Published paper

Gardiner, E.J., Holliday, J.D., O'Dowd, C. and Willett, P. (2011) *Effectiveness of 2D fingerprints for scaffold hopping.* Future Medicinal Chemistry, 3 (4). 405 - 414. ISSN 1756-8919

http://dx.doi.org/10.4155/FMC.11.4

Effectiveness of 2D Fingerprints for Scaffold Hopping

Eleanor J. Gardiner, John D. Holliday, Caroline O'Dowd and Peter Willett[1]
Information School, University of Sheffield,
Western Bank, Sheffield S10 2TN, UK

ABSTRACT

Background. It has been suggested that similarity searching using 2D fingerprints may not be suitable for scaffold hopping.

Methods. This paper reports a detailed evaluation of the effectiveness of six common types of 2D fingerprint when they are used for scaffold hopping similarity searches of MDDR, WOMBAT and MUV data.

Results. The results demonstrate that 2D fingerprints can be used for scaffold hopping, with novel scaffolds being identified in nearly every search that was carried out. The degree of enrichment depends on the structural diversity of the actives that are being sought, with the greatest enrichments often being obtained using ECFP4 fingerprints.

Conclusions. 2D fingerprints provide a simple, and computationally efficient, way of identifying novel chemotypes in lead-discovery programmes.

KEYWORDS

Enrichment. The ability of a virtual screening (*vide infra*) method to retrieve more bioactive molecules in a database search than would random selection.

Fingerprint. A compact representation of molecular structure widely used in chemoinformatics systems. A fingerprint is a (normally) binary vector string encoding the presence or absence of topological substructures in a molecule.

Scaffold hopping. A database search for molecules that have a different central ring system from that of an existing lead, such as might be used as the reference structure in a similarity search

Similarity search. Search of a chemical database that computes the similarity between each database structure and an input reference structure, and that returns as output the most similar molecules.

Virtual screening. A computational procedure that ranks the structures in a chemical database in order of decreasing probability of activity in a bioassay of interest.

[1] To whom all correspondence should be addressed: email p.willett@sheffield.ac.uk; telephone +44-114-2222633; fax +44-114-2780300

INTRODUCTION

Virtual screening plays an increasingly important role in the discovery of novel, bioactive molecules, such as pharmaceuticals and agrochemicals. We focus here on similarity searching, which is based on the idea that molecules that are structurally similar are likely to have similar properties [1-4]. It involves the identification of those molecules in a database that are most similar to a known active ligand (the *reference structure*) since these are expected to have the highest probabilities of exhibiting the activity of interest and will hence be candidates for biological testing.

At the heart of any similarity searching system is the similarity measure that forms the basis for the ranking of the database structures. A very common approach to determining the similarity between two molecules is a comparison of their corresponding *fingerprints*, where a fingerprint is a (normally) binary string encoding the presence or absence of topological substructures, or fragments, in a molecule. The comparison enables the subsequent calculation of the Tanimoto coefficient, which is based on the numbers of fragments common and non-common to the two fingerprints. This very simple approach to the quantification of molecular resemblances was first described some 25 years ago but is still very extensively used for ligand-based virtual screening since it has been found to be both efficient and effective in operation [1, 5].

Although widely used, the application of simple, topological fragments to the matching of two molecules has been criticised on the grounds that a similarity search is likely to retrieve only those molecules that have a close topological relationship to the reference structure. In particular, it has been suggested that 2D fingerprints are unlikely to be able to retrieve molecules that have a different central ring-system from that of the reference structure, i.e., that belong to different chemotypes. The ability to identify such novel ring-systems, an ability that is often referred to as *scaffold hopping* or *lead hopping*, is a key capability of an effective method for virtual screening [6-9]. Bohm *et al*. state that “The aim of scaffold hopping is to discover structurally novel compounds starting from known active compounds by modifying the central core structure of the molecule” [10]; and in like vein, Zhou defines scaffold hopping as “a computational technique that identifies a topologically different scaffold from the parent compound but with similar or improved activity and other properties from a given database” [11]. Scaffold hopping is important for three principal reasons: it can provide a back-up if the existing lead series in a project subsequently proves to exhibit poor ADMET properties; some positions around a given scaffold may be synthetically difficult, where as an alternative may enable the creation of the desired substitution pattern; most importantly, it provides a way of circumventing the structural coverage of an existing, competitor patent.

The use of 3D structural information lies at the heart of pharmacophore searching and docking, which are two of the most important current approaches to virtual screening, and many of the published methods for scaffold hopping use similarity measures that are based on 3D descriptors of various sorts [6]. This is entirely appropriate given that protein-ligand binding is an inherently 3D event. However, the geometry of a molecule is determined, in part at least, by its underlying topology, and it is hence of interest to study the extent to which such information, as encoded by 2D fingerprints, is applicable to virtual screening. We have previously reported a detailed study of the merits of several different 2D fingerprints for similarity-based virtual screening in general [12]; here we extend that work by considering the suitability of this popular structure representation for the specific task of scaffold hopping.

METHODS

The experiments simulated a typical virtual screening environment. A bioactive reference structure was matched against each database structure in turn, the similarities computed, and the database ranked in decreasing similarity order. A threshold was then applied – the top-1% of the ranking in our experiments – and then the effectiveness of the search determined on the basis of the activity, or otherwise, of the top-ranked molecules. This was possible since the three databases used here all contain data regarding the biological activity of their constituent structures. The three databases were the *MDL Drug Data Report* database (MDDR, from Accelrys Inc. [101]), the *World of Molecular Bioactivity* database (WOMBAT, from Sunset Molecular Discovery LLC [102]), and the Maximum Unbiased Validation database (MUV, from the Carolo-Wilhelmina Technical University in Braunschweig [103]). Full details of these three datasets are given by Gardiner *et al*. [13] and by Rohrer and Baumann [14].

