Mini-fingerprints Detect Similar Activity of Receptor Ligands Previously Recognized Only by Three-Dimensional Pharmacophore-Based Methods

Ling Xue, Florence L. Stahura, Jeffrey W. Godden, and Jürgen Bajorath*,†

New Chemical Entities, Inc., 18804 North Creek Parkway, Suite 100, Bothell, Washington 98011

Received September 12, 2000

Mini-fingerprints (MFPs) are short binary bit string representations of molecular structure and properties, composed of few selected two-dimensional (2D) descriptors and a number of structural keys. MFPs were specifically designed to recognize compounds with similar activity. Here we report that MFPs are capable of detecting similar activities of some druglike molecules, including endothelin A antagonists and α_1-adrenergic receptor ligands, the recognition of which was previously thought to depend on the use of multiple point three-dimensional (3D) pharmacophore methods. Thus, in these cases, MFPs and pharmacophore fingerprints produce similar results, although they define, in terms of their complexity, opposite ends of the spectrum of methods currently used to study molecular similarity or diversity. For each of the studied compound classes, comparison of MFP bit settings identified a consensus or signature pattern. Scaling factors can be applied to these bits in order to increase the probability of finding compounds with similar activity by virtual screening.

INTRODUCTION

Computational methods to evaluate molecular similarity and diversity or screen databases for compounds with desired biological specificity are of significant interest in pharmaceutical research. These approaches are based on a variety of scientific concepts. Consequently, computational tools to identify molecules with similar activity are highly diverse in their design, ranging from substructure[1] and structural fragment-type search strings[2] to much more complex two-[3] or three-dimensional[4,5] molecular representations, often encoded as binary bit strings or fingerprints.

There is considerable debate in the literature concerning the relative performance of simple or more complex methods in similarity/diversity analysis and virtual screening,[6,7] and the views are sometimes controversial. In part, these discussions focus on the question of whether 2D or 3D methods yield superior results. Some analyses suggest that 2D representations or descriptors of molecules may often be suitable,[8−11] whereas other studies support the advantages of 3D or combined 2D/3D approaches.[12−14] Since many of these investigations use different test cases and algorithms, they cannot be directly compared, and it is therefore difficult at present to draw more general conclusions.

One can certainly argue that 3D methods should a priori be more powerful than 2D approaches, since molecules definitely interact in three dimensions. However, 3D methods are often also more prone to errors, in particular, when attempting to accurately predict bioactive conformations of ligands. Moreover, in many cases, molecular features critical for a specific activity may be readily identified by comparison of 2D representations of active and inactive molecules. In any event, ultimately, the balance between resolution and error margins of different molecular similarity methods determines their relative performance.

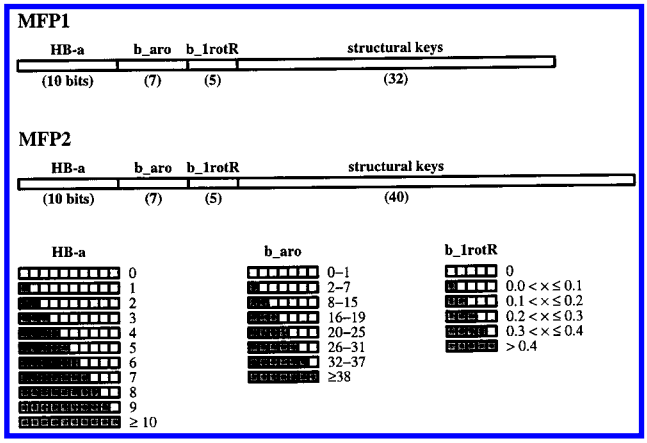


Figure 1. Mini-fingerprints. The design of two binary MFPs used in this study is schematically illustrated. MFP1 and MFP2 consist of only 54 and 62 bit positions, respectively. Both MFPs share three numerically encoded descriptors, HB-a (number of hydrogen bond acceptors in a molecule), b__aro (number of aromatic bonds), and b__1rotR (fraction of rotatable bonds). In addition, they contain different sets of structural keys. The bottom part of the figure illustrates how bit segments are used to encode numerical descriptors. Shaded bit positions are set on (i.e., 1).

Multiple point 3D pharmacophore methods and fingerprints were introduced as complex and high-resolution methods to study various aspects of molecular similarity or diversity.[6,13] Multiple point pharmacophores capture and compare the spatial arrangements of predefined functional sites or features in molecules that are known or thought to be relevant for specific interactions[15] based on complete, or nearly complete, conformational analysis. Pharmacophore fingerprints thus derived may literally consist of millions of bits.

