JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

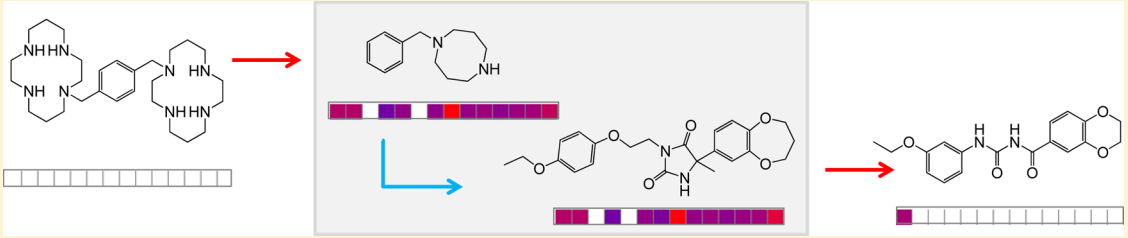
pubs.acs.org/jcim

Bioturbo Similarity Searching: Combining Chemical and Biological Similarity To Discover Structurally Diverse Bioactive Molecules

Anne Mai Wassermann,* Eugen Lounkine, and Meir Glick

In Silico Lead Discovery, Novartis Institutes for Biomedical Research Inc., 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

Supporting Information



ABSTRACT: Virtual screening using bioactivity profiles has become an integral part of currently applied hit finding methods in pharmaceutical industry. However, a significant drawback of this approach is that it is only applicable to compounds that have been biologically tested in the past and have sufficient activity annotations for meaningful profile comparisons. Although bioactivity data generated in pharmaceutical institutions are growing on an unprecedented scale, the number of biologically annotated compounds still covers only a minuscule fraction of chemical space. For a newly synthesized compound or an isolated natural product to be biologically characterized across multiple assays, it may take a considerable amount of time. Consequently, this chemical matter will not be included in virtual screening campaigns based on bioactivity profiles. To overcome this problem, we herein introduce bioturbo similarity searching that uses chemical similarity to map molecules without biological annotations into bioactivity space and then searches for biologically similar compounds in this reference system. In benchmark calculations on primary screening data, we demonstrate that our approach generally achieves higher hit rates and identifies structurally more diverse compounds than approaches using chemical information only. Furthermore, our method is able to discover hits with novel modes of inhibition that traditional 2D and 3D similarity approaches are unlikely to discover. Test calculations on a set of natural products reveal the practical utility of the approach for identifying novel and synthetically more accessible chemical matter.

INTRODUCTION

Virtual screening[1] has its origins in chemical similarity searching that uses a single chemical reference structure to find structurally similar molecules.[2] The idea behind this approach is that overall structurally similar molecules also tend to have similar biological activities, following the similarity-property principle formulated by Johnson and Maggiora in 1990.[3] However, this approach faces two major caveats: 1) The ability of the approach to discover novel active chemotypes is limited because, by definition, chemically similar compounds are prioritized. 2) Although the similarity-property principle is intuitive and widely accepted in medicinal chemistry,[4] it is also well-known that, in some instances, only small structural modifications can alter the bioactivity of a compound dramatically or render it completely inactive.[5] Such compound pairs that show high structural but low bioactivity similarity are generally referred to as activity cliffs[6] and constitute a major limitation for ligand structure-based virtual screening methods.[5] Over the past years, conceptually different approaches that circumvent these problems have been introduced: bioactivity-based methods describe compounds via their interactions with the proteome. In this case, molecular similarity for a pair of compounds is calculated by comparing their observed effects on defined targets or cell types.[7,8] In a pioneering study conducted at the NCI, compounds were represented by the growth inhibition effects that they showed across a panel of 60 different cancer cell lines.[9] It was demonstrated that compounds with similar mechanism of cell growth inhibition tended to produce similar profile vectors, and growth inhibition patterns were subsequently used to predict mechanisms-of-action for a variety of compounds.[10] So-termed *affinity fingerprints* introduced by Kauvar et al. compared molecules on the basis of pIC50 values against eight selected protein targets and were applied to predict novel ligand-protein interactions.[11] Similarly, Fliri et al. made use of the BioPrint database[12] of Cerep to encode about 1,600 compounds by their percent inhibition values from 92 ligand-binding assays and, for example, used these so-called *biospectra* to identify agonist and antagonist effect profiles of medicinal agents.[13,14] Cheng et al. used PubChem bioassay data to relate cellular effects of 37 compounds tested against the NCI-60 cell line panel to protein targets.[15] In many instances, compounds that had similar cellular