Several activity classes were available for each of the three datasets, as listed in Table 1(a)-(c). Each row of the table contains an activity class, the number of molecules in the database belonging to that class, the number of scaffolds present in the molecules belonging to that class (which we shall refer to subsequently as the *active scaffolds*), and the structural diversity of that class, and we now describe the entries in the second and third columns of each table.

Brown and Jacoby [6] and Xu and Johnson [15] review a range of definitions that have appeared in the literature, with the former noting that the molecular frameworks first reported by Bemis and Murcko [16] have been widely adopted. A molecular framework is obtained by pruning all acyclic parts of a molecule, whilst maintaining the atom and bond types for the ring system, and we have used this definition here, specifically the implementation available in the Murcko Scaffold routine in the Pipeline Pilot software.

The diversity value for a class was computed as the mean similarity when averaged over all pairs of molecules in the class, using Tripos Unity 2D binary fingerprints and the Tanimoto coefficient. It will be seen that the MDDR (Table 1a) and WOMBAT (Table 1b) datasets have examples both of classes involving actives that are structurally homogeneous and of classes involving actives that are structurally heterogeneous (i.e., structurally diverse). The use of homogeneous sets of actives can give overly-optimistic results (referred to as *analogue bias*) as to the effectiveness of topologically-based virtual screening methods [17, 18]. MUV has been designed specifically to assess the performance of virtual screening methods when diverse sets of actives are being sought. Thus, while the MDDR and WOMBAT datasets used here contained fixed sets of 102,516 molecules and 138,127 molecules, respectively, the MUV dataset contained 15,000 carefully selected inactives to go with each set of 30 actives; moreover, these actives have been chosen to be structurally diverse, with a mean of only 1.16 examples for each scaffold present in the set of actives.

In each case, ten representative reference structures were chosen from an activity class using a MaxMin diversity selection routine to ensure coverage of the full range of structural types within each activity class. The conventional approach to measuring the effectiveness of a similarity search is by using some function of the *recall*, i.e., the percentage of the active molecules retrieved at some cut-off point in the ranking (for which we have used the top-1% in our experiments). Examples of such measures include enrichment factors, cumulative recall, and ROC curves. Here, however, we wished to focus on the active scaffolds, and we hence adopted three different criteria for the evaluation of search performance. The first criterion was the percentage of the active scaffolds identified in all the molecules (not just the active molecules) retrieved in the top-1% of a ranking. The second, more stringent criterion was the percentage of the active scaffolds identified in the active molecules retrieved in the top-1% of a ranking. The third criterion was the percentage of the active molecules in the top-1% of the ranking that had a scaffold different from that of the reference structure. All three criteria were used to evaluate the MDDR and WOMBAT searches, and only criterion-3 for the MUV searches given the manner in which this dataset has been constructed. The screening performance for each activity class was obtained by calculating first the appropriate criterion value for each search and then the arithmetic mean when averaged over the ten chosen reference structures for that class.

The focus of this study is the use of 2D fingerprint representations of molecular structure, where a fingerprint is a binary vector encoding the presence or absence of substructural fragments. There are two main ways in which a fingerprint can be generated [19, 20]. *Dictionary-based* approaches involve a pre-defined list of fragments, with normally one fragment allocated to each position in the vector. A molecule is checked for the presence of each of the fragments in the dictionary, and bits are then set (or not set) depending on the presence (or absence) of that fragment. *Molecule-based* approaches involve the use of hashing algorithms to allocate multiple fragments to each bit-position. A fragment definition is provided, e.g., all chains of five bonded non-hydrogen atoms, and the presence identified within a molecule of all such fragments matching the definition. Each of the resulting fragments is then hashed to set multiple bits in the fingerprint.

The experiments involved testing six different types of fingerprint that are available in widely used chemoinformatics systems: ECFP4 (for Extended Connectivity Fingerprint encoding circular substructures of diameter four bonds) fingerprints from the Pipeline Pilot software (hashed to a fixed length of 1024 bits, and available from Accelrys Inc. [101]); FCFP4 (for Functional-Class Fingerprint encoding circular substructures of diameter four bonds) fingerprints (1024 bits, also available in the Pipeline Pilot software); Tripos Unity fingerprints (988 bits, available from Tripos LP [104]), BCI fingerprints (1052 bits, available from Digital Chemistry [105]), Daylight fingerprints (2048 bits, available from Daylight Chemical Information Systems Inc. [106]) and MDL key fingerprints (166 bits, available from Accelrys Inc. [101]). The BCI and MDL fingerprints are dictionary-based, the Daylight and Pipeline Pilot fingerprints are molecule-based, and the Unity fingerprints employ both types of generation method. The use of these and other types of fingerprints for similarity-based virtual screening are described by Hert *et al.* [12], Sastry *et al.* [21] and Duan *et al.* [22].