In independent studies, making use of similar yet distinct four-point pharmacophore concepts, several examples have

* To whom correspondence should be addressed at NCE, Inc. Telephone: (425) 424-7297. Fax: (425) 424-7299. E-mail: jbajorath@nce-mail.com.

† NCE and Department of Biological Structure, University of Washington, Seattle, WA 98195.

10.1021/ci000305x CCC: $20.00
Published on Web 01/30/2001

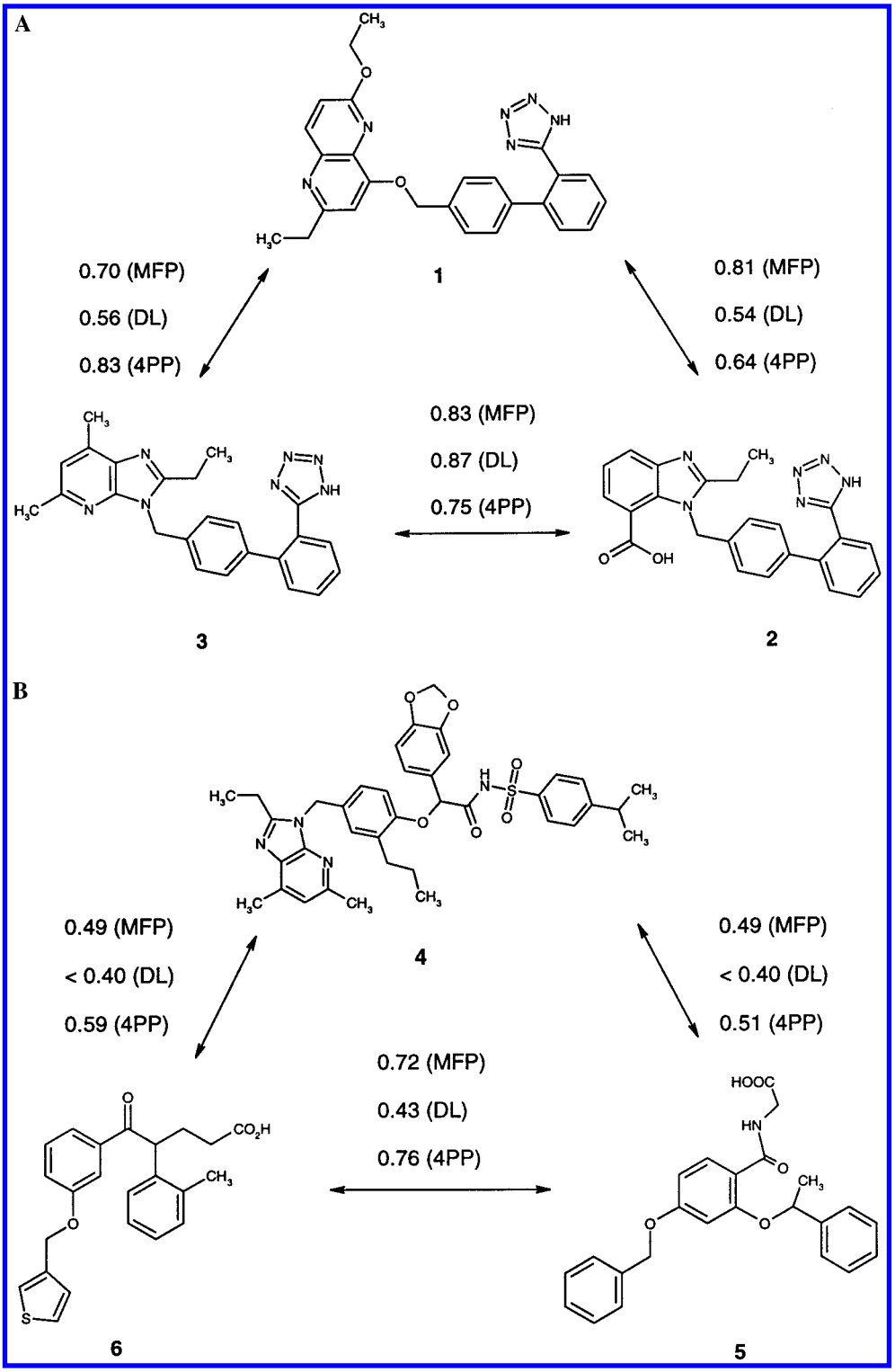
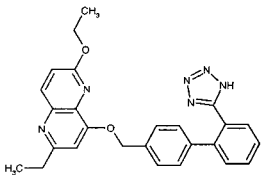
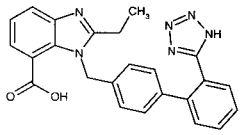
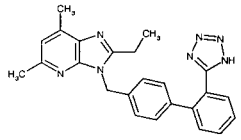


