JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Combinatorial Consensus Scoring for Ligand-Based Virtual Fragment Screening: A Comparative Case Study for Serotonin 5-HT$_3$A, Histamine H$_1$, and Histamine H$_4$ Receptors

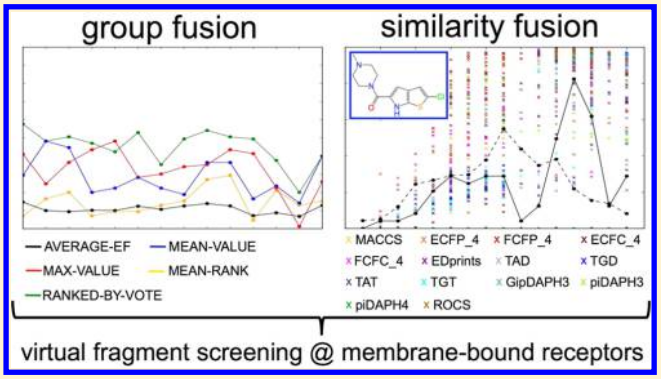
Sabine Schultes,[†,‡] Albert J. Kooistra,[†] Henry F. Vischer,[†] Saskia Nijmeijer,[†] Eric E. J. Haaksma,[†,‡] Rob Leurs,[†] Iwan J. P. de Esch,[†] and Chris de Graaf*[,†]

[†]Division of Medicinal Chemistry, Faculty of Sciences, Amsterdam Institute for Molecules, Medicines and Systems (AIMMS), VU University Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

[‡]Department of Medicinal Chemistry, Boehringer Ingelheim RCV GmbH & Co KG, Dr Boehringer-Gasse 5-11, 1121 Vienna, Austria

Supporting Information

ABSTRACT: In the current study we have evaluated the applicability of ligand-based virtual screening (LBVS) methods for the identification of small fragment-like biologically active molecules using different similarity descriptors and different consensus scoring approaches. For this purpose, we have evaluated the performance of 14 chemical similarity descriptors in retrospective virtual screening studies to discriminate fragment-like ligands of three membrane-bound receptors from fragments that are experimentally determined to have *no* affinity for these proteins (*true inactives*). We used a complete fragment affinity data set of experimentally determined ligands and inactives for two G protein-coupled receptors (GPCRs), the histamine H$_1$ receptor (H$_1$R) and the histamine H$_4$ receptor (H$_4$R), and one ligand-gated ion channel (LGIC), the serotonin receptor (5-HT$_3$AR), to validate our retrospective virtual screening studies. We have exhaustively tested consensus scoring strategies that combine the results of multiple actives (group fusion) or combine different similarity descriptors (similarity fusion), and for the first time systematically evaluated different *combinations* of *group* fusion and *similarity* fusion approaches. Our studies show that for these three case study protein targets both consensus scoring approaches can increase virtual screening enrichments compared to single chemical similarity search methods. Our cheminformatics analyses recommend to use a combination of both group fusion and similarity fusion for prospective ligand-based virtual fragment screening.



INTRODUCTION

Fragment-based drug discovery (FBDD) is a promising approach in the identification of novel, chemically tractable leads for the development of new therapeutic agents.[1−4] The key advantage of FBDD is that the respective chemical space described by low molecular weight compounds is much smaller than for drug-like compounds.[5−7] In addition, the lower complexity of small fragments is believed to increase the chance of a good match between ligand and protein.[8] However, the experimental identification of fragment hits is often more challenging than for drug-like compounds.[9] Fragments have in general lower affinities for their targets and therefore need special screening methods.[10] Virtual screening has evolved as an integral part in the drug discovery process to select compounds for experimental testing.[11] There are two general virtual screening methods, structure-based virtual screening (SBVS) and ligand-based virtual screening (LBVS). SBVS is frequently applied when a suitable structural model of the target is available. Ligand-based techniques provide alternative and complementary approaches particularly when the target structure is not available but active ligand molecules have been identified.[12] The applicability of virtual screening for the discovery of small fragment-like molecules is still a relatively unexplored research area for structure-based[13−18] and ligand-based[15,19] approaches. LBVS is based on the chemical similarity to one or more known active ligands of a specific protein target.[20] Information of the ligand can be encoded in a bit string representation, called a molecular fingerprint that can be used for comparison.[21] Such fingerprints can be classified into four different classes: binary (presence/absence) circular fingerprints, circular fingerprints considering counts, path-based and keyed fingerprints, and pharmacophore-based fingerprints.[22−24] Other methods compare the overlay of the 3D structures of two molecules,[25,26] or use molecular graph approaches[27,28] or feature trees[29,30] to determine the similarity between compounds.[31]

As LBVS methods are based on the similarity to known actives the challenge is to find novel structures.[20] In the context of

Received: November 20, 2014
Published: March 27, 2015

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044

ligand-based virtual *fragment* screening the size dependence of similarity measurements bears an additional challenge.[21,32] Moreover there is a lack of proper test sets for the evaluation of virtual fragment screening approaches. The low affinities of fragments provide a challenge for detection of fragment actives. Therefore, the number of known fragment actives is often very low. In the absence of true, experimentally determined, inactive molecules, *decoy* molecules (randomly selected molecules that have not been experimentally tested on a specific protein target) can be selected for the retrospective validation of virtual screening methods. The are however several caveats in the selection of active ligands and decoy sets for retrospective virtual screening studies: (i) decoys should not be distinguishable from actives by very simple molecular properties like the number of heavy atoms, H-bond donors, H-bond acceptors, and hydrophobicity (*artificial enrichment*),[33−35] (ii) actives should not be too similar/all share the same chemical scaffold (*analogue bias*),[35,36] (iii) the number of actives and ratio of active vs decoy molecules[33,37] should be representative for a real life screen. All these factors can have large effects the relative performance of ligand-based virtual screening approaches for different protein targets.[33−37] In dedicated benchmark data sets such as DUD[38]/DUD-E,[39] MUV,[40] and DEKOIS,[41] the decoys are selected in such a way that they resemble the molecular properties of the actives of specific targets. These benchmark data sets have been frequently used for the retrospective evaluation of mainly structure-based (but also ligand-based) virtual screening approaches.[38−45] However, the selection of decoys always bears the problem that assumed inactives may actually be actives,[46] and the ratio of actives vs decoys in designed virtual screening benchmark data sets may not be comparable to real life screening scenarios. Few virtual screening benchmark data sets have been reported that contain active and inactive molecules (extracted from the PubChem bioassay database).[47,48] Most of these high-throughput screening data sets contain a relatively lower ratio of active vs inactive molecules (1 active on every 2.557−22.466 inactives)[48] than what is found in most benchmark data sets (1 active on every 36−50 inactives)[38−40] and for several targets, actives can already largely be separated from inactive molecules based on simple physical−chemical properties.[48] Therefore, although these data sets represent real life screening data, they may not be sufficiently challenging to properly evaluate virtual screening methods for certain targets. Moreover, a truly systematic, comparative virtual screening assessment on different protein targets would require a complete biological activity data set in which the same set of molecules is tested against all targets. The possibilities to evaluate and compare virtual (fragment) screening approaches in a consistent manner for different protein targets is therefore not only limited by the low number of inactive (vs actives) molecules in publically available databases, but also by the relatively low number of molecules that have been tested against multiple proteins (Figure 1).[49−53]

This study represents a systematic comparative investigation of the challenges and applicability of different methods for ligand-based discrimination of fragment-like ligands from true (i.e., experimentally determined) inactive fragments. We have recently performed experimental screening studies of an in-house library of fragment-like molecules against different protein targets[51,54] This fragment−protein selectivity data set (containing true active and inactive molecules for different pharmaceutically relevant proteins) has enabled us to (i) perform chemogenomics analyses to identify unexpected similarities

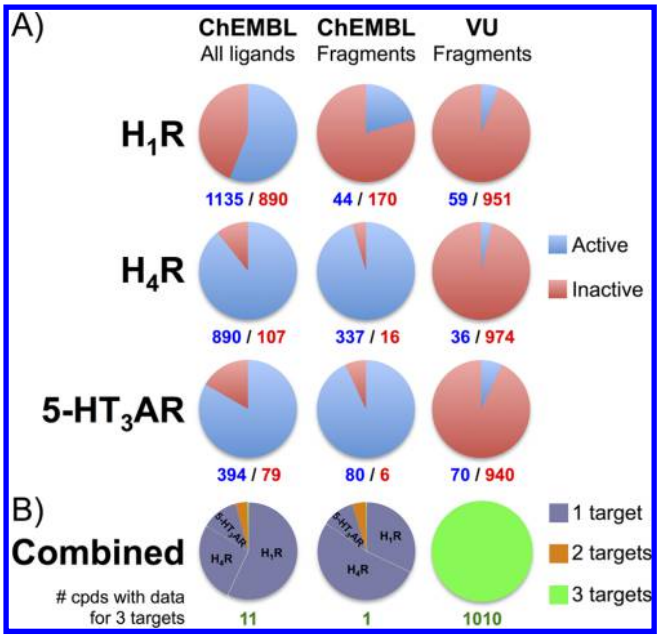


Figure 1. (A) Comparison of the available bioactivity data for each of the three case study protein targets histamine H_1 receptor (H_1R), histamine H_4 receptor (H_4R), and serotonin 5-HT_3A receptor (5-HT_3AR) extracted from ChEMBL,[59] fragment-like (heavy atom count (HA) ≤ 22, number of rotatable bonds (Nrot) ≤ 5, logP ≤ 3)[51,54] compounds from ChEMBL, and the VU fragment library used in the current study.[51] (B) Data completeness was assessed by combining the bioactivity data and counting the compounds with bioactivity data for one target (either H_1R, H_4R, or 5-HT_3AR), two targets (either H_1R/H_4R, H_1R/5-HT_3AR, H_4R/5-HT_3AR), or all three targets (H_1R/H_4R/5-HT_3AR). From ChEMBL version 19, all (p)K_i and (p)IC_{50} values for each of the three targets with a minimum confidence score of 8 were assessed.

and differences in ligand binding properties, and subtle ligand affinity and selectivity cliffs,[51,52] and (ii) to optimize and validate virtual screening methods to discover new bioactive molecules.[15] In the current studies we have used our fragment binding data sets for the histamine H_1 receptor (H_1R), the histamine H_4 receptor (H_4R), and the serotonin receptor (5-HT_3AR)[51,54] to validate our systematic retrospective virtual screening simulations. In order to arrive a ratio of about 1 active to 10 inactives and simulate typical hit rates observed in experimental fragment screening studies[9,51,55−58] the number of actives have been complemented with independent sets of active molecules extracted from ChEMBL.[59] All three case study proteins are of pharmaceutical relevance[104,105] and have been studied with fragment-based drug discovery approaches.[14,17,51,52,54,60−66] Both, H_1R and H_4R, are G protein-coupled receptors (GPCRs) and key players in immunological and allergic responses.[67] Antagonists for the H_1R receptor are known as the classical "antihistamines" and are used against allergic, inflammatory responses like rhinitis or pruritis.[68] The more recently identified H_4R is a promising drug target for diseases like asthma, allergic rhinitis, pruritis, and rheumatoid arthritis.[69−75] 5-HT_3AR is a ligand-gated ion channel and is found on neurons, where they trigger rapid depolarization due to a transient inward current.[76] 5-HT_3 receptor antagonists are in clinical use to treat nausea and vomiting in cancer patients receiving chemotherapy or radiation therapy.[77]