It should be noted that any comparison of virtual screening methods is inherently complex given the many variables that can affect the results, such as the choice of structure representation, of reference structure, of biological activity, of similarity coefficient, and of weighting scheme *inter alia*. The methods used here have been chosen specifically to represent those most commonly encountered, in both the published literature and currently available chemoinformatics software. For example, Holliday *et al.* have discussed the use of 22 different similarity coefficients for the matching of chemical fingerprints [23]; however, despite the many alternatives and despite the known limitations of the Tanimoto coefficient [3], it is this coefficient that continues to be the most widely used. As another example, Duan *et al.* note that fingerprints can often be implemented in multiple ways, with their extensive comparison of similarity methods for virtual screening involving 11 different parameterisations of the atoms involved in each substructural fragment encoded in a fingerprint [22]; the comparison here has used two popular representations (ECFP4 and FCFP4) in the Pipeline Pilot software to exemplify the use of alternative approaches to atom-typing. Other factors that may affect the effectiveness of fingerprint implementations include: the length of the fingerprint that is used, especially if hashing techniques are employed that can result in substantial numbers of collisions [21]; and whether incidence or occurrence data is used, i.e., whether the fingerprint encodes merely the presence of a fragment, its frequency of occurrence, or some standardised form of the latter [24].

The results that are presented and discussed in the following section are hence typical of those that might be obtained using the default implementations of much current chemoinformatics software. Increases in retrieval effectiveness could certainly be obtained by appropriate choice of method and/or parameterisation for specific activity classes and/or reference structures [4, 22]; however, the training data necessary for such tuning is often unavailable in the early stages of a discovery project when similarity methods are most commonly employed.

RESULTS AND DISCUSSION

The results of the searches, using all three evaluation criteria for the MDDR and WOMBAT datasets and criterion-3 for the MUV dataset, are shown in Tables 2-4. Each column in each of the tables is headed by an abbreviation of the name of the activity class as detailed in Table 1. The best performance in each column has been bold-faced and italicised for ease of identification.

The first, and most important, general conclusion that can be drawn from the data is that 2D fingerprints are indeed capable of being used for scaffold hopping searches. The figures listed in these tables are the percentages of active molecules or active scaffolds retrieved in just the top-1% of the ranking. If a search had been no better than random then one would expect this top-1% to contain approximately 1% of the active molecules or scaffolds; figures greater than that percentage hence demonstrate the ability of the screening method to enrich the search outputs, as compared to random selection. Inspection of the tables shows clearly that some degree of enrichment is obtained in all cases, with the sole exception of the MUV SF1A searches where sub-random performance is observed; in some cases there is a very considerable degree of enrichment indeed. Examples of some of the scaffold hops that were identified are shown in Figure 1, these being for MDDRD searches using reference structures in the HIV and 5HT1A activity classes. A second conclusion, and one that might be expected, is that the degree of enrichment decreases as one moves from criterion-1 to criterion-3 and as the strictness of the retrieval criterion is increased. Thirdly, there is a considerable degree of variation in the recall obtained with the different activity classes (and also a considerable degree of variation in the recall obtained using different reference structures for the same activity class [4]). One obvious factor that affects the recall is the diversity of an activity class: one would expect that a search for a class that is tightly clustered in chemical space would retrieve more actives than it would when those actives are more widely dispersed. This is observed to some extent in practice. For example, the best results in the MDDR and WOMBAT datasets were normally obtained in the renin searches, and this is the most homogeneous class; excellent results are also obtained for the PKC searches of WOMBAT, where this is the second most homogeneous class. As another example, the results for the MUV dataset are noticeably inferior to those for the other two datasets, in line with the fact that this dataset has been carefully designed to involve only highly diverse sets of actives. That said, there are obvious discrepancies; for example, the MDDR D2 searches give noticeably worse results than do the corresponding COX searches, despite the latter class being more diverse.

A further conclusion from the data in Tables 2-4 is that there is a considerable degree of variation in the recall obtained with the different fingerprints, with the Pipeline Pilot circular fingerprints (ECFP4 or sometimes FCFP4) generally giving the best performance using all three criteria for the MDDR and WOMBAT datasets. The significance, if any, of the differences in performance was tested with Kendall's *W* test of statistical significance, which is used to evaluate the consistency of *k* different sets of ranked judgements of the same set of *N* different objects. Specifically, each of the activity classes was considered as a judge ranking the different similarity measures in order of decreasing recall. Thus, for the MDDR searches, the eleven activity classes were used to rank the six fingerprints, so that k=11 and N=6. Converting the recall values in each sub-table to ranks one can then compute *W* and test the significance of the observed value using tables provided by Siegel and Castellan [25]. If a statistically significant value is obtained then these authors suggest that the best overall ranking of the *N* objects can be obtained using their mean ranks when averaged over the *k* judges. The *W* values (all statistically significant at $p <= 0.01$) and resulting mean ranks for the MDDR and WOMBAT searches are listed in Table 5 where, as before, the best performance is bold-faced and italicised. Two recent comparisons of 2D fingerprints have shown the general effectiveness of circular substructures for similarity applications [21, 22], and the results in the table show that this is also the case here for the ECFP4 and FCFP4 substructures [26] when used with the MDDR and WOMBAT datasets. The MUV searches have been excluded from Table 5 as the *W* value of 0.03 was not statistically significant at the 0.01 level, i.e., there was no degree of consistency in the ranking of the different fingerprints. Similar comments apply to the MDDR and WOMBAT datasets if attention is restricted to just those activity classes with mean similarities < 0.40 in Tables 1a and 1b.

While this article was being prepared for publication, we became aware of a very recent report by Vogt *et al*. that is closely related to our work [27]. They used sets of known actives for 17 biological targets, adding these to ZINC and ChEMBL datasets each containing ca. 500K presumed inactives and then carrying out scaffold hopping searches that were evaluated in a manner similar to our criterion-2. They used five different types of fingerprint, including the MDL and ECFP4 fingerprints employed in our experiments, and carried out each search using a 1-NN (or group fusion) strategy. Here, five different actives that shared a common scaffold were taken in turn as the reference structure, and then each database structure ranked on the basis of the largest similarity with any of these five reference structures. They found that retrieving the top-1% of a ranking resulted in the retrieval of 30-40% of the active Murcko scaffolds using the most effective fingerprints, with lower recall levels being achieved when less precise ring definitions were employed. They also reported extensive studies of the precise similarity values required for adequate scaffold retrieval with the different types of fingerprint, and noted that the best results were generally obtained with the ECFP4 fingerprint. It is not possible to compare the two sets of results directly owing to the different data and search strategies that were employed; however, the two reports are at one in concluding that 2D fingerprints can indeed be used for scaffold hopping searches.