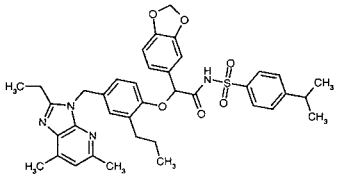
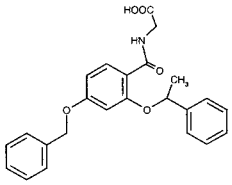
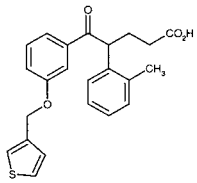
Figure 2. Comparison of angiotensin II and endothelin A antagonists. In A (top), angiotensin II antagonists are shown, and endothelin A antagonists are shown in B (bottom). Structures are numbered as in Table 1. Tc (-like) values are repoted for the Daylight[3] (DL) and four-point pharmacophore[16] (4PP) fingerprints and the best MFP according to Table 1.

recently been presented where these 3D methods succeeded to recognize activity relationships of compounds that were otherwise difficult, if not impossible, to detect.[16,17] In these studies, the well-known Daylight fingerprint (DL)[3] was used as a 2D benchmark method for comparison. Accordingly, these test cases, including endothelin A antagonists[16] and α_1-adrenergic receptor ligands,[17] were thought to provide significant support for the opportunities of these pharmacophore approaches and their superior performance, as compared to 2D methods.

In this study, we have analyzed these examples using mini-fingerprints (MFPs) that represent only a limited number of 2D descriptors. These descriptors were specifically selected on the basis of their ability to correctly classify compounds according to their biological activity.[18,19] MFPs were designed as short binary bit strings, consisting of only 50−60 bit positions.[20,21] By contrast, DL, which captures connectivity pathways through molecules, consists of 2048 bits. Thus,

Table 1: Comparison of Angiotensin II and Endothelin A Antagonists Using MFPs and MACCS[a]

Structures	1	2	Fingerprint
1 			MFP1
			MFP2
			MACCS
2 	**0.78**		MFP1
	0.81		MFP2
	0.64		MACCS
3 	**0.70**	**0.83**	MFP1
	0.69	**0.82**	MFP2
	0.62	0.81	MACCS

Structures	4	5	Fingerprint
4 			
5 	0.49		MFP1
	0.46		MFP2
	0.41		MACCS
6 	0.49	**0.72**	MFP1
	0.38	**0.71**	MFP2
	0.37	0.57	MACCS

[a] Tc values are reported for pairwise comparison of compounds. For MFPs, values greater than their similarity threshold are shown in boldface.

even compared to standard 2D fingerprints, MFPs are simple (yet rational) in their design. Moreover, MFPs are, in terms of their complexity, fundamentally different from four-point pharmacophore fingerprints.

In our analysis, we found that MFPs successfully recognized a number of similarity relationships, the detection of which was previously thought to be limited to 3D pharmacophore analysis. Importantly, these findings illustrate that approaches essentially occupying opposite ends of the spectrum of similarity search methods (from simple 2D to extensive 3D fingerprints) can handle "difficult" test cases in a comparable manner. We also show that MFP signature motifs can be derived by comparison of compounds within an activity class.

MATERIALS AND METHODS

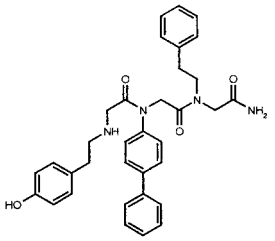
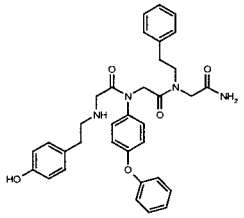
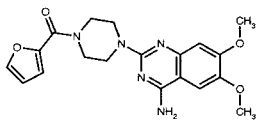
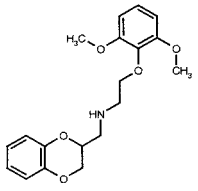
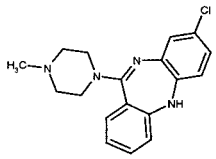
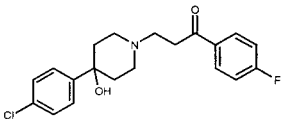
Compounds with specific activities analyzed in this study were taken from Mason et al.[16] (angiotensin II antagonists, endothelin A antagonists) and Bradley et al.[17] (α_1-adrenergic receptor ligands) and studied using two MFPs (1 and 2) that we implemented into the Molecular Operating Environment.[22] Pairwise comparisons of compounds were carried out by calculating the Tanimoto coefficient (Tc)[23] for fingerprint overlap, defined as $Tc = Bc/(B1 + B2 - Bc)$. Bc is the number of common bits set on (equal to 1) in molecules 1 and 2, and B1 and B2 are the number of bits set on in molecules 1 and 2, respectively.