An advantage of our comparative study is that we used *true inactives* that are experimentally validated in in-house assays following the same assay for each target and using the same

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044

Table 1. Similarity Descriptors

Substructure Descriptors	
MACCS[a]:[101] (2D)	Encode presence/absence of 166 predefined structural keys
ECFP_4[b]: atom type based extended connectivity fingerprint (2D)[102] FCFP_4[b]: functional class based extended connectivity fingerprint (2D)[102] ECFC_4[b]: atom type based extended connectivity count (2D) FCFC_4[b]: functional class based extended connectivity count (2D)	Topological fingerprints that encode information on atom centered fragments. The number (4) describes a certain diameter of chemical bonds within that information on the surrounding atoms is encoded.
Property Descriptor	
EDprints: electron density fingerprint (2D)[96]	Encodes electron density information (calculated chemical shifts of ^{1}H and ^{13}C and partial charges).
Pharmacophore-Based Descriptors	
TAD[a]: typed atom distances, spatial 2-point pharmacophore based fingerprint (3D) TGD[a]: typed graph distances, graph based 2-point pharmacophore based fingerprint (2D) TAT[a]: typed atom triangles, spatial 3-point pharmacophore based fingerprint (3D) TGT[a]: typed graph triangles, graph based 3-point pharmacophore based fingerprint (2D)	Encodes atom types (Donor, Acceptor, Polar, Anion, Cation, Hydrophobe) and graph (2D) or spatial (3D) distances between 2 or 3 atoms.
GpiDAPH3[a]: graph based 3-point pharmacophore based fingerprint (2D) piDAPH3[a]: spatial 3-point pharmacophore based fingerprint (3D) piDAPH4[a]: spatial 4-point pharmacophore based fingerprint (3D)	Encodes atom types that are calculated based on three atomic properties: "in pi system", "is donor", "is acceptor", and graph (2D) or spatial (3D) distances between 3 or 4 atoms.
Shape-Based Descriptors	
ROCS[c]: rapid overlay of chemical structures (3D)	Conformational overlay of shape and pharmacophore features

[a]Implemented in MOE v. 2010.10 (Chemical Computing Group Inc, Montreal). [b]Implemented in Pipeline Pilot v. 8.0 (Accelrys Software Inc., San Diego). [c]OpenEye Scientific Software, Inc., Santa Fe, NM, USA.

chemically diverse library of fragment-like molecules.[51,54] In most other virtual screening evaluations decoy molecules, randomly selected from large chemical databases, are considered as inactives for a specific protein target, although these molecules have *not* been experimentally determined to be inactive for this protein.[38,42−44] It should be noticed that the number of fragment-like molecules available in ChEMBL (the largest curated publically available bioactivity database[59]) that have been reported to be inactive at the case study protein targets H_1R, H_4R, and 5-HT_3AR is 6 to 156 times lower than the number of experimentally determined inactive fragments for these proteins in the VU data set (Figure 1). Moreover, the numbers of fragment-like molecules in ChEMBL that are experimentally studied at multiple protein targets are significantly lower than in the VU fragment set (68−336 times lower for two targets, 1010 times lower for all three targets). The challenge of constructing consistent biological activity data sets for comparative virtual screening studies against different protein targets is further illustrated by previous analyses of ChEMBL which indicated that (i) biological activity data for least three different proteins within the same protein family is available for only a few thousand molecules[78] (e.g., GPCRs H_1R and H_4R), (ii) biological activity data for molecules at multiple target families (e.g., H_1R/H_4R GPCRs vs 5-HT_3AR LGIC) are still scarce.[53]

In our comparative retrospective virtual screening study we tested each similarity descriptor individually for each reference active for each of the three case study protein targets (H_1R, H_4R, and 5-HT_3AR) and applied different consensus scoring approaches. Consensus scoring combines the data from individual virtual screens with the aim to generate a result superior to the individual screens. Combining the results from individual reference actives[79−88] was previously defined as group fusion.[87,89] In the same way consensus scoring approaches can be applied for combining the results of the screens using different similarity descriptors,[22,24,44,90−94] also called similarity fusion.[87,89] Most of the similarity fusion approaches that have been reported to date involve the use of single reference structures.[22,24,44,90−93] In the current study we have to the best of our knowledge for the first time systematically evaluated different *combinations* of *group* fusion and *similarity* fusion approaches. Furthermore, the abilities of different similarity search methods to identify compounds that do not contain chemical scaffolds present in the respective reference actives were evaluated. Overall, the current study gives insights into some of the strengths and challenges of ligand-based virtual fragment screening that can be useful for future FBDD studies.

METHODS

Datasets. Experimentally validated active and inactive fragment-like molecules from in-house experimental screening campaigns of 1010 fragments[51] at H_1R, H_4R, and 5-HT_3AR were complemented with active fragment-like molecules, with high ECFP_4 diversity and $pK_i \geq 7.0$ selected from the ChEMBL database[59] to achieve an active:inactive ratio of approximately 1:10 (Table S1). All molecules obey the following fragment rules: heavy atom count (HA) ≤ 22, number of rotatable bonds (Nrot) ≤ 5, logP ≤ 3.[51,54] The descriptors were calculated using MOE v. 2010.10 (Chemical Computing Group Inc, Montreal).

Chemical Similarity Descriptors. The molecular similarities according to the Tanimoto similarity coefficient[95] were calculated using the following descriptors: MACCS, TGT, TGD, GpiDAPH3, TAT, TAD, piDAPH3, piDAPH4 implemented in MOE v. 2010.10 (Chemical Computing Group Inc, Montreal), FCFP_4, ECFP_4, FCFC_4, and ECFC_4 available in Pipeline Pilot v. 8.0 (Accelrys Software Inc., San Diego). EDprints[96] similarities were determined using an in-house script, while 3D shape-based ROCS v. 2.3.2 (OpenEye Scientific Software, Inc., Santa Fe, NM, USA) alignments were ranked according to ComboScore, which combines the Shape Tanimoto score (3D

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044

overlapping of shapes) and the Color Score (common pharmacophore features within the reference query and the test compounds in the screening database).[45] For the 3D methods (TAD, TAT, piDAPH3, piDAPH4, and ROCS), the 3D conformations of the reference query and the test compounds (actives and inactives) in the screening database were generated using Corina v. 3.46 (Molecular Networks GmbH, Erlangen, Germany). For the test compounds conformational databases were generated using OMEGA OMEGA v. 2.3.2 (OpenEye Scientific Software, Inc., Santa Fe, NM, USA). At this step 16 compounds could not be converted and were therefore removed from the test set. It should be noted that the results of 3D ligand structure-based virtual screening studies can depend on the 3D conformations of the query and test compounds in the screening database.[43,97] The 3D conformation−generation procedure applied in the current study is in line with previously reported virtual screening studies.[15,97,98] Table 1 gives and overview of the properties of the different similarity descriptors used in this study.

Scaffold Similarity Analysis. To determine scaffold similarity between compounds, the freely available program strip-it v. 1.0.1 (Silicos-it Schilde, Belgium) was used for the extraction of predefined scaffolds from organic small molecules. In this study, we generated "ring_with_linkers_1"[48] scaffolds, herein further called scaffold 1 (SCF1), and "murcko_2"[48] scaffolds, herein further called scaffold 2 (SCF2). SCF1 are extracted by removing all side chains.[74] SCF1 were further converted to SCF2 by exchanging all heteroatoms to carbon atoms, changing all bond orders to single bonds and finally reducing all linkers between ring systems to one single bond (Figure 2).[99,100]

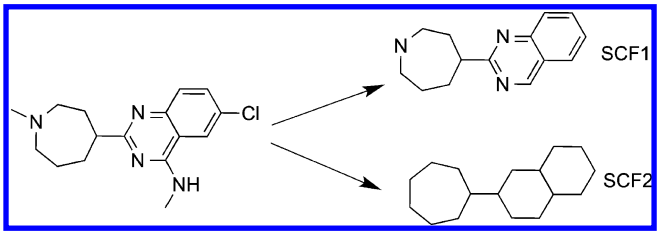


Figure 2. Demonstration of generation of SCF1 and SCF2.[48]

Analysis Retrospective Virtual Screening Studies. Similarity matrices were generated using all actives as reference actives and all actives and inactives as test compounds. Enrichment factors (EF) were calculated based on the 1% fraction of enriched false positives (FP) (excluding self-comparisons).[37]

$$EF = \text{fractionTP}/\text{fractionFP}$$

In order to test the performance of the virtual fragment screening approaches in retrieving active compounds with scaffolds differing from the respective reference active, we also generated similarity matrices where scores from test-active comparisons were excluded if the compounds had similar scaffolds. As we had several actives for each target the enrichment for each reference active was calculated individually. By averaging all these enrichment factors the average enrichment factor was obtained.[27] The individual similarity scores for each reference active and for each similarity descriptor were additionally combined using consensus scoring approaches.

Consensus Scoring Approaches. The following consensus scoring methods were applied:

Mean-Value. For each test compound the mean similarity score over all reference actives was calculated.[44,79,80,83−86,88]

Max-Value. For each test compound the maximum similarity score to any of the reference actives was selected.[30,79,80,82−86,88,92,103]

Mean-Rank. First, the ranks of the test compounds for each reference active were normalized (0−1), and second, the mean of the normalized ranks for each test compound was calculated[22,44,80,88,90−93]

Ranked-by-Vote. Initially the ranks of the test compounds were normalized (0−1) and subsequently the rate of the occurrence in the top 1% (normalized rank ≥ 0.99) was calculated for each test compound.[44,90,92]

Table 2 demonstrates how these consensus scoring approaches work based on theoretical examples of virtual screening hit lists.

RESULTS AND DISCUSSION

We have retrospectively validated the applicability of a large range of different ligand-based similarity descriptors for virtual

Table 2. Demonstration of Different Consensus Scoring Methods for the Case Where All Compounds Are Taken into Account (All) and for the Case Where Only Compounds with Scaffolds Dissimilar to the Reference Are Taken into Account[a]

all[b]	ref 1	ref 2	ref 3	mean-value	max-value	mean-rank	ranked-by-vote
test 1	0.70 (1.00)	0.80 (1.00)	0.90 (1.00)	0.80	0.90	1.00	1.00
test 2	0.60 (0.63)	0.50 (0.57)	0.40 (0.38)	0.50	0.60	0.52	0.00
test 3	0.10 (0.00)	0.20 (0.00)	0.30 (0.00)	0.20	0.30	0.00	0.00
test 4	0.60 (0.63)	0.80 (1.00)	0.40 (0.38)	0.60	0.80	0.66	0.33
test 5	0.50 (0.25)	0.30 (0.29)	0.70 (0.75)	0.50	0.70	0.43	0.00
SCF1/SCF2[c]	ref 1	ref 2	ref 3	mean-value	max-value	mean-rank	ranked-by-vote
test 1		0.80 (1.00)	0.90 (1.00)	0.85	0.90	1.00	1.00
test 2	0.60 (1.00)		0.40 (0.50)	0.50	0.60	0.75	0.50
test 3	0.10 (0.00)	0.20 (0.00)	0.30 (0.00)	0.20	0.30	0.00	0.00
test 4	0.60 (1.00)		0.40 (0.50)	0.50	0.60	0.75	0.50
test 5	0.50 (0.40)	0.30 (0.50)		0.40	0.70	0.45	0.00

[a]The similarity values between reference (ref) and test (test) compounds are theoretical values. The numbers in brackets are the normalized ranks (best 1, worst 0) of the test compounds compared to the respective reference compound. Mean-value and max-value are calculated by taking the mean or the maximum of the similarity values. Mean-rank is the mean of the normalized ranks and ranked-by-vote is the rate of the occurrence (1 always, 0 never) in the top 1% of the ranks (normalized rank ≥ 0.99). [b]All: all similarity scores were included for group fusion. [c]SCF1/SCF2: only similarity values resulting from comparisons of reference and test compounds with dissimilar SCF1 or SCF2 were used for group fusion.