FUTURE PERSPECTIVE

Similarity searching using 2D fingerprints is a simple approach to virtual screening that is very widely used. It has, however, been criticised on two, related grounds. First, it has been suggested that the topological nature of the fragment substructures encoded in a fingerprint will mean that such searches are unlikely to exhibit any significant capacity for scaffold hopping; second, that the apparently good results obtained by 2D fingerprints in retrospective studies of screening performance have been due to analogue bias in the databases that are being searched. In this paper, we have reported similarity-based virtual screening searches that have been evaluated so as to minimise the effects of analogue bias. Our results show clearly that at least some of the fingerprints studied here can be used in scaffold hopping applications, even when structurally diverse sets of actives are sought. The enrichments are often not large but they are consistently superior to those obtained from random screening. Given the simplicity of the approach and its very limited computational requirements, we conclude that 2D fingerprints provide a viable way of scanning a database for novel scaffolds.

That said, the use of 3D methods provides an alternative, and potentially complementary, source of information. An obvious approach is hence to combine the outputs of different scaffold-hopping methods that exploit both types of information. This can be done by applying data fusion to ranked search outputs [28] or, if appropriate training data are available, by using belief theory [29]. Thus, Muchmore *et al*. have described the use of the latter approach on Abbott internal data, and demonstrated that effective scaffold-hopping can be achieved by combining 2D and 3D similarity measures (specifically fingerprints encoding ECFP6 circular substructures, analogous to, but larger than, the ECFP4 ones used here, and the ROCS shape-similarity software from OpenEye Scientific [107]). We believe that such combined approaches are likely to become increasingly important as computational costs continue to fall (making multiple searches increasingly feasible) and as new virtual screening methods are developed.

ACKNOWLEDGEMENTS.

We thank Accelrys Software Inc., Digital Chemistry Ltd., Sunset Molecular Discovery LLC, and Tripos Inc. for software and data, and the Royal Society and the Wolfson Foundation for laboratory support.

EXECUTIVE SUMMARY

- Similarity searching using 2D fingerprints is a simple approach to virtual screening that is very widely used. However, the focus on the 2D nature of molecules might suggest that it would not be appropriate for scaffold-hopping applications
- Experiments with three standard datasets and six widely used fingerprints show that this is not the case, with enrichments consistently superior to those obtained from random screening.
- The best results are obtained with fingerprints based on circular substructures

FINANCIAL AND COMPETING INTEREST DISCLOSURE

The authors declare no conflicting interests, and have not received writing assistance in the preparation of this article.

REFERENCES

1. Eckert H, Bajorath J: Molecular similarity analysis in virtual screening: foundations, limitation and novel approaches. *Drug Discov. Today* 12, 225-233 (2007).
2. Glen RC, Adams SE: Similarity metrics and descriptor spaces - which combinations to choose? *QSAR Combin. Sci*. 25, 1133-1142 (2006).
3. Willett P: Similarity methods in chemoinformatics. *Ann. Review Inf. Sci. Technol*. 43, 3-71 (2009). (* The most recent comprehensive review of chemoinformatics applications of molecular similarity)

4. Sheridan RP: Chemical similarity searches: when is complexity justified? *Expert Opin. Drug Discov*. 2, 423-430 (2007). (** An excellent critical analysis of similarity searching approaches, both simple and complex)

5. Willett P: Similarity-based virtual screening using 2D fingerprints. *Drug Discov. Today*, 11, 1046-1053 (2006).

6. Brown N, Jacoby E: On scaffolds and hopping in medicinal chemistry. *Mini-Rev. Med. Chem*. 6, 1217-1229 (2006). (** An excellent comprehensive review of scaffold hopping techniques)

7. Martin YC, Muchmore S: Beyond QSAR: lead hopping to different structures. *QSAR Combin. Sci.* 28, 797-801 (2009).

8. Renner S, Schneider G: Scaffold hopping potential of ligand-based similarity coencepts. *ChemMedChem* 1:181-185 (2006).

9. Schneider G, Schneider P, Renner S: Scaffold-hopping: how far can you jump? *QSAR Combin. Sci.* 25, 1162-1171 (2006).

10. Böhm H-J, Flohr A, Stahl M: Scaffold hopping. *Drug Discov. Today: Technologies* 1, 217-224 (2004).

11. Zhou H: Scaffold selection and scaffold hopping in lead generation: a medicinal chemistry approach. *Drug Discov. Today* 12, 149-154 (2006).

12. Hert J, Willett P, Wilton DJ, *et al*.: Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org. Biomol. Chem*. 2, 3256-3266 (2004).

13. Gardiner EJ, Gillet VJ, Haranczyk M, *et al*.: Turbo similarity searching: Effect of fingerprint and dataset on virtual-screening performance. *Stat. Anal. Data Mining* 2, 103-114 (2009).

14. Rohrer SG, Baumann K: Maximum Unbiased Validation (MUV) data sets for virtual screening based on PubChem bioactivity data. *J. Chem. Inf. Model*. 49, 169-184 (2009).