The design of MFP1 and MFP2, as described in detail previously,[19,21] is illustrated in Figure 1. Both MFPs consist of three numerically encoded descriptors and a set of 32 and 40 single structural fragments or keys, respectively, which were selected on the basis of compound classification calculations.[18,19,21] MFP1 and MFP2 have only four structural fragments in common. However, in similarity search blind test calculations targeting a number of compound classes with distinct activity, MFP1 and MFP2 displayed comparable performance,[21] as discussed below in more detail.

MFP design is "keyed", which means that each bit position is associated with a fraction of a value of numerically encoded descriptors or the presence or absence of a structural fragment. By contrast, DL is a representative example of a hashed or folded fingerprint where properties or patterns are mapped to overlapping bit segments. For comparison with MFPs, Tc values obtained for DL were taken from the original publications.[16,17] As an additional reference fingerprint, we also used a set of 166 MACCS keys.[24]

MFP signature profiles were obtained by comparing the bit settings for each compound belonging to a specific antagonist or ligand class and identifying those bit positions that were set on in all fingerprints of the same class. In our approach, these positions define a signature pattern distributed over an MFP. Applying a scaling factor to each bit position of the signature motif effectively multiplies the number of common bits for comparison of MFPs and thus yields higher Tc values (which we call "sTc" for scaled Tc) for compounds sharing the signature pattern. This "fingerprint scaling" technique may therefore be used to increase the probability of identifying compounds with similar activity by database searching, provided meaningful signature motifs can be derived for specific compound classes. The definition of MFP signature motifs is somewhat similar to our "fingerprint profiling" method,[25] where a consensus profile for a specific activity class is obtained by determining the relative bit occupancy at each position of a fingerprint. Following our definition, signature motifs consist of all bit positions with relative occupancy of 1.0.

Table 2: Comparison of α_1-Adrenergic Receptor Ligands Using MFPs and MACCS[a]

Structures	7	8	9	10	11	Fingerprint
7  NSG trimer (α_1-adrenergic receptor ligand)						
8  NSG trimer (α_1-adrenergic receptor ligand)	**0.93**					**MFP1**
	0.93					**MFP2**
	0.94					**MACCS**
9  Prazosin	**0.68**	**0.62**				**MFP1**
	0.69	**0.69**				**MFP2**
	0.65	0.67				**MACCS**
10 	**0.64**	**0.63**	**0.64**			**MFP1**
	0.66	**0.71**	0.56			**MFP2**
	0.55	0.60	0.56			**MACCS**
11  Clozapine	0.45	0.43	0.45	0.44		**MFP1**
	0.43	0.39	0.49	0.31		**MFP2**
	0.46	0.43	0.53	0.37		**MACCS**
12  Haloperidol	0.45	0.43	0.41	0.40	**0.73**	**MFP1**
	0.48	0.44	0.42	0.46	0.59	**MFP2**
	0.58	0.54	0.46	0.50	0.49	**MACCS**

[a] "NSG" stands for N-substituted glycine. For MFPs, values greater than their similarity threshold are shown in boldface.

RESULTS AND DISCUSSION

Detection of Molecular Similarity. A variety of metrics and criteria can be applied to evaluate the similarities of compounds.[23] Our approach relies on pairwise comparison of MFPs and calculation of conventional Tc values, which is currently the most commonly used metric. As a similarity cutoff value, a Tc of 0.85 is often used in the literature, regardless of the particular search tools. This value has its