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044

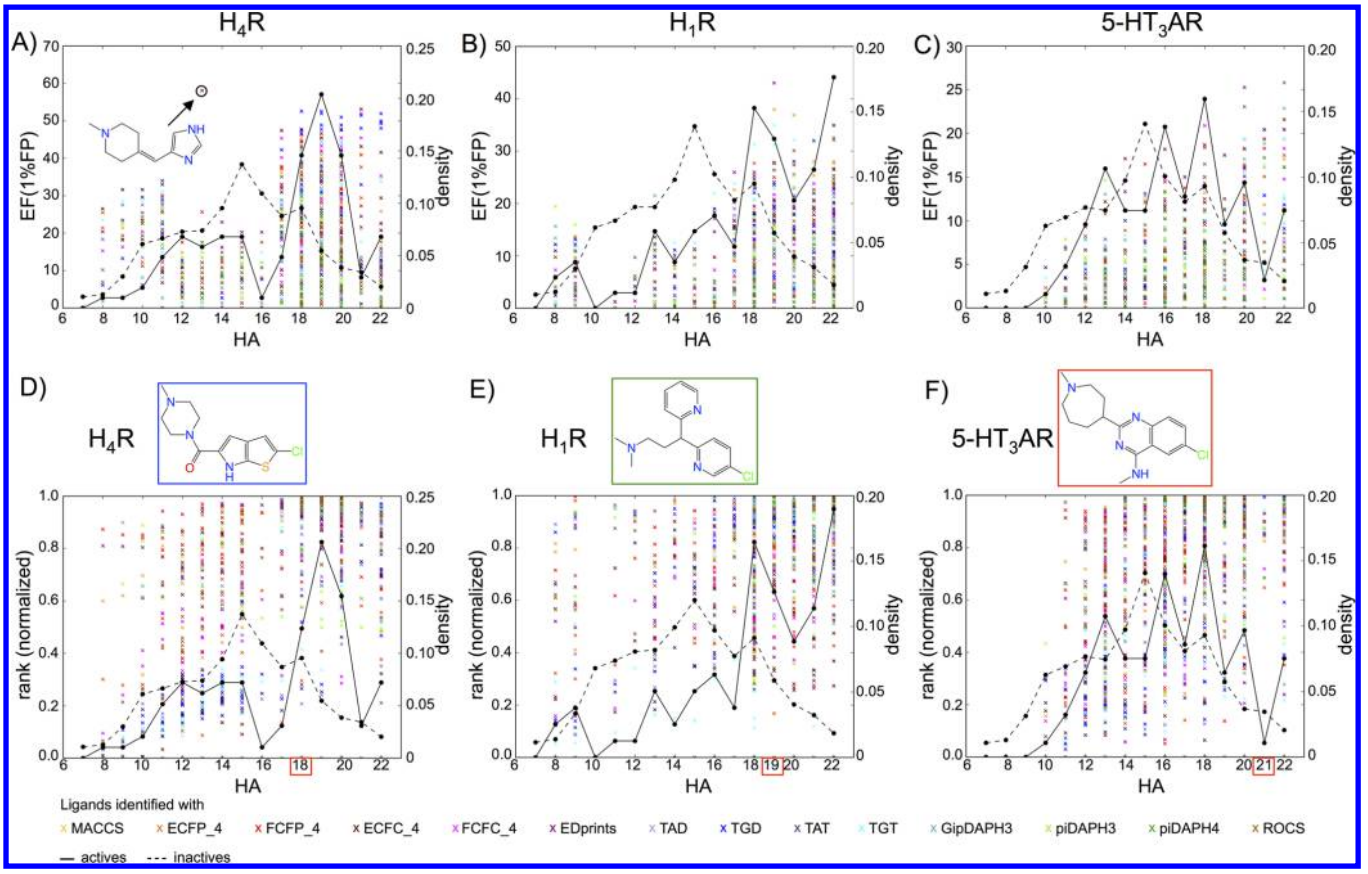


Figure 3. Enrichment at 1% FP (left axis) obtained for each reference ligand using different similarity descriptors (colored crosses). The right axis shows the relative amount of active compounds (density) with a certain heavy atom (HA) count. Test compounds sharing similar SCF2 scaffolds with the respective reference ligand were excluded from the analyses (A–C). In panel A the methimepip analogue is indicated as a relatively small reference ligand (13 heavy atoms) that gives a relatively high virtual screening enrichment factor using ECFC_4. Panels D–F show how reference ligands that achieved highest average enrichment over all similarity methods (shown above panels D–F) rank other SCF2 dissimilar actives using different similarity descriptors (colored crosses). Ranks were normalized from 0 to 1 where the first ranked compound has the rank 1 and the last ranked compound has the rank 0.

fragment screening. For this comparative virtual screening assessment we used fragment-like actives and inactives for three pharmaceutically relevant case study protein targets, H_1R, H_4R, and 5-HT_3AR,[104,105] that have been studied with fragment-based drug discovery approaches.[14,51,52,54,60−66] We first investigated the performance of each reference active individually for each of the three case study protein targets. Subsequently we tested different consensus scoring methods to combine the screening results from the actives (group fusion)[79−88] and also to combine the results from multiple similarity descriptors (similarity fusion).[22,44,90−93]

Performance of Individual Reference Actives. We first analyzed the performance of the individual reference ligands in retrieving actives that have dissimilar SCF2. Figure 3A–C shows the enrichment factors at 1% FP in retrospective virtual screening runs against H_1R, H_4R, and 5-HT_3AR test sets based on different similarity descriptors and using reference ligands with different numbers of heavy atoms (HAs). For all three tested targets chemical similarity searches against relatively larger fragment-like reference ligands (HA count 16–22) give higher enrichments of active over inactive molecules in the retrospective virtual screening runs. It should be noted however that there is one single small H_4R reference ligand (HA count: 13) that gives exceptional high enrichment of 58.1 using ECFC_4 (Figure 3A). The general better performance of the relatively larger actives seems to be independent of the relative amount active molecules compared to inactive molecules. For example for H_4R the number of actives with a HA count of 21 or 22 is relatively low, but nevertheless high enrichments are observed (Figure 3A). The same is true for actives for 5-HT_3AR with a HA count of 20, 21, or 22 (Figure 3C).

One reason for the general better performance of larger actives might be that small changes in the structure of small compounds are expected to lead to larger dissimilarities than for bigger compounds. Except from that part also the SAR might change more dramatically, leading to so-called activity cliffs.[106,107] Moreover the Tanimoto similarity measurement is described to be size dependent.[21,32] These factors might be reasons for the lower enrichments of actives over inactives in virtual screening if ligands with low number of HAs are used as reference compounds (Figure 3). Figure 3D–F shows how reference ligands, that give the highest average enrichment (across all similarity descriptors), rank compounds with different numbers of HAs. Virtual screening runs against these references ligands, that contain 18, 19 and 21 HAs for H_4R, H_1R and $5HT_{3A}R$ respectively, generally yield higher rankings for fragments that are relatively larger (HA count 16–22) which is largely independent of the HA distributions of actives and inactives (Figure 3D–F). It should be noted that especially for H_1R and H_4R data sets the relative number of small actives and ratio between actives and inactives is lower for molecules with less than 16 HAs (Figure 3D–F), which affects virtual screening

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044

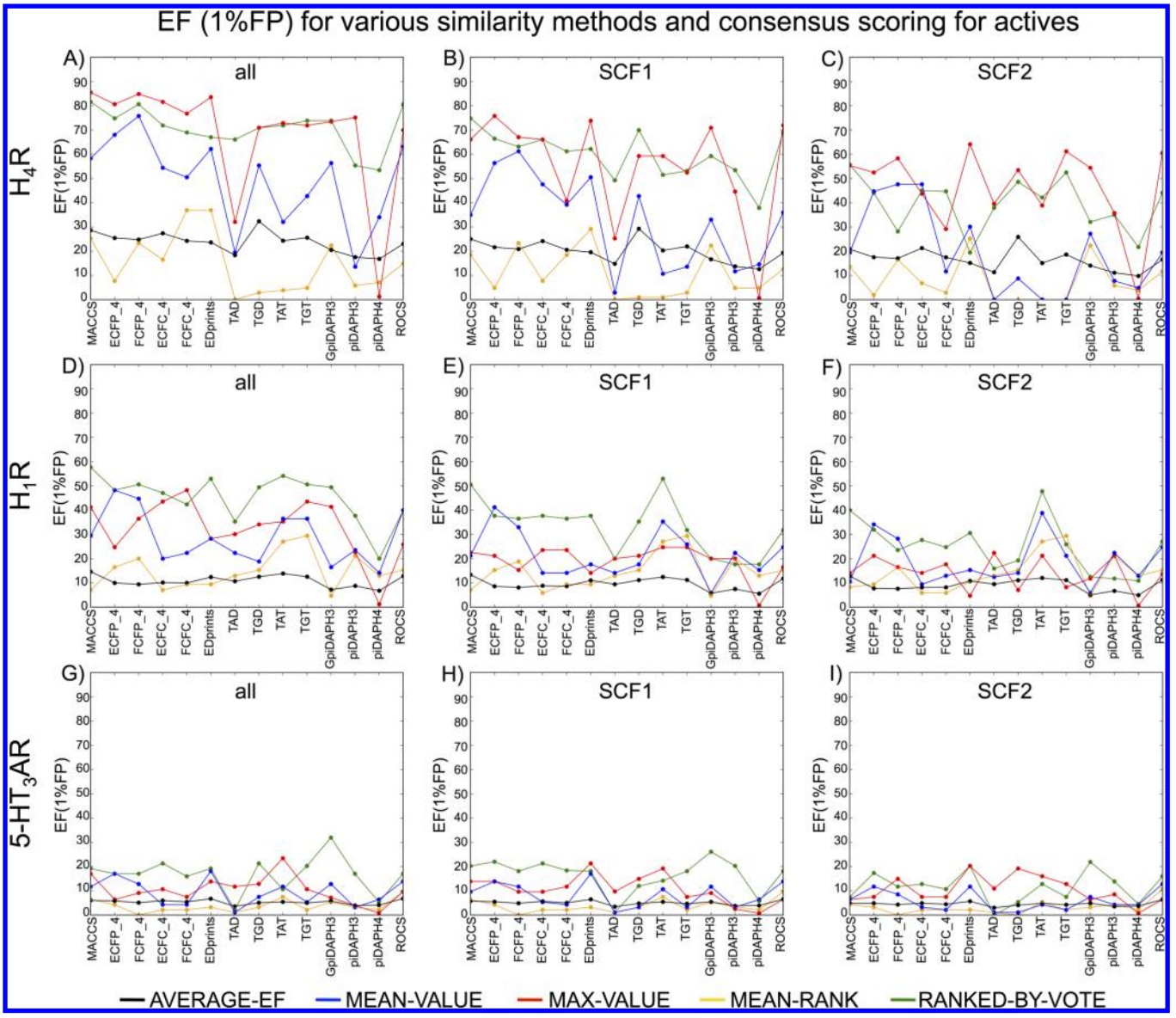


Figure 4. Enrichment factors at 1% FP of various similarity descriptors using different group fusion methods (blue *mean-value*, red *max-value*, yellow *mean-rank*, green *ranked-by-vote*) and the average enrichment (black) for H_4R (A–C), H_1R (D–F), and 5-HT_3AR (G–I). Enrichment factors in panel A, D, and G were calculated using all similarity values (except self-similarity values), enrichment factors in panel B, C, E, F, H, and I were calculated leaving out similarity values resulting from comparison of reference and test compounds that have the same SCF1 (B, E, and H) or SCF2 (C, F, and I).

ranking and enrichments for these smaller compounds. Overall our retrospective virtual screening analyses seem to indicate that the discrimination of very small (< 16 HAs) active fragments from inactive fragments is an especially challenging task in virtual fragment screening.