15. Xu Y-J, Johnson, M: Using molecular equivalence numbers to visually explore structural features that distinguish chemical libraries. *J. Chem. Inf. Comp. Sci*. 42, 912-926 (2002).

16. Bemis GW, Murcko MA: The properties of known drugs, 1. Molecular frameworks. *J. Med. Chem*. 39, 2887-2893 (1996). (* A widely used definition of a scaffold, and that used here in the work reported here)

17. Cleves AE, Jain AN: Effects of inductive bias on computational evaluations of ligand-based modeling and on drug discovery. *J. Comput.-Aided Mol. Design* 22, 147-159 (2008).

18. Good AC, Oprea TI: Optimization of CAMD techniques. 3. Virtual screening enrichment studies: a help or hindrance in tool selection? *J. Comput.-Aided Mol. Design* 22, 169-178 (2008).

19. Gasteiger J, Engel T (eds.): *Chemoinformatics: A Textbook*. Weinheim, Wiley-VCH (2003).

20. Leach AR, Gillet VJ: *An Introduction to Chemoinformatics* 2nd edition. Dordrecht: Kluwer (2007).

21. Sastry M, Lowrie JF, Dixon SL, *et al.*: Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. *J. Chem. Inf. Model*. 50, 771-748 (2010).

22. Duan J, Dixon SL, Lowrie JE, Sherman W: Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. *J. Mol. Graph. Model*. 29, 157-170 (2010).

23. Holliday JD, Hu C-Y, Willett P: Grouping of coefficients for the calculation of inter-molecular similarity and dissimilarity using 2D fragment bit-strings. *Combin. Chem. High-Throughput Screen*. 5, 155-166 (2002).

24. Arif SM, Holliday JD, Willett P: Analysis and use of fragment occurrence data in similarity-based virtual screening. *J. Com.-Aided Mol. Design* 23, 655-668 (2009).

25. Siegel S, Castellan NJ: *Nonparametric Statistics for the Behavioural Sciences* 2nd Edition. New York: McGraw-Hill (1988).

26. Rogers D, Hahn M: Extended-connectivity fingerprints. *J. Chem. Inf. Model*. 50, 742-754 (2010).

27. Vogt M, Stumpfe D, Geppert H, *et al.*: Scaffold hopping using two-dimensional fingerprints: true potential, black magic, or a hopeless endeavor? Guidelines for virtual screening. *J. Med. Chem*. 53, 5707-5715 (2010). (** A study that is very similar in design to that reported here)

28. Willett P: Enhancing the effectiveness of ligand-based virtual screening using data fusion. *QSAR Combin. Sci.* 25, 1143-1152 (2006).

29. Muchmore SW, Debe DA, Metz JT, Brown SP, Martin YC, Hajduk PJ: Application of belief theory to similarity data fusion for use in analog searching and lead hopping. *J. Chem. Inf. Model*. 48, 941-948 (2008). (* An elegant approach to the combination of evidence for virtual screening)

WEBSITES

101. Accelrys Inc. at http://www.accelrys.com

102. Sunset Molecular Discovery LLC at http://sunsetmolecular.com/

103. Maximum Unbiased Validation database at http://www.pharmchem.tu-bs.de/lehre/baumann/MUV.html

104. Tripos LP [104] at http://www.tripos.com

105. Digital Chemistry at http://www.digitalchemistry.co.uk

106. Daylight Chemical Information Systems Inc. at http://www.daylight.com

107. OpenEye Scientific Software Inc. at http://www.eyesopen.com/

CAPTIONS

Figure caption.

Figure 1. Active scaffolds, together with ECFP4 Tanimoto similarities, retrieved in the top-1% of the rankings in searches of the MDDR database for the HIV (a) and 5HT1A (b) activity classes using the reference structure in the centre of the figure.

Table captions

Table 1 Activity classes used in virtual screening with (a) MDDR, (b), WOMBAT and (c) MUV. The mean similarity for each class here was obtained by averaging over all pairs of molecules in the class, using Tripos Unity 2D binary fingerprints and the Tanimoto coefficient.

Table 2. Scaffold-hopping searches on the MDDR dataset using (a) criterion-1, (b) criterion-2 and (c) criterion-3

Table 3. Scaffold-hopping searches on the WOMBAT dataset using (a) criterion-1, (b) criterion-2 and (c) criterion-3

Table 4. Scaffold-hopping searches on the MUV dataset using criterion-3

Table 5 Kendall *W* analyses for MDDR and WOMBAT searches using criteria 1-3

Activity class	Active molecules	Active scaffolds	Mean similarity
5HT1A agonists	827	271	0.34
5HT3 antagonists	752	237	0.35
5HT reuptake inhibitors	359	126	0.35
Angiotensin II AT1 antagonists	943	285	0.40
Cyclooxygenase inhibitors	636	139	0.27
D2 antagonists	395	187	0.35
HIV protease inhibitors	750	331	0.45
Protein kinase C inhibitors	453	134	0.32
Renin inhibitors	1125	339	0.57
Substance P antagonists	1246	380	0.40
Thrombin inhibitors	803	295	0.42

(a)

Activity class	Active molecules	Active scaffolds	Mean similarity
5HT1A antagonists	592	135	0.40
5HT3 antagonists	220	68	0.38
Acetylcholine esterase inhibitors	503	150	0.37
Angiotensin II AT1 antagonists	724	154	0.44
Cyclooxygenase inhibitors	965	93	0.32
D2 antagonists	910	191	0.37
Factor Xa inhibitors	842	181	0.39
HIV protease inhibitors	1128	314	0.44
Matrix metalloprotease inhibitors	694	164	0.44
Phosphodiesterase inhibitors	596	161	0.36
Protein kinase C inhibitors	142	23	0.57
Renin inhibitors	474	124	0.59
Substance P antagonists	558	110	0.43
Thrombin inhibitors	421	138	0.42