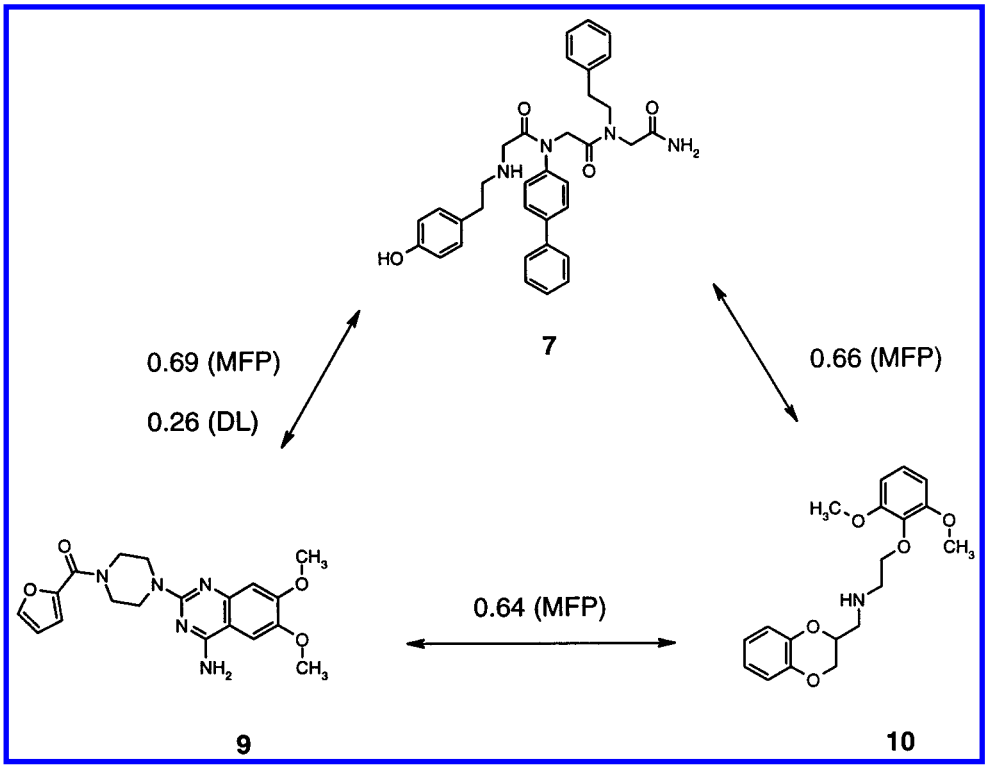


Figure 3. Comparison of α_1-adrenergic receptor ligands. Structures are numbered as in Table 2, and the representation corresponds to the one in Figure 2.

origin in the analysis of "neighborhood behavior"[12] but is in many applications more or less arbitrarily applied. With the development of MFPs, we have attempted to systematically determine similarity cutoff values that yield the best results specifically for the application of our fingerprints. Therefore, we have tested the performance of MFPs by extensive calculations in benchmark databases consisting of compounds belonging to different activity classes. Each compound was searched against the remainder of the database under systematic variation of Tc cutoff values.[20,21] In these and subsequent benchmark calculations, we found that MFPs displayed overall the best performance at Tc cutoff values between 0.65 and 0.7. Within this "similarity interval", MFPs have, on average, an approximately 60% chance of correctly identifying all molecules with biological activity similar to a query compound and recognize only 1−2% false positives. In virtual screening calculations, we therefore use a Tc threshold value of 0.6 as an MFP-specific similarity criterion and initially regard any match having a Tc of 0.6 or greater as a "hit".

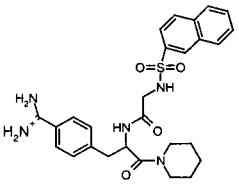
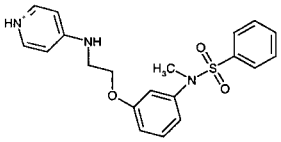
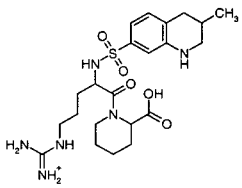
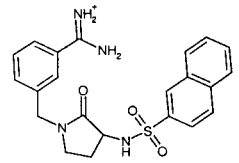
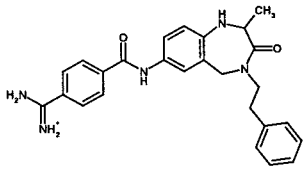
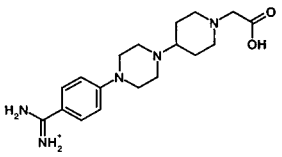
Tc values or similar metrics may or may not be applied when comparing pharmacophore fingerprints. For example, Mason et al.[16] use a Tc-like value to compare four-point pharmacophores of compounds (see below), whereas Bradley et al.[17] do not report coefficients to evaluate molecular similarity but apply a pharmacophore filter that best separates active from inactive molecules. For the examples described below, this filter corresponded to a 40% match of the complete pharmacophore ensemble.

Angiotensin II and Endothelin A Antagonists. When describing their four-point pharmacophore method, Mason et al. presented angiotensin II and endothelin A antagonists as examples of increasing difficulty.[16] Structures of the angiotensin II antagonists (compounds **1**−**3**) and endothelin A antagonists (compounds **4**−**6**) are shown in Figure 2A,B, respectively, together with calculated Tc values for DL and Tc-like values for four-point pharmacophore fingerprints (4PP). We have calculated Tc values for pairwise comparison of these compounds using MFP1, MFP2, and MACCS. The results are reported in Table 1A,B, respectively. Best results obtained with our MFPs are also shown in Figure 2A,B for comparison.