Average Performance of Individual Chemical Similarity Search Methods. The average enrichment factor of a method for a specific target is calculated by dividing the sum of all enrichment factors obtained with all actives for a specific target by the total number actives.[27] Comparison of the average enrichment factor at 1% FP (Figure 4, Table S2) shows that the performance of the different similarity descriptors decreases with decreasing scaffold similarity of the test compounds compared to the reference actives (all > SCF1 > SCF2). Retrospective virtual screening enrichments are generally somewhat higher for H_4R, than for H_1R, while enrichments for 5-HT_3AR are relatively low. This is also the case if similar scaffolds (SCF1 and SCF2) are not considered in the analysis (Figure 4 and Table S2). This is demonstrated in Figure 5, which shows the average enrichment curves observed for EDprints for all three targets. In this plot the average early enrichment curve is the highest for H_4R followed by H_1R and 5-HT_3AR. A comparison of the MACCS-similarities of the actives of each data set revealed that the actives of H_4R are more similar to each other than actives of H_1R, while the actives of 5-HT_3AR are relatively dissimilar to each other (Figure 6). The observation that the performance of the similarity descriptors decreases with decreasing self-similarity of the actives is in line with other studies.[80]

Comparing the average enrichment factor at 1% FP (Figure 4, Table S2) shows that for all three tested targets (H_4R, H_1R and 5-HT_3AR) there is no method that always leads to best average enrichments. For example MACCS is within the three best performing similarity descriptors for H_4R and H_1R while ROCS is within the best performing similarity descriptors for H_1R and 5-HT_3AR. piDAPH3 and piDAPH4 are among worst performing descriptors for all three data sets. This observation is

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030–1044

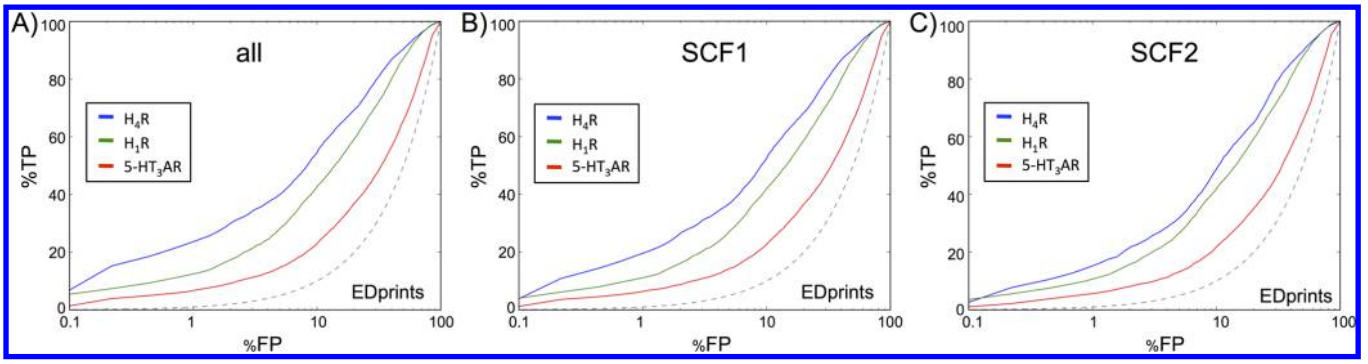


Figure 5. Average enrichment curves observed for EDprints for H_4R, H_1R, and 5-HT_3AR including (A) all similarity values and single similarity values resulting from comparison of compounds with (B) dissimilar SCF1 and (C) dissimilar SCF2.

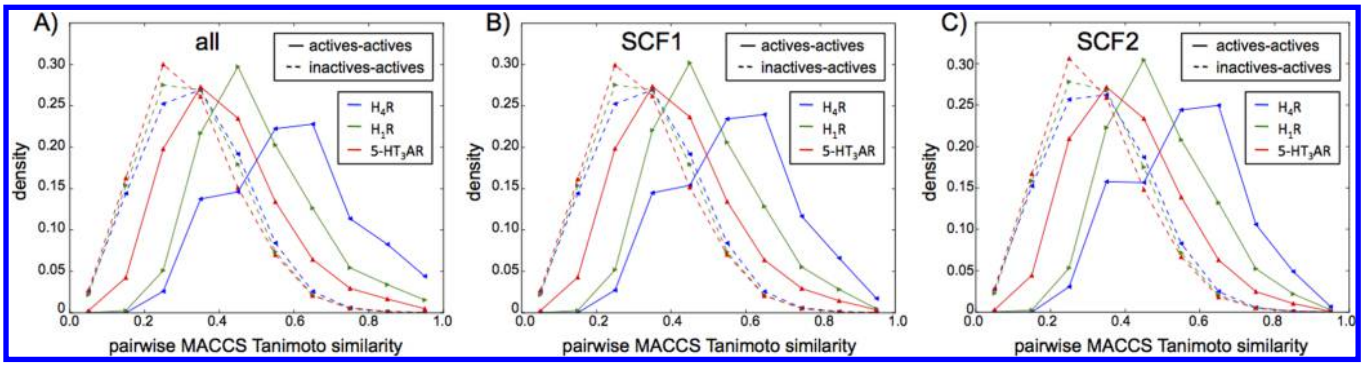


Figure 6. Frequency of pairwise MACCS Tanimoto similarities resulting from comparisons between actives−actives (solid lines) and inactives−actives (dashed lines) including (A) all similarity values or including only similarity values resulting from comparison of compounds with (B) dissimilar SCF1 or (C) dissimilar SCF2.

independent whether if similarity values for all test compounds are taken into account (Figure 4A, D, G, Table S2) or if similarity values of test compounds that have similar SCF1 or SCF2 to the respective reference active are excluded. (Figure 4B, C, E, F, H, I, Table S2). The low virtual screening accuracy of piDAPH3 and piDAPH4 compared to other methods has been observed in previous retrospective virtual screening evaluations.[22] However, it has to be noted that especially in the case of the 5-HT_3AR, were the worst performance of all similarity descriptors compared to the other data sets is observed, there is hardly any difference in the average enrichment factors of the individual methods (Figure 4G, H, and I, Table S2).

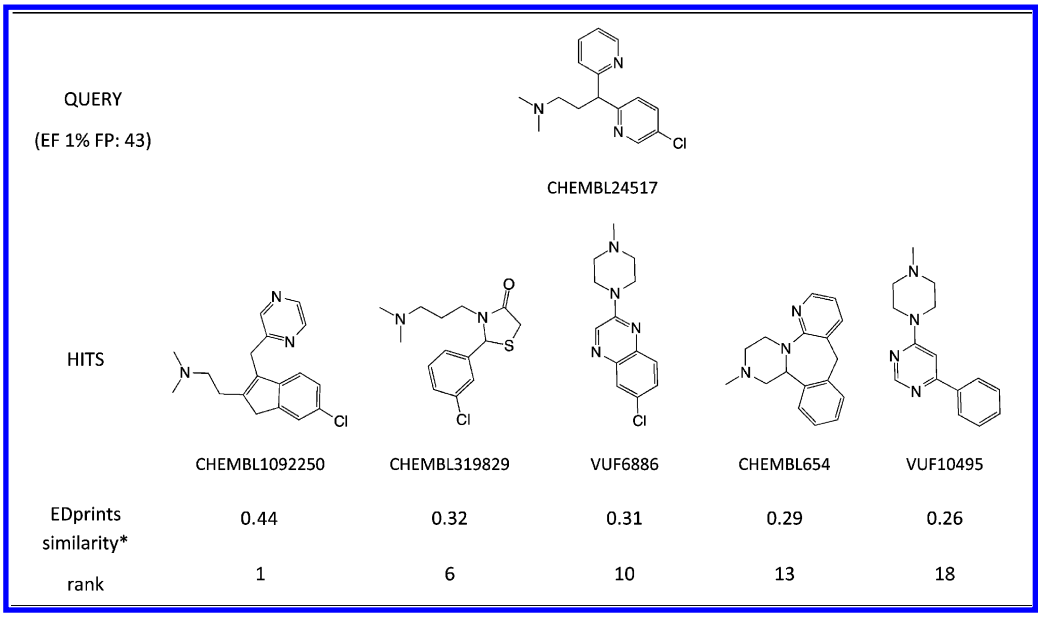
Consensus Scoring. *Combination of Similarity Values: Group Fusion.* If multiple actives are available, consensus scoring approaches to combine similarity values of test compounds can be applied to enhance the effectiveness of the virtual screening methods.[79−89] In general if quite an amount of actives are known also more sophisticated machine learning methods like naive Bayesian classifiers or a support vector machines can be used to discriminate actives from inactives.[23,94,108−112] In this study we evaluated the performance of four distinct consensus scoring approaches to combine similarity values or the ranks of the test compounds obtained for multiple actives (also called group fusion[87,89]): mean-value, max-value, mean-rank, and ranked-by-vote. A demonstration how these consensus scoring approaches work is shown in Table 2 in the Methods section. Mean-value and max-value are calculated by taking the mean or the maximum of the similarity values. Mean-rank is the mean of the normalized ranks, and ranked-by-vote is the rate of the occurrence (1 always, 0 never) in the top 1% of the ranks (normalized rank $\geq$ 0.99).

Figure 4 shows that certain consensus scoring methods achieve better enrichments than the average enrichment. For all three tested targets both max-value or ranked-by-vote consensus methods most often achieve the best results. Ranked-by-vote achieves better enrichments (at 1% FP) than average enrichment except for TAD for 5-HT_3AR. Max-value achieves better enrichment than the average enrichment, except for piDAPH4 for all targets and additionally for EDprints, TGD and TGT for H_1R in the SCF2 similarity scenario and for piDAPH3 for 5-HT_3AR in the SCF1 scaffold similarity scenario. So far using the max-value was already reported to be successful for group fusion.[30,79,80,82−86,88,89,92,103] However, to the best of our knowledge, studies about the use of ranked-by-vote for group fusion have not been reported so far.

The performance of consensus scoring of the actives is not related to the performance of the average enrichment. For example for the H_1R test set the GpiDAPH3 method gives very low average enrichment factor compared to all other similarity descriptors in retrospective virtual screening runs (Figure 4D, Table S2). However, when ranked-by-vote is used as consensus scoring, enrichment with GpiDAPH3 at 1% FP is 7 times higher and comparable with the best methods (Figure 4D, Table S2). But especially this method performs poorly for H_1R if the scaffold similarity of the test compounds compared to the respective reference active is reduced (Figure 4E and F, Table S2). For 5-HT_3AR GpiDAPH3 is always among the best methods using ranked-by-vote as group fusion for all similarity scenarios. This shows that the relative performance of the individual methods to identify new scaffolds (i.e., considering all compounds or SCF1 or SCF2 scaffold similarity scenarios) is in fact target dependent.