(b)

Activity class	Active molecules	Active scaffolds	Mean similarity
S1P1 receptor agonists	30	28	0.29
PKA inhibitors	30	27	0.29
SF1 inhibitors	30	24	0.29
Rho-Kinase2 inhibitors	30	27	0.27
HIV RT-RNase inhibitors	30	27	0.26
Eph receptor A4 inhibitors	30	29	0.27
SF1 agonists	30	30	0.25
HSP 90 inhibitors	30	27	0.26
ER-a-Coactivator binding inhibitors	30	26	0.26
ER-β-Coactivator binding inhibitors	30	28	0.27
ER-a-Coactivator binding potentiators	30	28	0.30
FAK inhibitors	30	28	0.28
Cathepsin G inhibitors	30	28	0.32
FXIa inhibitors	30	21	0.28
FXIIa inhibitors	30	24	0.30
D1 receptor allosteric modulators	30	24	0.25
M1 receptor allosteric inhibitors	30	29	0.28

(c)

Table 1

	5HT1A	5HT3	5HTRE	ANGIO	COX2	D2	HIVP	PKC	RENIN	SUBP	THROM
BCI	18.11	15.30	18.00	22.39	20.29	9.73	9.12	10.83	20.89	6.25	10.61
Daylight	21.04	13.81	22.08	17.89	21.52	11.94	12.33	10.23	19.14	8.34	14.08
ECFP4	21.00	***20.81***	22.80	***26.34***	***28.77***	14.46	***15.06***	***16.32***	***33.88***	***10.95***	18.71
FCFP4	***23.59***	18.69	***23.60***	17.92	28.26	***15.54***	13.73	***16.32***	26.01	9.29	***19.97***
MDL	18.63	16.36	21.44	19.12	28.20	13.98	11.06	12.78	18.08	9.37	15.14
Unity	22.11	16.23	23.20	17.57	23.77	12.37	12.15	11.50	22.54	10.26	16.50

(a)

	5HT1A	5HT3	5HTRE	ANGIO	COX2	D2	HIVP	PKC	RENIN	SUBP	THROM
BCI	10.93	10.34	8.24	19.26	6.45	3.92	6.27	5.34	17.99	4.49	6.56
Daylight	12.22	8.90	***10.24***	15.18	4.78	4.78	8.67	4.21	15.83	6.62	7.38
ECFP4	12.78	***15.38***	8.24	***23.24***	***8.12***	4.30	***11.09***	***6.17***	***31.33***	***8.50***	11.12
FCFP4	***14.07***	13.01	9.60	15.46	7.17	***6.02***	9.33	***6.17***	22.84	6.65	***12.31***
MDL	9.89	11.31	7.76	15.07	7.08	5.32	7.12	4.74	15.21	6.94	9.05
Unity	12.22	11.31	9.60	14.79	5.22	4.25	8.33	5.34	19.38	8.05	9.90

(b)

	5HT1A	5HT3	5HTRE	ANGIO	COX2	D2	HIVP	PKC	RENIN	SUBP	THROM
BCI	8.07	8.30	5.57	15.55	3.87	3.92	4.86	4.10	16.59	2.76	5.73
Daylight	8.62	6.14	7.31	13.03	2.82	4.78	6.79	3.09	12.89	5.20	6.13
ECFP4	10.06	***13.29***	5.27	***20.21***	***4.19***	4.30	***9.49***	***5.67***	***32.76***	***6.66***	10.08
FCFP4	***10.58***	10.15	***7.54***	13.60	3.58	***6.02***	7.32	4.04	21.54	5.30	***10.28***
MDL	6.58	7.68	4.74	11.75	3.66	5.32	5.41	2.74	11.34	4.08	6.37
Unity	7.70	7.91	7.13	11.91	2.48	4.25	6.23	4.70	15.62	6.09	8.15

(c)

Table 2.

	5HT1A	5HT3	AChE	ANGIO	COX	D2	FXA	HIVP	MMP1	PDE4	PKC	RENIN	SUBP	THROM
BCI	18.21	26.42	17.92	29.28	25.00	17.16	14.56	14.19	25.58	16.56	45.45	29.67	12.11	11.75
Daylight	18.58	18.66	18.39	30.78	24.78	17.37	16.17	13.10	25.34	17.19	45.45	30.00	17.89	14.67
ECFP4	23.73	***28.36***	***22.21***	***37.39***	***35.98***	21.37	***24.17***	***19.39***	33.74	***21.25***	***60.00***	***58.54***	***19.27***	***23.65***
FCFP4	***26.87***	25.67	19.53	30.46	32.61	***22.53***	19.33	18.82	30.37	20.31	52.27	49.35	17.16	21.39
MDL	20.60	26.12	22.01	22.68	31.20	18.89	19.11	14.15	***36.63***	16.13	44.09	28.54	15.41	17.37
Unity	19.93	19.40	19.87	26.86	27.17	19.11	16.50	16.17	30.80	18.63	46.36	39.35	19.17	17.30

(a)