The angiotensin II antagonists are structurally related to an extent that 2D fingerprints such as DL and MACCS can at least weakly detect their similarity. As can be seen, 4PP comparisons identify these similarities more clearly. However, all of these similarities are also clearly detected using MFPs, with best Tc values ranging from 0.70 to 0.83 (Figure 2A). The endothelin A antagonists shown in Figure 2B represent a more challenging test case. Here similarities are not obvious, and 2D fingerprints fail to detect these relationships. Tc-like values calculated using 4PP are also lower than for angiotensin II antagonists, except for comparison of compounds **5** and **6**. However, this similarity is also detected using both MFPs, with Tc values greater than 0.7.

α_1-Adrenergic Receptor Ligands. Bradley et al.[17] discussed a particularly challenging test case for similarity methods, α_1-adrenergic receptor ligands. Their activity relationship could be successfully identified using the four-point pharmacophore filter technique but not by DL-based comparison.[17] Structures of these compounds are shown in Table 2. Compounds **7** and **8** (NSG trimers), which are closely related, represent molecules that were identified from a combinatorial library as α_1-adrenergic receptor ligands with low to high nanomolar activity. Their similarity to known ligands with low nanomolar activity, taken from the literature, was detected using the pharmacophore filter test. Four examples of literature compounds (**9**−**12**) are shown in Table

Table 3: Comparison of Different Serine Protease Inhibitors and Fibrinogen Receptor Antagonists Using MFPs and MACCS[a]

Structures	NAPAP	Boeh. Mann	MQPA	RPR 118071	MDDR 192259	Fingerprint
 NAPAP	1					MFP1
	1					MFP2
	1					MACCS
 Boeh. Mann	**0.64**	1				MFP1
	0.75	1				MFP2
	0.61	1				MACCS
 MQPA	**0.71**	0.54	1			MFP1
	0.76	**0.65**	1			MFP2
	0.84	0.61	1			MACCS
 RPR 118071	**0.83**	0.59	**0.62**	1		MFP1
	0.88	**0.70**	**0.67**	1		MFP2
	0.87	0.62	0.8	1		MACCS
 MDDR 192259 (FRA-1)	**0.66**	0.47	0.5	**0.73**	1	MFP1
	0.60	0.53	0.47	**0.65**	1	MFP2
	0.58	0.42	0.56	0.65	1	MACCS
 MDDR 199187 (FRA-2)	0.5	0.41	0.55	0.5	**0.67**	MFP1
	0.53	0.5	0.53	0.53	**0.74**	MFP2
	0.57	0.42	0.57	0.57	0.73	MACCS

[a] Tc values are reported for pairwise comparison of compounds. For MFPs, values greater than their similarity threshold are shown in boldface. Compounds were taken from ref 26. The top four structures shown are inhibitors of different serine proteases. The first three compounds are thrombin inhibitors, and RPR 118071 is an inhibitor of factor Xa. The other two structures are different fibrinogen receptor antagonists (FRAs).

2. Compounds **7** and **8** are larger and more complex than literature examples **9**−**12,** and it is certainly difficult to "see" that these two groups of compounds are similar or predict that they are related in terms of their activity. In pairwise comparisons, DL consistently produced very low Tc values, on average 0.27.[17] Our calculations reveal that both MACCS keys and MFPs perform better than DL in these cases. As a trend, increasing complexity of these 2D fingerprints appears to lead to lower Tc values, consistent with the idea that complex fingerprints become more sensitive to small structural variations, which may not be relevant for maintaining specific activity.[20] Importantly, using MFPs, we clearly identify the relationship between NSG trimers and compounds **9** and **10**, with Tc values consistently greater than

the MFP similarity threshold, but fail to do so for compounds **11** and **12** (clozapine and haloperidol, respectively). However, using MFP1, we also detect similarity between these two compounds, which is again not at all obvious. Figure 3 summarizes calculations for compounds **9** (prazosin) and **10**. For both compounds, MFPs indicate the similarity to NSG trimers. In addition, MFP1 (but not MFP2) suggests that these two molecules are similar to each other.

Serine Protease Inhibitors and Fibrinogen Receptor Antagonists. As an additional and conceptually different example, we have compared sets of serine protease inhibitors and fibrinogen receptor antagonists, taken from the literature.[26] The structures are shown in Table 3. Here the task was not only to recognize similarity of molecules but also to differentiate between structurally related compounds having distinct activities. In previous studies, four-point 3D pharmacophore fingerprints could differentiate between the selectivity of these protease inhibitors (however, only when information about the enzyme binding sites was taken into account) and, furthermore, structurally related fibrinogen receptor antagonists were correctly predicted to be inactive against these targets.[26] Results of our calculations, reported in Table 3, again demonstrate that 2D fingerprints successfully capture at least some of these findings. MFPs clearly recognize the similarity of all protease inhibitors and the similarity of the two fibrinogen receptor antagonists. Overall MFP2 performed somewhat better then MFP1 in these calculations, and MACCS keys also performed well. As expected, none of these fingerprints distinguished the factor Xa-specific inhibitor from the thrombin inhibitors. However, it was possible to partly differentiate between protease inhibitors and fibrinogen receptor antagonists on the basis of our calculations. Whereas incorrect similarity between antagonist FRA-1 and two of the four protease inhibitors was detected, FRA-2 was completely distinguished from these inhibitors. FRA-1 could only be differentiated from the serine protease inhibitors by application of four-point (but not three-point) pharmacophores capturing features of protease active sites.[26]