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044

Table 3. Active Hits of an EDprints Similarity Search against the H_1R Active That Showed the Best Enrichment Factor at 1% FP as a Query Compound[a]



[a]As test compounds only compounds that have a dissimilar SCF2 to the query compound were used. *EDprints similarities were calculated using the McConnaughey similarity measure (which ranges from −1 to 1).[96]

The decrease in performance of the similarity descriptors with decreasing self-similarity is a challenge for the discovery of novel scaffolds. Our study shows, however, that for each target there are some promising methods for the early retrieval of compounds that have dissimilar SCF2 compared to the respective reference compounds. The best performing methods if SCF2 dissimilar compounds compared to the reference ligands are enriched are the property descriptor EDprints using max-value for the H_4R data set, the pharmacophore-based descriptor TAT using ranked-by-vote for the H_1R data set and the pharmacophore-based descriptor GpiDAPH3 using ranked-by-vote for the 5-HT_3AR data set (Figure 4C, F, and I, Table S2). Although the performance of the similarity descriptors in enriching SCF2 dissimilar scaffolds is target dependent we could show that EDprints is not only most successful in enriching of active compounds for the H_4R data set that have dissimilar SCF2 compared to the respective reference actives but also that EDprints is the only descriptor that is within the best four performing methods for the SCF2 scenario for all tested targets. In other studies substructure descriptors,[103,113] extended connectivity descriptors,[103] pharmacophore-based descriptors,[88,103,113,114] and shape-based descriptors[115] have been already reported to be successful in scaffold hopping or in enriching actives among a diverse set of compounds.[116] EDprints, which encodes information about electron densities and charges, have been previously shown to be more robust toward subtle rearrangements of chemical groups and more suitable for screening against reference molecules with fused ring systems compared to other similarity descriptors.[96] For the H_1R data set the highest enrichment of 43.0 at a 1% FP rate for a specific H_1R active, using test compounds with dissimilar SCF2 compared to the respective reference active, was observed using EDprints (see Figure 3B). Table 3 shows this H_1R active and different compounds that are among the top 20 ranked compounds (which are all actives), which demonstrates the scaffold hopping ability of ligand based approaches.

Combination of Similarity Descriptors: Similarity Fusion. In the same way as the results from different actives can be fused into a combined *score* (*group fusion*[87,89]), these combined scores of the different similarity descriptors can be again fused (*similarity fusion*[87,89]).[22,44,90−94,117] For this task the group fused scores from each similarity descriptor were normalized. For the different similarity descriptors only scores that were obtained with the same consensus scoring approach for group fusion were further combined with a certain consensus scoring approach. Enrichment factors for all combinations of the 14 investigated similarity descriptors (16 369 possible combinations), combined with a certain group and similarity fusion approach (4 × 4 = 16 possible combinations) were calculated. In this way we tested for each of the three protein targets which combination of similarity descriptors combined with a certain combination of group and similarity fusion method (261 904 possible combinations) yields the highest virtual screening enrichment.

Table 4 shows the combinations of similarity descriptors that lead to highest enrichment. For H_4R using all test compounds and for 5-HT_3AR using test compounds with a dissimilar SCF2 compared to the actives several combinations of similarity descriptors lead to best enrichments (Table S3 and Table S4). In Table 4 the combination(s) with the smallest number of similarity descriptors are shown.

For all tested targets and similarity scenarios the enrichment factor (1% FP) of the best combination of similarity descriptors is always better than the enrichment factor of the best individual method (using the same group fusion method). Although the best combination mostly includes the similarity descriptor that gives the highest enrichment this is not necessarily always the case as demonstrated for H_4R (SCF1), H_1R (SCF1) and for 5-HT_3AR (SCF2). For H_4R (SCF1) for example the best combination (group fusion max-value, similarity fusion mean-value) gives an enrichment factor of 93.2 (at 1% FP) and includes MACCS, TGT, GpiDAPH3 and ROCS. However, ECFP_4, which results individually in the highest enrichment using max-value as group fusion method, is not included. If the best four

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044

Table 4. Combinations of (the Smallest Number of) Similarity Descriptors That Lead to Best Enrichment (Marked with X) and the Similarity Descriptor That Leads on Its Own to Best Enrichment with the Same Group Fusion Method (in Bold, if Included in the Best Combination Marked with X; if Not, Marked with O)

target	similarity scenario	EF (1% FP) best single method	EF (1% FP) best combination	group fusion	similarity fusion	MACCS	ECFP_4	FCFP_4	ECFC_4	FCFC_4	EDprints	TAD	TGD	TAT	TGT	GpiDAPH3	piDAPH3	piDAPH4	ROCS
H_4R	All[a]	85.4	91.3	max-value	mean-value	**X**									X	X			
	All[a]	85.4	91.3	max-value	mean-value	**X**						X				X			
	SCF1[b]	75.7	93.2	max-value	mean-value	X	**O**								X	X			X
	SCF2[b]	64.1	85.4	max-value	mean-value	X					**X**	X	X		X		X		
H_1R	All[a]	57.6	71.8	ranked-by-vote	mean-rank	**X**	X							X					
	SCF1[b]	52.9	69.4	ranked-by-vote	mean-value	X			X			X		**O**	X				X
	SCF2[b]	47.7	65.9	ranked-by-vote	mean-value	X					X			**X**	X				
5-HT_3AR	All[a]	31.9	39.4	ranked-by-vote	mean-rank	X					X			X		**X**	X		X
	SCF1[b]	26.1	34.0	ranked-by-vote	mean-rank	X			X							**X**	X		X
	SCF2[b]	20.2	27.7	max-value	mean-rank	X					**O**		X			X			

[a]All: All similarity scores were included for group fusion. [b]SCF1/SCF2: Only similarity values resulting from comparisons of reference and test compounds with dissimilar SCF1 or SCF2 were used for group fusion.

J. Chem. Inf. Model. 2015, 55, 1030−1044
DOI: 10.1021/ci500694c

similarity descriptors, ECFP_4, EDprints, GpiDAPH3, and ROCS (using max-value as group fusion) are combined then the best similarity fusion approach (mean-value and mean-rank) results in an enrichment of only 88.3. Interestingly, for all tested virtual screening scenarios MACCS is always included in the (minimal) combination of similarity descriptors that lead to the best enrichment (Table 4).

An analysis of the frequency of how often a similarity descriptor is included within the combinations of similarity descriptors that lead to the best 1% of enrichments (out of all combinations of similarity descriptors combined with all combinations of group and similarity fusion methods which is 261904 combinations) shows that MACCS, ROCS, and TGD are very frequently included (Table 5). They occur ≥53% (MACCS), ≥47.7% (ROCS), and ≥44.9% (TGD) for all targets and similarity scenarios. However, the success of other similarity descriptors is very dependent on the target and/or on the tested similarity scenario. The EDprints similarity descriptor for example occurs in 76.6% of the cases in combinations that give best enrichments for 5-HT_3AR (all). The same EDprints method however is only part of 19.2% of the best combinations for H_4R (all). Another example is ECFC_4 that occurs in 51.4% of the cases in the top 1% combinations for H_4R when SCF1 dissimilar test compounds compared to the actives are included, but only in 3.6% of the cases if all test compounds are included (Table 5).

Table 6 shows the frequency how often a certain combination of a group and a similarity fusion method leads to the best (top 1%) of enrichments (out of all 261 904 tested combinations). Max-value or ranked-by-vote as group fusion method most often lead to top enrichments. Max-value leads in more than 95% of the cased to top enrichments for H_4R whereby ranked-by-vote leads in more than 99% of the cased to top enrichments for H_1R and in more than 96% (all and SCF1) or 60% (SCF2) of the cases for 5-HT_3AR.

In contrast to group fusion, for similarity fusion the mean-value or mean-rank most often leads to the best (top 1%) enrichments (Table 6). Mean-value leads most often to top 1% enrichments for SCF1 and SCF2 similarity scenarios over all tested targets. Especially for H_1R (SCF1 and SCF2) more than 99% of the top 1% consensus scoring were obtained by using mean-value as similarity fusion method. If all similarity values are included, mean-rank more often leads to top 1% enrichments. Especially for 5-HT_3AR mean-rank as similarity fusion method leads in more than 81% of the cases to top 1% enrichments.

Analysis of the enrichments achieved with similarity fusion methods over all possible combinations of similarity descriptors (for each group fusion methods) revealed that mean-value in more than 89.3% and mean-rank in more than 93.7% of the cases for the tested enrichment scenarios achieved enrichments that are higher than the average of the enrichments obtained with the individual similarity descriptors that were included in the similarity fusion model (Table S5). Moreover the enrichments obtained with the mean-value or mean-rank similarity fusion method were also frequently higher than the enrichments of any of the individual similarity descriptors (Table S6).

These results of our retrospective virtual fragment screening studies against the three case study protein targets, H_1R, H_4R, and 5-HT_3AR recommend the use of mean-value or mean-rank similarity fusion consensus scoring, as the performance of the individual similarity descriptors is rather target-dependent. Although extension of a truly comparative virtual screening assessment to more protein targets would require biological activity profiles of the same set of molecules against more protein

Table 5. Occurrence (%) of the Different Similarity Descriptor Combinations That Give the (Top 1%) Best Enrichments (at a 1% FP Rate)[a]

target	Similarity scenario	MACCS	ECFP_4	FCFP_4	ECFC_4	FCFC_4	EDprints	TAD	TGD	TAT	TGT	GpiDAPH3	piDAPH3	piDAPH4	ROCS
H_4R	All[b]	69.1	38.6	53.7	3.6	14.4	19.2	32.1	50.4	54.7	65.9	69.6	57.2	48.9	76.4
	SCF1[c]	53.0	56.8	19.6	51.4	18.4	55.7	32.2	72.0	56.4	56.7	78.2	22.3	33.6	72.8
	SCF2[c]	88.6	74.2	47.5	31.5	28.8	56.4	67.6	83.1	47.3	71.2	44.6	38.2	21.4	49.9
H_1R	All[b]	81.7	51.5	16.8	55.2	45.9	37.8	23.8	62.7	55.2	67.9	51.0	34.1	10.9	58.6
	SCF1[c]	76.5	70.7	39.3	65.9	46.3	43.6	63.0	81.0	73.1	69.5	41.3	30.0	19.7	58.5
	SCF2[c]	89.0	57.4	42.5	59.5	56.3	52.3	49.0	67.6	85.6	76.4	38.5	41.0	40.0	47.7
5-HT_3AR	All[b]	77.6	43.3	33.9	36.4	50.3	76.6	32.4	44.9	36.0	45.6	63.3	75.9	29.0	60.6
	SCF1[c]	62.1	53.5	50.7	52.7	39.7	63.2	11.3	47.6	40.4	55.1	80.4	83.7	29.4	51.3
	SCF2[c]	60.9	38.4	35.3	30.2	36.7	49.3	23.6	49.8	34.2	40.7	49.7	53.2	39.6	56.6

[a]Enrichments were calculated for of all possible combination of similarity descriptors (16 369) and all 16 (4 × 4) combinations of group and similarity fusion methods (which is 261 904 combinations). [b]All: All similarity scores were included for group fusion. [c]SCF1/SCF2: Only similarity values resulting from comparisons of reference and test compounds with dissimilar SCF1 or SCF2 were used for group fusion.