	5HT1A	5HT3	AChE	ANGIO	COX	D2	FXA	HIVP	MMP1	PDE4	PKC	RENIN	SUBP	THROM
BCI	9.93	***18.51***	10.60	27.25	11.96	8.89	10.50	10.06	14.85	9.94	43.64	26.02	9.36	6.64
Daylight	9.33	11.19	10.13	29.67	9.35	7.16	11.72	8.27	13.56	11.69	43.64	25.61	14.04	8.91
ECFP4	13.21	16.42	10.07	***34.77***	***14.02***	10.05	***17.78***	***12.88***	16.93	12.25	***52.73***	***55.61***	***14.95***	***17.08***
FCFP4	***16.49***	13.73	8.72	28.69	12.72	***10.26***	14.22	12.59	15.89	***12.63***	45.00	45.69	12.94	13.80
MDL	11.12	16.57	***10.81***	19.67	9.89	7.53	14.61	8.91	***21.60***	8.56	39.55	24.07	11.83	10.51
Unity	11.49	10.90	10.47	24.97	10.43	7.79	12.33	11.25	18.28	12.25	44.09	34.31	14.86	10.73

(b)

	5HT1A	5HT3	AChE	ANGIO	COX	D2	FXA	HIVP	MMP1	PDE4	PKC	RENIN	SUBP	THROM
BCI	6.69	***17.76***	***9.74***	26.52	6.47	6.57	8.53	7.94	11.03	8.01	40.95	20.83	6.46	3.54
Daylight	8.74	10.77	8.59	29.27	3.50	5.23	9.77	6.67	8.81	9.75	41.76	20.98	10.31	5.53
ECFP4	9.06	11.63	7.26	***33.57***	***8.58***	**6.88**	***14.64***	***10.69***	11.68	8.84	***48.71***	***55.23***	***11.45***	***14.23***
FCFP4	***12.04***	10.61	6.80	28.23	6.12	6.65	11.96	9.68	10.92	***9.76***	41.39	41.59	9.35	10.21
MDL	6.77	13.09	6.99	16.07	4.54	4.77	9.88	5.92	***16.84***	5.27	29.32	16.76	8.28	6.78
Unity	9.10	10.76	7.44	23.01	4.05	5.44	9.57	8.36	12.80	9.19	41.83	25.51	11.00	6.60

(c)

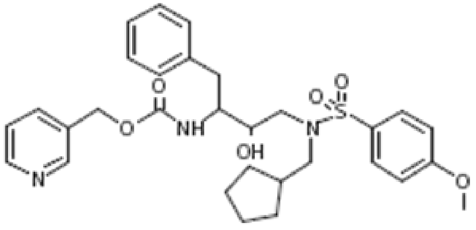
Table 3.

	S1P1	PKA	SF1I	RK2	HIV	ERA4	SF1A	HSP	ERaI	ERbI	ERaP	FAK	CG	FX1a	FXIIa	D1	M1
BCI	2.67	8.67	2.33	4.00	1.33	***6.00***	***0.67***	***12.67***	***7.33***	***5.00***	1.33	3.33	12.00	7.00	***14.67***	2.33	2.33
Daylight	3.00	9.33	3.33	3.67	1.33	5.33	0.33	2.67	2.00	3.00	2.33	4.00	15.33	5.00	***14.67***	2.00	2.33
ECFP4	***3.33***	9.00	3.00	5.33	2.00	4.67	***0.67***	4.67	1.67	4.33	1.67	2.67	9.33	***9.33***	10.67	2.00	1.33
FCFP4	***3.33***	10.33	2.67	***7.33***	***2.33***	4.33	***0.67***	4.33	2.00	3.67	2.33	2.67	11.67	8.67	13.00	2.00	***2.67***
MDL	2.00	3.67	2.67	4.67	***2.33***	3.67	***0.67***	4.00	2.67	4.00	***5.00***	***4.33***	11.00	5.00	11.67	2.67	1.33
Unity	3.00	***10.67***	***4.00***	3.33	1.67	4.00	0.33	3.67	1.33	4.00	2.00	3.33	***16.67***	4.67	13.00	***3.00***	***2.67***

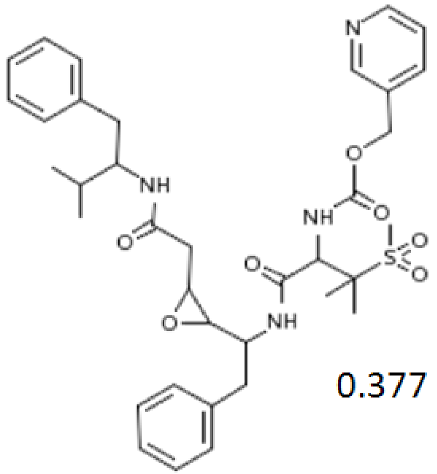
Table 4.

	MDDR			WOMBAT		
	1	2	3	1	2	3
BCI	5.27	4.72	4.09	5.07	4.18	4.14
Daylight	4.72	4.32	4.09	4.71	4.82	3.93
ECFP4	***1.68***	***1.82***	***1.82***	***1.21***	***1.82***	***1.86***
FCFP4	1.95	2.00	2.18	2.50	2.64	3.07
MDL	3.91	4.41	4.90	4.00	4.14	4.57
Unity	3.45	3.72	3.90	3.50	3.39	3.15
Kendall *W*	0.60	0.45	0.42	0.59	0.36	0.26