Signature Motifs and MFP Scaling. In Figure 4, we show a comparison of MFP2 bit settings for angiotensin II antagonists **1**−**3** (Figure 4A) and α_1-adrenergic receptor ligands **7**, **9**, and **10** (Figure 4B). In both cases, a consensus or signature bit pattern is obtained by identifying those bit positions that are consistently set on within an activity class. As expected, these patterns differ significantly for both compound classes. Similar observations were made when calculating fingerprint profiles (capturing the relative occupancy of MFP bit positions) for different classes of enzyme inhibitors and receptor antagonists.[25] The angiotensin II antagonist consensus pattern is stronger than the one of α_1-adrenergic receptor ligands, consistent with the fact that angiotensin II antagonists are more similar to each other. When a scaling factor of 2 is applied to all signature bits, the incidence of these particular bits is effectively doubled when the Tanimoto coefficient is calculated. Hence, if the scaled Tanimoto coefficient (sTc) is calculated for a pair of MFPs that contain consensus bits, then sTc will be greater than the (unscaled) Tc. Larger scaling factors increase the magnitude of these effects. Thus, "difficult" similarity relationships may be easier to detect with scaled MFPs, as illustrated in Figure 4B. For example, in a database search,

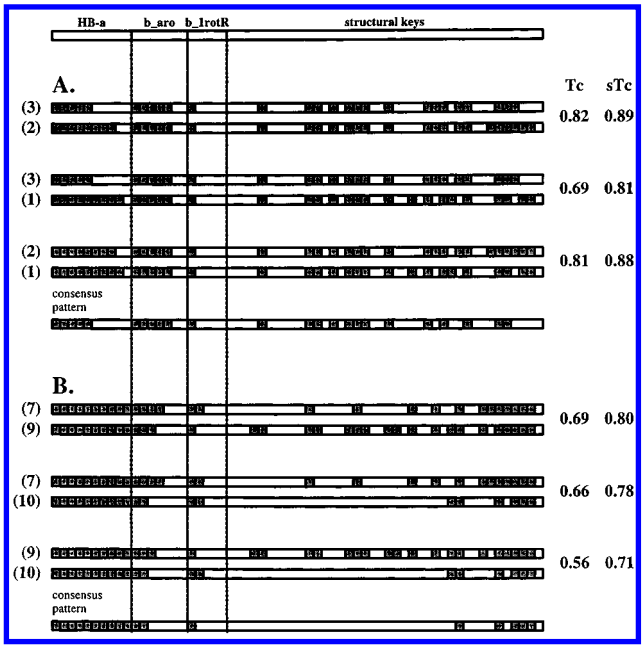


Figure 4. MFP comparison and consensus patterns. Bit settings of MFP2 (see also Figure 1) are compared for angiotensin II antagonists (A) and α_1-adrenergic receptor ligands (B). Compounds are numbered according to Tables 1 and 2. For each ligand class, the consensus bit pattern (or signature motif) is shown. Tc values are reported for pairwise comparison, and corresponding scaled Tc (sTc) values are obtained by applying a scaling factor of 2 to the consensus bits of each class.

the similarity of compounds **9** and **10** would only be detected with MFP1 but not MFP2, since the latter Tc value is slightly lower (0.56) than our similarity threshold value (0.60). However, when the signature motif is taken into account and scaled, the similarity of compounds **9** and **10** is recognized, since an sTc value of 0.71 is obtained. Therefore, if meaningful signature motifs can be identified, MFP scaling is a useful modification of the approach for virtual screening applications.

Implications for Similarity Analysis. Our calculations reveal that 2D and 3D fingerprints of very different design and complexity can produce similar results in the study of difficult test cases. These examples have been selected since their successful identification was previously thought to be possible only by applications of 3D pharmacophore fingerprints. Thus, the comparisons made provide some evidence that increasing complexity of fingerprint representations does not necessarily correlate with better performance in the analysis of molecular similarity. The findings also suggest that combinations of a limited number of relatively simple descriptors are sufficient, in many cases, to capture and distinguish molecular features responsible for specific activities. It is a critical task to identify those descriptors, regardless of their complexity, that are best suited to capture and differentiate molecular features that determine a specific activity. Moreover, even for specific test cases, it is difficult to conclude that complex 3D pharmacophore fingerprints are in principle superior to simpler 2D approaches. It follows that generally preferred approaches to the analysis of molecular similarity may be difficult to establish.