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044

Table 6. Frequency (%) of a Certain Combination of a Group and a Similarity Fusion Method Leading to Enrichments within the Top 1%[a]

H_4R												
similarity scenario	all[b]				SCF1[c]				SCF2[c]			
	similarity fusion											
group fusion	mean-value	max-value	mean-rank	ranked-by-vote	mean-value	max-value	mean-rank	ranked-by-vote	mean-value	max-value	mean-rank	ranked-by-vote
mean-value	0	0	0	0	0	0	0.0	0	0	0	0	0
max-value	41.5	0	56.8	0	62.3	0	33.3	0	69.5	0	29.3	0
mean-rank	0	0	0	0	0	0	0	0	0	0	0	0
ranked-by-vote	0	0	1.7	0	2.3	0	2.1	0	0	0	1.2	0
H_1R												
similarity scenario	all[b]				SCF1[c]				SCF2[c]			
	similarity fusion											
group fusion	mean-value	max-value	mean-rank	ranked-by-vote	mean-value	max-value	mean-rank	ranked-by-vote	mean-value	max-value	mean-rank	ranked-by-vote
mean-value	0.1	0	0	0	0	0	0	0	0	0	0	0
max-value	0	0	0	0	0	0	0	0	0	0	0	0
mean-rank	0	0	0	0	0	0	0	0	0	0	0	0
ranked-by-vote	46.5	4.7	48.7	0	99.4	0	0.6	0	99.4	0	0.6	0
5-HT_3AR												
similarity scenario	all[b]				SCF1[c]				SCF2[c]			
	similarity fusion											
group fusion	mean-value	max-value	mean-rank	ranked-by-vote	mean-value	max-value	mean-rank	ranked-by-vote	mean-value	max-value	mean-rank	ranked-by-vote
mean-value	2.1	0	0.6	0	2.3	0	0.6	0	4.0	0	9.1	0
max-value	0.1	0	0.3	0	0	0	0.4	0	7.9	11.9	6.5	0.5
mean-rank	0	0	0	0	0	0	0	0	0	0	0.1	0
ranked-by-vote	11.7	1.1	81.0	3.1	70.3	0.6	17.0	8.8	35.3	0.4	24.3	0

[a]Enrichments (at 1% FP) were calculated for of all possible combination of similarity descriptors (16 369) and all 16 (4 × 4) combinations of group and similarity fusion methods (which is 261 904 combinations). [b]All: All similarity scores were included for group fusion. [c]SCF1/SCF2: Only similarity values resulting from comparisons of reference and test compounds with dissimilar SCF1 or SCF2 were used for group fusion.

targets, it should be noted that also other comparative virtual screenings at other proteins suggest the use of similarity fusion methods.[22,90] Mean-rank was already reported to be successful for similarity fusion.[22,44,90−93] To the best of our knowledge, this study is the first study showing that also the use of the mean-value is successful for similarity fusion.

CONCLUSIONS

This study presents a comparative retrospective LBVS study of fragment-like compounds based on a set of experimentally determined *actives* and *inactives* at three different case study protein targets: G protein-coupled receptors H_1R and H_4R and ligand-gated ion channel 5-HT_3AR. Here we have shown that LBVS methods are in fact suitable for the efficient enrichment of active fragments from a database of active and inactive fragments. Moreover, even scaffold hopping can be achieved by employing these methods. However, our study also pointed out that the enrichment of very small fragments (<16 HAs) can be an especially challenging task. The use of a data set with true actives and true inactives enabled us to perform an evaluation based on a consistent data set of solid experimental data. We tested several similarity descriptors with and without consensus scoring approaches. We have shown that for the three investigated protein targets consensus scoring methods are a very effective means to increase enrichment compared to the average enrichment. Ranked-by-vote and also (although less often) the max-value performed best for group fusion; mean-rank and mean-value performed best for similarity fusion. Our retrospective ligand-based virtual fragment screening analyses highly recommend the simultaneous use of both group fusion and similarity fusion methods if multiple actives and similarity descriptors are available. For future research it will be of high interest to investigate whether the evaluated consensus scoring approaches are also successful for other protein targets and to test the combination of group and similarity fusion approaches in a prospective virtual screening.

ASSOCIATED CONTENT

Supporting Information

Additional analyses and information regarding the virtual screening studies: data set composition, enrichment factors, highest-scoring similarity method combinations, frequencies of increased performance by group and similarity fusion combinations. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: c.de.graaf@vu.nl.

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.J.K., H.F.V., S.N., R.L., I.J.P.d.E., and C.d.G. participate in the European Cooperation in Science and Technology (COST) Action CM1207: GPCR-Ligand Interactions, Structures, and Transmembrane Signalling: a European Research Network (GLISTEN). This research was financially supported by The Netherlands Organization for Scientific Research (NWO VENI Grant 700.59.408 to C.d.G. and NWO TOP-PUNT Grant to R.L.).

■ REFERENCES

(1) Murray, C. W.; Rees, D. C. The Rise of Fragment-Based Drug Discovery. *Nat. Chem.* **2009**, *1*, 187−192.
(2) Warr, W. A. Fragment-Based Drug Discovery. *J. Comput. Aided Mol. Des.* **2009**, *23*, 453−458.
(3) de Kloe, G. E.; Bailey, D.; Leurs, R.; de Esch, I. J. Transforming Fragments into Candidates: Small Becomes Big in Medicinal Chemistry. *Drug Discovery Today* **2009**, *14*, 630−646.
(4) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. Recent Developments in Fragment-Based Drug Discovery. *J. Med. Chem.* **2008**, *51*, 3661−3680.
(5) Fink, T.; Reymond, J. L. Virtual Exploration of the Chemical Universe up to 11 Atoms of C, N, O, F: Assembly of 26.4 Million Structures (110.9 Million Stereoisomers) and Analysis for New Ring Systems, Stereochemistry, Physicochemical Properties, Compound Classes, and Drug Discovery. *J. Chem. Inf. Model.* **2007**, *47*, 342−353.
(6) Loving, K.; Alberts, I.; Sherman, W. Computational Approaches for Fragment-Based and De Novo Design. *Curr. Top. Med. Chem.* **2010**, *10*, 14−32.
(7) Crisman, T. J.; Bender, A.; Milik, M.; Jenkins, J. L.; Scheiber, J.; Sukuru, S. C. K.; Fejzo, J.; Hommel, U.; Davies, J. W.; Glick, M. ‚Virtual Fragment Linking': An Approach to Identify Potent Binders from Low Affinity Fragment Hits. *J. Med. Chem.* **2008**, *51*, 2481−2491.
(8) Hann, M. M.; Leach, A. R.; Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856−864.
(9) Murray, C. W.; Verdonk, M. L.; Rees, D. C. Experiences in Fragment-Based Drug Discovery. *Trends Pharmacol. Sci.* **2012**, *33*, 224−232.
(10) Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-Based Drug Discovery. *J. Med. Chem.* **2004**, *47*, 3463−3482.
(11) Matter, H.; Sotriffer, C. Applications and Success Stories in Virtual Screening. In *Virtual Screening*; Wiley-VCH Verlag GmbH & Co. KGaA: 2011; pp 319−358.
(12) Koeppen, H.; Kriegl, J.; Lessel, U.; Tautermann, C. S.; Wellenzohn, B. Ligand-Based Virtual Screening. In *Virtual Screening*; Wiley-VCH Verlag GmbH & Co. KGaA: 2011; pp 61−85.
(13) Chen, Y.; Shoichet, B. K. Molecular Docking and Ligand Specificity in Fragment-Based Inhibitor Discovery. *Nat. Chem. Biol.* **2009**, *5*, 358−364.
(14) de Graaf, C.; Kooistra, A. J.; Vischer, H. F.; Katritch, V.; Kuijer, M.; Shiroishi, M.; Iwata, S.; Shimamura, T.; Stevens, R. C.; de Esch, I. J.; Leurs, R. Crystal Structure-Based Virtual Screening for Fragment-Like Ligands of the Human Histamine H(1) Receptor. *J. Med. Chem.* **2011**, *54*, 8195−8206.
(15) Sirci, F.; Istyastono, E. P.; Vischer, H. F.; Kooistra, A. J.; Nijmeijer, S.; Kuijer, M.; Wijtmans, M.; Mannhold, R.; Leurs, R.; de Esch, I.; de Graaf, C. Virtual Fragment Screening: Discovery of Histamine H3 Receptor Ligands Using Ligand-Based and Protein-Based Molecular Fingerprints. *J. Chem. Inf. Model.* **2012**, *52*, 3308−3324.
(16) Barelier, S.; Eidam, O.; Fish, I.; Hollander, J.; Figaroa, F.; Nachane, R.; Irwin, J. J.; Shoichet, B. K.; Siegal, G. Increasing Chemical Space Coverage by Combining Empirical and Computational Fragment Screens. *ACS Chem. Biol.* **2014**, *9*, 1528−1535.
(17) Vass, M.; Schmidt, E.; Horti, F.; Keseru, G. M. Virtual Fragment Screening on Gpcrs: A Case Study on Dopamine D3 and Histamine H4 Receptors. *Eur. J. Med. Chem.* **2014**, *77*, 38−46.
(18) Chen, D.; Ranganathan, A.; AP, I. J.; Siegal, G.; Carlsson, J. Complementarity between in Silico and Biophysical Screening Approaches in Fragment-Based Lead Discovery against the a(2a) Adenosine Receptor. *J. Chem. Inf. Model.* **2013**, *53*, 2701−2714.
(19) van Linden, O. P.; Farenc, C.; Zoutman, W. H.; Hameetman, L.; Wijtmans, M.; Leurs, R.; Tensen, C. P.; Siegal, G.; de Esch, I. J. Fragment Based Lead Discovery of Small Molecule Inhibitors for the Epha4 Receptor Tyrosine Kinase. *Eur. J. Med. Chem.* **2012**, *47*, 493−500.
(20) Eckert, H.; Bajorath, J. Molecular Similarity Analysis in Virtual Screening: Foundations, Limitations and Novel Approaches. *Drug Discovery Today* **2007**, *12*, 225−233.
(21) Flower, D. R. On the Properties of Bit String-Based Measures of Chemical Similarity. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 379−386.
(22) Bender, A.; Jenkins, J. L.; Scheiber, J.; Sukuru, S. C.; Glick, M.; Davies, J. W. How Similar Are Similarity Searching Methods? A Principal Component Analysis of Molecular Descriptor Space. *J. Chem. Inf. Model.* **2009**, *49*, 108−119.
(23) Bender, A. Bayesian Methods in Virtual Screening and Chemical Biology. In *Chemoinformatics and Computational Chemical Biology*, Bajorath, J. r., Ed.; Humana Press: 2011; Vol. *672*, Chapter 7, pp 175−196.
(24) Koutsoukas, A.; Paricharak, S.; Galloway, W. R.; Spring, D. R.; Ijzerman, A. P.; Glen, R. C.; Marcus, D.; Bender, A. How Diverse Are Diversity Assessment Methods? A Comparative Analysis and Benchmarking of Molecular Descriptor Space. *J. Chem. Inf. Model.* **2014**, *54*, 230−242.
(25) Grant, J. A.; Gallardo, M. A.; Pickup, B. T. A Fast Method of Molecular Shape Comparison: A Simple Application of a Gaussian Description of Molecular Shape. *J. Comput. Chem.* **1996**, *17*, 1653−1666.
(26) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2006**, *50*, 74−82.
(27) Gillet, V. J.; Willett, P.; Bradshaw, J. Similarity Searching Using Reduced Graphs. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 338−345.
(28) Harper, G.; Bravi, G. S.; Pickett, S. D.; Hussain, J.; Green, D. V. The Reduced Graph Descriptor in Virtual Screening and Data-Driven Clustering of High-Throughput Screening Data. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2145−2156.
(29) Lessel, U.; Wellenzohn, B.; Lilienthal, M.; Claussen, H. Searching Fragment Spaces with Feature Trees. *J. Chem. Inf. Model.* **2009**, *49*, 270−279.
(30) Kiss, R.; Sandor, M.; Gere, A.; Schmidt, E.; Balogh, G. T.; Kiss, B.; Molnar, L.; Lemmen, C.; Keseru, G. M. Discovery of Novel Histamine H4 and Serotonin Transporter Ligands Using the Topological Feature Tree Descriptor. *J. Chem. Inf. Model.* **2012**, *52*, 233−242.
(31) Bajorath, J. Methods for Ligand-Based Virtual Screening. *Front. Med. Chem.* **2009**, *4*, 1.
(32) Holliday, J. D.; Salim, N.; Whittle, M.; Willett, P. Analysis and Display of the Size Dependence of Chemical Similarity Coefficients. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 819−828.
(33) Nicholls, A. What Do We Know and When Do We Know It? *J. Comput. Aided Mol. Des.* **2008**, *22*, 239−255.
(34) Bender, A.; Glen, R. C. A Discussion of Measures of Enrichment in Virtual Screening: Comparing the Information Content of Descriptors with Increasing Levels of Sophistication. *J. Chem. Inf. Model.* **2005**, *45*, 1369−1375.
(35) Rohrer, S. G.; Baumann, K. Impact of Benchmark Data Set Topology on the Validation of Virtual Screening Methods: Exploration and Quantification by Spatial Statistics. *J. Chem. Inf. Model.* **2008**, *48*, 704−718.
(36) Good, A. C.; Oprea, T. I. Optimization of Camd Techniques 3. Virtual Screening Enrichment Studies: A Help or Hindrance in Tool Selection? *J. Comput. Aided Mol. Des.* **2008**, *22*, 169−178.
(37) Jain, A. N.; Nicholls, A. Recommendations for Evaluation of Computational Methods. *J. Comput. Aided Mol. Des.* **2008**, *22*, 133−139.