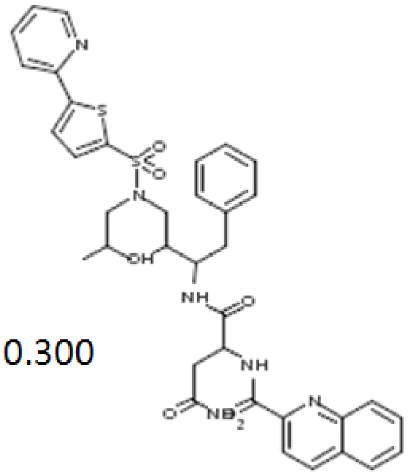
Table 5



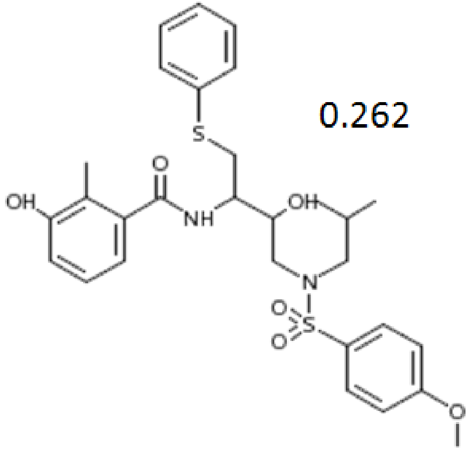
0.516



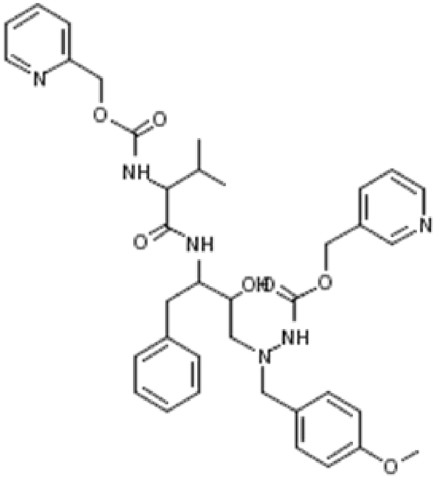
0.377

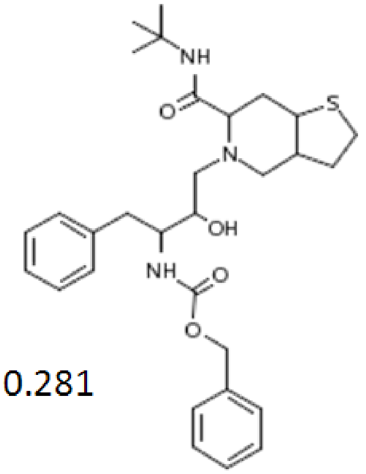


0.300



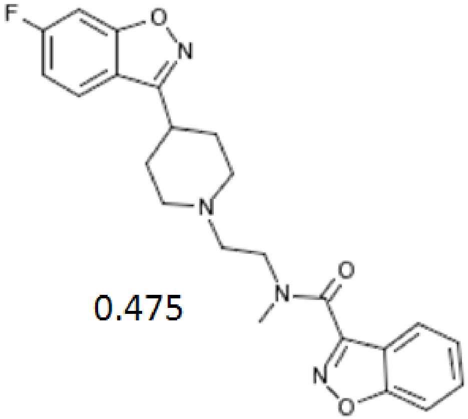
0.262



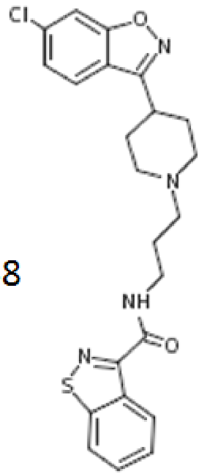


0.281

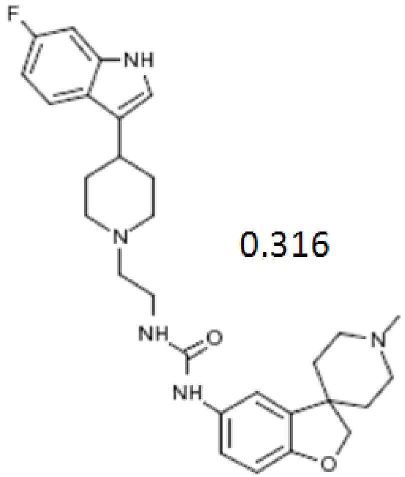
Figure 1a



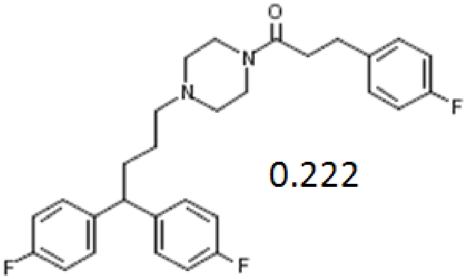
0.475



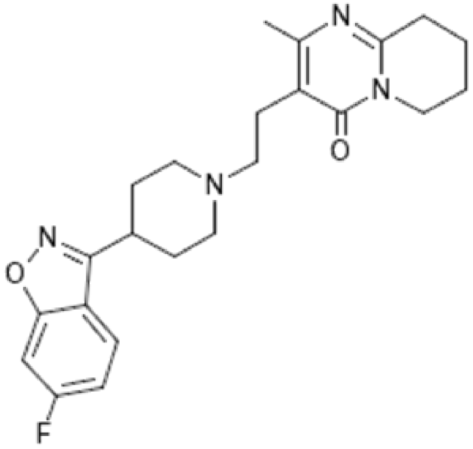
0.348

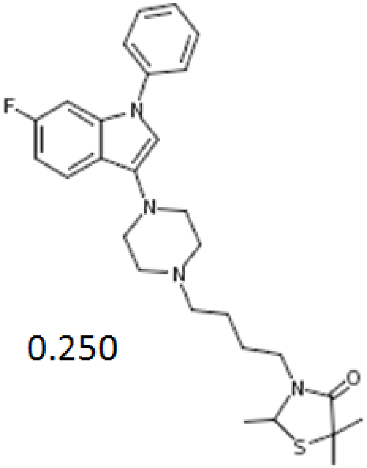


0.316



0.222





0.250

Figure 1b