CONCLUSIONS

In this study, we have analyzed the structural and functional similarity of molecules that were previously presented as challenging test cases for conventional 2D methods. However, in our calculations, MFPs and, to some extent, MACCS keys successfully recognized a number of these relationships. These findings have implications for the study of molecular similarity and are expected to add to the evaluation of methods in this area of computational research. In addition, we show, using MFPs as an example, that descriptor components dominating the detection of similarity can be easily determined in keyed fingerprints and that it is possible to identify signature bit patterns for molecules with specific activity. Scaling factors can be applied to put more weight on bits in such signature motifs, and we predict that this will increase the probability of identifying similar molecules by virtual screening.

REFERENCES AND NOTES

(1) Barnard, J. M. Substructure searching methods. Old and new. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 532−538.
(2) McGregor, M. J.; Pallai, P. V. Clustering of large databases of compounds: Using MDL "keys" as structural descriptors. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 443−448.
(3) James, C. A.; Weininger, D. Daylight fingerprint. *Daylight theory manual*; Daylight Chemical Information Systems, Inc.: Irvine, CA, 1995.
(4) Pickett, S. D.; Mason, J. S.; McLay, I. M. Diversity profiling and design using 3D pharmacophores: Pharmacophore-derived queries. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1214−1223.
(5) 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.
(6) Mason, J. S.; Hermsmeier, M. A. Diversity assessment. *Curr. Opin. Chem. Biol.* **1999**, *3*, 342−349.
(7) Gorse, D.; Lahana, R. Functional diversity of compound libraries. *Curr. Opin. Chem. Biol.* **2000**, *4*, 287−294.
(8) 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.
(9) 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.
(10) Matter, H. Selecting optimally diverse compounds from structure databases: A validation study of two-dimensional and three-dimensional molecular descriptors. *J. Med. Chem.* **1997**, *40*, 1219−1229.
(11) Matter, H.; Pötter, T. Comparing 3D pharmacophore triplets and 2D fingerprints for selecting diverse compound subsets. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1211−1225.
(12) Patterson, D. E.; Cramer, R. D., Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood behavior: A useful concept for validation of "molecular diversity" descriptors. *J. Med. Chem.* **1996**, *39*, 3049−3059.
(13) Mason, J. S.; Cheney, D. L. Library design and virtual screening using multiple 4-point pharmacophore fingerprints. *Pac. Symp. Biocomput.* **2000**, *5*, 576−587.
(14) Schuffenhauer, A.; Gillett, V. J.; Willett, P. Similarity searching in files of three-dimensional chemical structures: Analysis of the BIOSTER database using two-dimensional fingerprints and molecular field descriptors. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 295−307.
(15) Greene, J.; Kahn, S.; Savoj, H.; Sprague, P. Chemical functional queries for 3D database search. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1297−1308.
(16) 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.
(17) Bradley, E. K.; Beroza, P.; Penzotti, J. E.; Grootenhuis, P. D. J.; Spellmeyer, D. C.; Miller, J. L. A rapid computational method for lead evolution: Description and application to α_1-adrenergic antagonists. *J. Med. Chem.* **2000**, *43*, 2770−2774.
(18) Xue, L.; Godden, J.; Gao, H.; Bajorath, J. Identification of a preferred set of molecular descriptors for compound classification based on principal component analysis. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 699−704.
(19) Xue, L.; Bajorath, J. Molecular descriptors for effective classification of biologically active compounds based on principal component analysis identified by a genetic algorithm. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 801−809.
(20) Xue, L.; Godden, J.; Bajorath, J. Database searching for compounds with similar biological activity using short binary bit string representations of molecules. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 881−886.
(21) Xue, L.; Godden, J. W.; Bajorath, J. Evaluation of descriptors and mini-fingerprints for the identification of molecules with similar activity. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1227−1234.
(22) *MOE (Molecular Operating Environment)*, version 1999.05; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 1999.
(23) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.
(24) *MACCS keys*; MDL Information Systems Inc.: San Leandro, CA.
(25) Godden, J. W.; Xue, L.; Stahura, F. L.; Bajorath, J. Searching for molecules with similar biological activity: Analysis by fingerprint profiling. *Pac. Symp. Biocomput.* **2000**, *5*, 566−575.
(26) Mason, J. S.; Cheney, D. L. Ligand−receptor 3-D similarity studies using multiple 4-point pharmacophores. *Pac. Symp. Biocomput.* **1999**, *4*, 456−467.

CI000305X