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044

(38) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.
(39) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (Dud-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582−6594.
(40) Rohrer, S. G.; Baumann, K. Maximum Unbiased Validation (Muv) Data Sets for Virtual Screening Based on Pubchem Bioactivity Data. *J. Chem. Inf. Model.* **2009**, *49*, 169−184.
(41) Vogel, S. M.; Bauer, M. R.; Boeckler, F. M. Dekois: Demanding Evaluation Kits for Objective in Silico Screening–a Versatile Tool for Benchmarking Docking Programs and Scoring Functions. *J. Chem. Inf. Model.* **2011**, *51*, 2650−2665.
(42) Hu, G.; Kuang, G.; Xiao, W.; Li, W.; Liu, G.; Tang, Y. Performance Evaluation of 2d Fingerprint and 3d Shape Similarity Methods in Virtual Screening. *J. Chem. Inf. Model.* **2012**, *52*, 1103−1113.
(43) von Korff, M.; Freyss, J.; Sander, T. Comparison of Ligand- and Structure-Based Virtual Screening on the DUD Data Set. *J. Chem. Inf. Model.* **2009**, *49*, 209−231.
(44) Svensson, F.; Karlen, A.; Skold, C. Virtual Screening Data Fusion Using Both Structure- and Ligand-Based Methods. *J. Chem. Inf. Model.* **2012**, *52*, 225−232.
(45) Tiikkainen, P.; Markt, P.; Wolber, G.; Kirchmair, J.; Distinto, S.; Poso, A.; Kallioniemi, O. Critical Comparison of Virtual Screening Methods against the MUV Data Set. *J. Chem. Inf. Model.* **2009**, *49*, 2168−2178.
(46) Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martinez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. Recognizing Pitfalls in Virtual Screening: A Critical Review. *J. Chem. Inf. Model.* **2012**, *52*, 867−881.
(47) Butkiewicz, M.; Lowe, E. W., Jr.; Mueller, R.; Mendenhall, J. L.; Teixeira, P. L.; Weaver, C. D.; Meiler, J. Benchmarking Ligand-Based Virtual High-Throughput Screening with the Pubchem Database. *Molecules* **2013**, *18*, 735−756.
(48) Lindh, M.; Svensson, F.; Schaal, W.; Zhang, J.; Skold, C.; Brandt, P.; Karlen, A. Toward a Benchmarking Data Set Able to Evaluate Ligand- and Structure-Based Virtual Screening Using Public Hts Data. *J. Chem. Inf. Model.* **2015**, *55*, 343−353.
(49) Mestres, J.; Gregori-Puigjane, E.; Valverde, S.; Sole, R. V. Data Completeness–the Achilles Heel of Drug-Target Networks. *Nat. Biotechnol.* **2008**, *26*, 983−984.
(50) Brianso, F.; Carrascosa, M. C.; Oprea, T. I.; Mestres, J. Cross-Pharmacology Analysis of G Protein-Coupled Receptors. *Curr. Top. Med. Chem.* **2011**, *11*, 1956−1963.
(51) de Graaf, C.; Vischer, H. F.; de Kloe, G. E.; Kooistra, A. J.; Nijmeijer, S.; Kuijer, A.; Verheij, M. H.; England, P.; van Muijlwijk-Koezen, J. E.; Leurs, R.; de Esch, I. J. P. Small and Colourful Tesserae Make Beautiful Mosaics: Fragment-Based Chemogenomics. *Drug Discovery Today* **2013**, *18*, 323−330.
(52) Kooistra, A. J.; Kuhne, S.; de Esch, I. J.; Leurs, R.; de Graaf, C. A Structural Chemogenomics Analysis of Aminergic GPCRs: Lessons for Histamine Receptor Ligand Design. *Br. J. Pharmacol.* **2013**, *170*, 101−126.
(53) Hu, Y.; Bajorath, J. Introduction of Target Cliffs as a Concept to Identify and Describe Complex Molecular Selectivity Patterns. *J. Chem. Inf. Model.* **2013**, *53*, 545−552.
(54) Verheij, M. H.; de Graaf, C.; de Kloe, G. E.; Nijmeijer, S.; Vischer, H. F.; Smits, R. A.; Zuiderveld, O. P.; Hulscher, S.; Silvestri, L.; Thompson, A. J.; van Muijlwijk-Koezen, J. E.; Lummis, S. C.; Leurs, R.; de Esch, I. J. Fragment Library Screening Reveals Remarkable Similarities between the G Protein-Coupled Receptor Histamine H_4 and the Ion Channel Serotonin 5-Ht_{3A}. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 5460−5464.
(55) Albert, J. S.; Blomberg, N.; Breeze, A. L.; Brown, A. J.; Burrows, J. N.; Edwards, P. D.; Folmer, R. H.; Geschwindner, S.; Griffen, E. J.; Kenny, P. W.; Nowak, T.; Olsson, L. L.; Sanganee, H.; Shapiro, A. B. An Integrated Approach to Fragment-Based Lead Generation: Philosophy, Strategy and Case Studies from Astrazeneca's Drug Discovery Programmes. *Curr. Top. Med. Chem.* **2007**, *7*, 1600−1629.
(56) Chen, I. J.; Hubbard, R. E. Lessons for Fragment Library Design: Analysis of Output from Multiple Screening Campaigns. *J. Comput. Aided Mol. Des.* **2009**, *23*, 603−620.
(57) Schuffenhauer, A.; Ruedisser, S.; Marzinzik, A. L.; Jahnke, W.; Blommers, M.; Selzer, P.; Jacoby, E. Library Design for Fragment Based Screening. *Curr. Top. Med. Chem.* **2005**, *5*, 751−762.
(58) Doak, C. D.; Morton, C. J.; Simpson, J. S.; Scanlon, M. J. Design and Evaluation of the Performance of an NMR Screening Fragment Library. *Aust. J. Chem.* **2013**, *66*, 1465−1472.
(59) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. Chembl: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107.
(60) Istyastono, E. P.; Kooistra, A. J.; Vischer, H. F.; Kuijer, M.; Roumen, L.; Nijmeijer, S.; Smits, R. A.; de Esch, I. J.; Leurs, R.; de Graaf, C. Structure-based virtual screening for fragment- like ligands of the G protein-coupled histamine H4 receptor. *Med. Chem Commun.* **2015**, DOI: 10.1039/C5MD00022J.
(61) Smits, R. A.; Lim, H. D.; Hanzer, A.; Zuiderveld, O. P.; Guaita, E.; Adami, M.; Coruzzi, G.; Leurs, R.; de Esch, I. J. Fragment Based Design of New H4 Receptor-Ligands with Anti-Inflammatory Properties in Vivo. *J. Med. Chem.* **2008**, *51*, 2457−2467.
(62) Smits, R. A.; de Esch, I. J.; Zuiderveld, O. P.; Broeker, J.; Sansuk, K.; Guaita, E.; Coruzzi, G.; Adami, M.; Haaksma, E.; Leurs, R. Discovery of Quinazolines as Histamine H4 Receptor Inverse Agonists Using a Scaffold Hopping Approach. *J. Med. Chem.* **2008**, *51*, 7855−7865.
(63) Smits, R. A.; Lim, H. D.; van der Meer, T.; Kuhne, S.; Bessembinder, K.; Zuiderveld, O. P.; Wijtmans, M.; de Esch, I. J.; Leurs, R. Ligand Based Design of Novel Histamine H(4) Receptor Antagonists; Fragment Optimization and Analysis of Binding Kinetics. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 461−467.
(64) Verheij, M. H.; Thompson, A. J.; van Muijlwijk-Koezen, J. E.; Lummis, S. C.; Leurs, R.; de Esch, I. J. Design, Synthesis, and Structure-Activity Relationships of Highly Potent 5-Ht(3) Receptor Ligands. *J. Med. Chem.* **2012**, *55*, 8603−8614.
(65) Thompson, A. J.; Verheij, M. H.; van Muijlwijk-Koezen, J. E.; Lummis, S. C.; Leurs, R.; de Esch, I. J. Structure-Activity Relationships of Quinoxaline-Based 5-HT3A and 5-HT3AB Receptor-Selective Ligands. *ChemMedChem.* **2013**, *8*, 946−955.
(66) Visegrady, A.; Keseru, G. M. Fragment-Based Lead Discovery on G-Protein-Coupled Receptors. *Expert Opin Drug Discov* **2013**, *8*, 811−820.
(67) Thurmond, R. L.; Gelfand, E. W.; Dunford, P. J. The Role of Histamine H1 and H4 Receptors in Allergic Inflammation: The Search for New Antihistamines. *Nat. Rev. Drug Discovery* **2008**, *7*, 41−53.
(68) Simons, F. E. R. Advances in H1-Antihistamines. *New Engl. J. Med.* **2004**, *351*, 2203−2217.
(69) Coruzzi, G.; Adami, M.; Guaita, E.; de Esch, I. J.; Leurs, R. Antiinflammatory and Antinociceptive Effects of the Selective Histamine H4-Receptor Antagonists JNJ7777120 and VUF6002 in a Rat Model of Carrageenan-Induced Acute Inflammation. *Eur. J. Pharmacol.* **2007**, *563*, 240−244.
(70) Cowden, J. M.; Riley, J. P.; Ma, J. Y.; Thurmond, R. L.; Dunford, P. J. Histamine H4 Receptor Antagonism Diminishes Existing Airway Inflammation and Dysfunction Via Modulation of Th2 Cytokines. *Respir. Res.* **2010**, *11*, 86.
(71) Dunford, P. J.; Williams, K. N.; Desai, P. J.; Karlsson, L.; McQueen, D.; Thurmond, R. L. Histamine H4 Receptor Antagonists Are Superior to Traditional Antihistamines in the Attenuation of Experimental Pruritus. *J. Allergy Clin. Immunol.* **2007**, *119*, 176−183.
(72) Hsieh, G. C.; Chandran, P.; Salyers, A. K.; Pai, M.; Zhu, C. Z.; Wensink, E. J.; Witte, D. G.; Miller, T. R.; Mikusa, J. P.; Baker, S. J.; Wetter, J. M.; Marsh, K. C.; Hancock, A. A.; Cowart, M. D.; Esbenshade, T. A.; Brioni, J. D.; Honore, P. H4 Receptor Antagonism Exhibits Anti-Nociceptive Effects in Inflammatory and Neuropathic Pain Models in Rats. *Pharmacol., Biochem. Behav.* **2010**, *95*, 41−50.
(73) Lim, H. D.; Smits, R. A.; Leurs, R.; De Esch, I. J. The Emerging Role of the Histamine H4 Receptor in Anti-Inflammatory Therapy. *Curr. Top. Med. Chem.* **2006**, *6*, 1365−1373.

