[29] M. Rarey, J.S. Dixon, Feature trees: a new molecular similarity measure based on tree matching, J. Comput. Aided Mol. Des. 12 (1998) 471–490.
[30] D.J. Wild, C.J. Blankley, Comparison of 2D fingerprint types and hierarchy level selection methods for structural grouping using Ward's clustering, J. Chem. Inf. Comput. Sci. 40 (2000) 155–162.
[31] L. Xue, F.L. Stahura, J.W. Godden, J. Bajorath, Fingerprint scaling increases the probability of identifying molecules with similar activity in virtual screening calculations, J. Chem. Inf. Comput. Sci. 41 (2001) 746–753.
[32] Q. Zhang, I. Muegge, Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring, J. Med. Chem. 49 (2006) 1536–1548.
[33] Canvas, Schrödinger, LLC, New York, NY, 2009.
[34] X. Chen, C.H. Reynolds, Performance of similarity measures in 2D fragment-based similarity searching: comparison of structural descriptors and similarity coefficients, J. Chem. Inf. Comput. Sci. 42 (2002) 1407–1414.
[35] R.C. Glen, S.E. Adams, Similarity metrics and descriptor spaces—which combinations to choose? QSAR Comb. Sci. 25 (2006) 1133–1142.
[36] J. Hert, P. Willett, D.J. Wilton, P. Acklin, K. Azzaoui, E. Jacoby, et al., Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures, Org. Biomol. Chem. 2 (2004) 3256–3266.
[37] P. Willett, Similarity-based virtual screening using 2D fingerprints, Drug Discov. Today 11 (2006) 1046–1053.
[38] A. Bender, J.L. Jenkins, J. Scheiber, S.C. Sukuru, M. Glick, J.W. Davies, How similar are similarity searching methods? A principal component analysis of molecular descriptor space, J. Chem. Inf. Model. 49 (2009) 108–119.
[39] M. Madhavi, J.F. Lowrie, S.L. Dixon, W. Sherman, Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments, J. Chem. Inf. Model. 50 (2010) 771–784.
[40] H. Briem, I.D. Kuntz, Molecular similarity based on DOCK-generated fingerprints, J. Med. Chem. 39 (1996) 3401–3408.
[41] H. Briem, U.F. Lessel, In vitro and in silico affinity fingerprints: finding similarity beyond structural classes, Perspect. Drug Discov. 20 (2000) 231–244.
[42] U.F. Lessel, H. Briem, Flexsim-X: a method for the detection of molecules with similar biological activity, J. Chem. Inf. Comput. Sci. 40 (2000) 246–253.
[43] MACCS-II, MDL Information Systems/Symyx, Santa Clara, CA, 1984.
[44] Daylight Fingerprint Toolkit, Daylight Chemical Systems Inc., Aliso Viejo, CA, version 4.9, 2008.
[45] Unity, Tripos L.P., St. Louise, MO, version 4.4, 2003.
[46] Daylight Theory Manual, Daylight Chemical System Inc., Aliso Viejo, CA, 2008.
[47] M.S. Lajiness, Dissimilarity-based compound selection techniques, Perspect. Drug Discov. 7/8 (1997) 65–84.
[48] D. Rogers, R.D. Brown, M. Hahn, Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up, J. Biomol. Screen. 10 (2005) 682–686.
[49] D. Rogers, M. Hahn, Extended-connectivity fingerprints. J. Chem. Inf. Model. 50 (2010) 724–754.
[50] H.L. Morgan, The generation of a unique machine description for chemical structure—a technique developed at Chemical Abstract Service, J. Chem. Doc. 5 (1965) 107–113.
[51] R.E. Carhart, D.H. Smith, R. Venkataghavan, Atom pairs as molecular features in structure-activity studies: definition and applications, J. Chem. Inf. Comput. Sci. 25 (1985) 65–73.
[52] R. Nilakantan, N. Bauman, J.S. Dixon, R. Venkataghavan, Topological torsions: a new molecular descriptos for SAR applications. Comparison with other descriptors, J. Chem. Inf. Comput. Sci. 27 (1987) 82–85.
[53] A. Bender, H.Y. Mussa, R.C. Glen, S. Reiling, Molecular similarity searching using atom environments, information-based feature selection, and a naive Bayesian classifier, J. Chem. Inf. Comput. Sci. 44 (2004) 170–178.
[54] A. Bender, H.Y. Mussa, R.C. Glen, S. Reiling, Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): evaluation of performance, J. Chem. Inf. Comput. Sci. 44 (2004) 1708–1718.
[55] SMARTS – Language for describing molecular patterns, Daylight Chemical Information Systems Inc., Aliso Viejo, CA, 2008.
[56] D. Weininger, A. Weininger, J.L. Weininger, SMILES 2. Algorithm for generation of unique SMILES notation, J. Chem. Inf. Comput. Sci. 29 (1989) 97–101.
[57] MDL Drug Data Report, MDL Information Systems/Symyx, Santa Clara, CA.
[58] A. Evers, G. Hessler, H. Matter, T. Klabunde, Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols, J. Med. Chem. 48 (2005) 5448–5465.
[59] C. Gerlach, H. Broughton, A. Zaliani, FTree query construction for virtual screening: a statistical analysis, J. Comput. Aided Mol. Des. 22 (2008) 111–118.
[60] M. Rarey, M. Stahl, Similarity searching in large combinatorial chemistry spaces, J. Comput. Aided. Mol. Des. 15 (2001) 497–520.
[61] D. Vidal, M. Thormann, M. Pons, A novel search engine for virtual screening of very large databases, J. Chem. Inf. Model. 46 (2006) 836–843.
[62] R.P. Sheridan, S.K. Kearsley, Why do we need so many chemical similarity search methods? Drug Discov. Today 7 (2002) 903–911.
[63] M. Feher, Consensus scoring for protein–ligand interactions, Drug Discov. Today 11 (2006) 421–428.
[64] J.C. Baber, W.A. Shirley, Y. Gao, M. Feher, The use of consensus scoring in ligand-based virtual screening, J. Chem. Inf. Model. 46 (2006) 277–288.
[65] J.M. Yang, Y.F. Chen, T.W. Shen, B.S. Kristal, D.F. Hsu, Consensus scoring criteria for improving enrichment in virtual screening, J. Chem. Inf. Model. 45 (2005) 1134–1146.
[66] R. Teramoto, H. Fukunishi, Consensus scoring with feature selection for structure-based virtual screening, J. Chem. Inf. Model. 48 (2008) 288–295.
[67] N.E. Shemetulskis, D. Weininger, C.J. Blankley, J.J. Yang, C. Humblet, Stigmata: an algorithm to determine structural commonalities in diverse datasets, J. Chem. Inf. Comput. Sci. 36 (1996) 862–871.
[68] J. Hert, P. Willett, D.J. Wilton, P. Acklin, K. Azzaoui, E. Jacoby, et al., Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures, J. Chem. Inf. Comput. Sci. 44 (2004) 1177–1185.
[69] Sybyl, version 8.1.1, Tripos L.P., St. Louise, MO, 2009.
[70] L.B. Kier, L.H. Hall, Molecular Structure Description: The Electrotopological State, San Diego, CA, Academic Press, 1999.