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044

(74) Rossbach, K.; Wendorff, S.; Sander, K.; Stark, H.; Gutzmer, R.; Werfel, T.; Kietzmann, M.; Baumer, W. Histamine H4 Receptor Antagonism Reduces Hapten-Induced Scratching Behaviour but Not Inflammation. *Exp. Dermatol.* **2009**, *18*, 57−63.
(75) Thurmond, R. L.; Desai, P. J.; Dunford, P. J.; Fung-Leung, W. P.; Hofstra, C. L.; Jiang, W.; Nguyen, S.; Riley, J. P.; Sun, S.; Williams, K. N.; Edwards, J. P.; Karlsson, L. A Potent and Selective Histamine H4 Receptor Antagonist with Anti-Inflammatory Properties. *J. Pharmacol. Exp. Ther.* **2004**, *309*, 404−413.
(76) Derkach, V.; Surprenant, A.; North, R. A. 5-HT3 Receptors Are Membrane Ion Channels. *Nature* **1989**, *339*, 706−709.
(77) Aapro, M. S. 5-Ht3 Receptor Antagonists. An Overview of Their Present Status and Future Potential in Cancer Therapy-Induced Emesis. *Drugs* **1991**, *42*, 551−568.
(78) Hu, Y.; Bajorath, J. Growth of Ligand-Target Interaction Data in Chembl Is Associated with Increasing and Activity Measurement-Dependent Compound Promiscuity. *J. Chem. Inf. Model.* **2012**, *52*, 2550−2558.
(79) Chen, B.; Mueller, C.; Willett, P. Combination Rules for Group Fusion in Similarity-Based Virtual Screening. *Mol. Inf.* **2010**, *29*, 533−541.
(80) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of Fingerprint-Based Methods for Virtual Screening Using Multiple Bioactive Reference Structures. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1177−1185.
(81) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Enhancing the Effectiveness of Similarity-Based Virtual Screening Using Nearest-Neighbor Information. *J. Med. Chem.* **2005**, *48*, 7049−7054.
(82) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. New Methods for Ligand-Based Virtual Screening: Use of Data Fusion and Machine Learning to Enhance the Effectiveness of Similarity Searching. *J. Chem. Inf. Model.* **2006**, *46*, 462−470.
(83) Schuffenhauer, A.; Floersheim, P.; Acklin, P.; Jacoby, E. Similarity Metrics for Ligands Reflecting the Similarity of the Target Proteins. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 391−405.
(84) Tovar, A.; Eckert, H.; Bajorath, J. Comparison of 2d Fingerprint Methods for Multiple-Template Similarity Searching on Compound Activity Classes of Increasing Structural Diversity. *ChemMedChem* **2007**, *2*, 208−217.
(85) Whittle, M.; Gillet, V. J.; Willett, P.; Alex, A.; Loesel, J. Enhancing the Effectiveness of Virtual Screening by Fusing Nearest Neighbor Lists: A Comparison of Similarity Coefficients. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1840−1848.
(86) Whittle, M.; Gillet, V. J.; Willett, P.; Loesel, J. Analysis of Data Fusion Methods in Virtual Screening: Similarity and Group Fusion. *J. Chem. Inf. Model.* **2006**, *46*, 2206−2219.
(87) Willett, P., Fusing Similarity Rankings in Ligand-Based Virtual Screening. *Comput. Struct. Biotechnol. J.* **2013**, *5*.
(88) Zhang, Q.; Muegge, I. Scaffold Hopping through Virtual Screening Using 2d and 3d Similarity Descriptors: Ranking, Voting, and Consensus Scoring. *J. Med. Chem.* **2006**, *49*, 1536−1548.
(89) Willett, P. Combination of Similarity Rankings Using Data Fusion. *J. Chem. Inf. Model.* **2013**, *53*, 1−10.
(90) Baber, J. C.; Shirley, W. A.; Gao, Y.; Feher, M. The Use of Consensus Scoring in Ligand-Based Virtual Screening. *J. Chem. Inf. Model.* **2006**, *46*, 277−288.
(91) Muchmore, S. W.; Debe, D. A.; Metz, J. T.; Brown, S. P.; Martin, Y. C.; Hajduk, P. J. Application of Belief Theory to Similarity Data Fusion for Use in Analog Searching and Lead Hopping. *J. Chem. Inf. Model.* **2008**, *48*, 941−948.
(92) Wiggers, H. J.; Rocha, J. R.; Cheleski, J.; Montanari, C. A. Integration of Ligand- and Target-Based Virtual Screening for the Discovery of Cruzain Inhibitors. *Mol. Inf.* **2011**, *30*, 565−578.
(93) Costanzi, S. On the Applicability of Gpcr Homology Models to Computer-Aided Drug Discovery: A Comparison between in Silico and Crystal Structures of the Beta2-Adrenergic Receptor. *J. Med. Chem.* **2008**, *51*, 2907−2914.
(94) Riniker, S.; Fechner, N.; Landrum, G. A. Heterogeneous Classifier Fusion for Ligand-Based Virtual Screening: Or, How Decision Making by Committee Can Be a Good Thing. *J. Chem. Inf. Model.* **2013**, *53*, 2829−2836.
(95) Todeschini, R.; Consonni, V.; Xiang, H.; Holliday, J.; Buscema, M.; Willett, P. Similarity Coefficients for Binary Chemoinformatics Data: Overview and Extended Comparison Using Simulated and Real Datasets. *J. Chem. Inf. Model.* **2012**, *52*, 2884−2901.
(96) Kooistra, A. J.; Binsl, T. W.; van Beek, J. H.; de Graaf, C.; Heringa, J. Electron Density Fingerprints (EDprints): Virtual Screening Using Assembled Information of Electron Density. *J. Chem. Inf. Model.* **2010**, *50*, 1772−1780.
(97) Kirchmair, J.; Ristic, S.; Eder, K.; Markt, P.; Wolber, G.; Laggner, C.; Langer, T. Fast and Efficient in Silico 3D Screening: Toward Maximum Computational Efficiency of Pharmacophore-Based and Shape-Based Approaches. *J. Chem. Inf. Model.* **2007**, *47*, 2182−2196.
(98) de Graaf, C.; Rognan, D. Selective Structure-Based Virtual Screening for Full and Partial Agonists of the Beta2 Adrenergic Receptor. *J. Med. Chem.* **2008**, *51*, 4978−4985.
(99) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.
(100) Xu, Y. J.; Johnson, M. Using Molecular Equivalence Numbers to Visually Explore Structural Features That Distinguish Chemical Libraries. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 912−926.
(101) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of Mdl Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273−1280.
(102) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(103) Vogt, M.; Stumpfe, D.; Geppert, H.; Bajorath, J. Scaffold Hopping Using Two-Dimensional Fingerprints: True Potential, Black Magic, or a Hopeless Endeavor? Guidelines for Virtual Screening. *J. Med. Chem.* **2010**, *53*, 5707−5715.
(104) Leurs, R.; Vischer, H. F.; Wijtmans, M.; de Esch, I. J. En Route to New Blockbuster Anti-Histamines: Surveying the Offspring of the Expanding Histamine Receptor Family. *Trends Pharmacol. Sci.* **2011**, *32*, 250−257.
(105) Thompson, A. J.; Lummis, S. C. The 5-Ht3 Receptor as a Therapeutic Target. *Expert Opin. Ther. Targets* **2007**, *11*, 527−540.
(106) Bajorath, J.; Peltason, L.; Wawer, M.; Guha, R.; Lajiness, M. S.; Van Drie, J. H. Navigating Structure-Activity Landscapes. *Drug Discovery Today* **2009**, *14*, 698−705.
(107) Maggiora, G. M. On Outliers and Activity Cliffs–Why Qsar Often Disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535.
(108) Chen, B.; Harrison, R. F.; Papadatos, G.; Willett, P.; Wood, D. J.; Lewell, X. Q.; Greenidge, P.; Stiefl, N. Evaluation of Machine-Learning Methods for Ligand-Based Virtual Screening. *J. Comput. Aided Mol. Des.* **2007**, *21*, 53−62.
(109) Geppert, H.; Horvath, T.; Gartner, T.; Wrobel, S.; Bajorath, J. Support-Vector-Machine-Based Ranking Significantly Improves the Effectiveness of Similarity Searching Using 2d Fingerprints and Multiple Reference Compounds. *J. Chem. Inf. Model.* **2008**, *48*, 742−746.
(110) Hammann, F.; Gutmann, H.; Baumann, U.; Helma, C.; Drewe, J. Classification of Cytochrome P(450) Activities Using Machine Learning Methods. *Mol. Pharmaceutics* **2009**, *6*, 1920−1926.
(111) Reynolds, C. R.; Amini, A. C.; Muggleton, S. H.; Sternberg, M. J. E. Assessment of a Rule-Based Virtual Screening Technology (Inddex) on a Benchmark Data Set. *J. Phys. Chem. B* **2012**, *116*, 6732−6739.
(112) Riniker, S.; Fechner, N.; Landrum, G. A. Heterogeneous Classifier Fusion for Ligand-Based Virtual Screening: Or, How Decision Making by Committee Can Be a Good Thing. *J. Chem. Inf. Model.* **2013**, *53*, 2829−2836.
(113) Renner, S.; Schneider, G. Scaffold-Hopping Potential of Ligand-Based Similarity Concepts. *ChemMedChem.* **2006**, *1*, 181−185.
(114) Hessler, G.; Baringhaus, K.-H. The Scaffold Hopping Potential of Pharmacophores. *Drug Discovery Today: Technologies* **2010**, *7*, e263−e269.
(115) Martin, Y. C.; Muchmore, S. Beyond Qsar: Lead Hopping to Different Structures. *QSAR & Combinatorial Science* **2009**, *28*, 797−801.

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044

(116) Schneider, G.; Schneider, P.; Renner, S. Scaffold-Hopping: How Far Can You Jump? *QSAR & Combinatorial Science* **2006**, *25*, 1162−1171.
(117) Sastry, G. M.; Inakollu, V. S. S.; Sherman, W. Boosting Virtual Screening Enrichments with Data Fusion: Coalescing Hits from Two-Dimensional Fingerprints, Shape, and Docking. *J. Chem. Inf. Model.* **2013**, *53*, 1531−1542.

DOI: 10.1021/ci500694c
J. Chem. Inf. Model. 2015, 55, 1030−1044