Received: December 17, 2012
Published: March 5, 2013

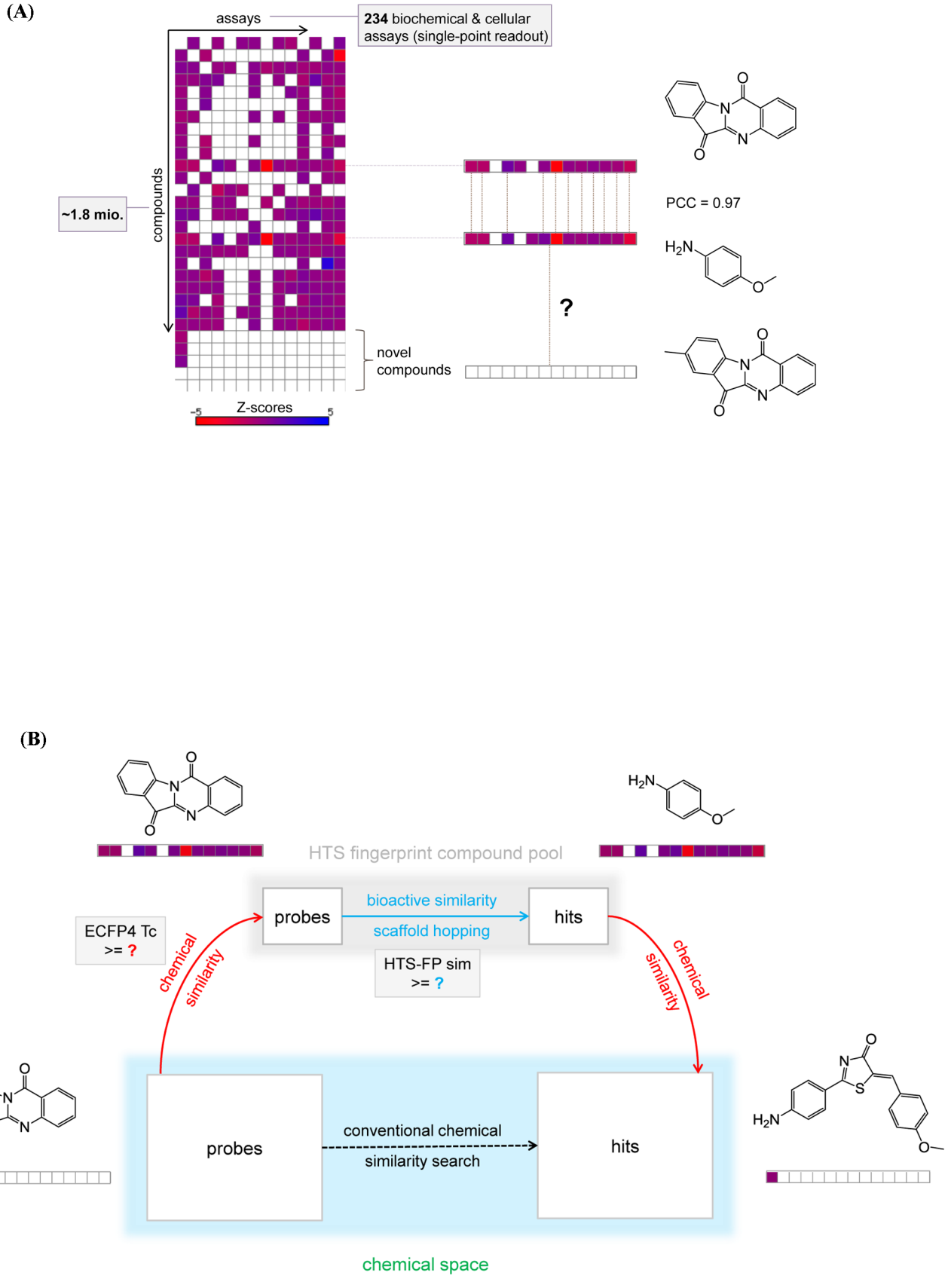


Figure 1. continued

dx.doi.org/10.1021/ci300607r | J. Chem. Inf. Model. 2013, 53, 692−703

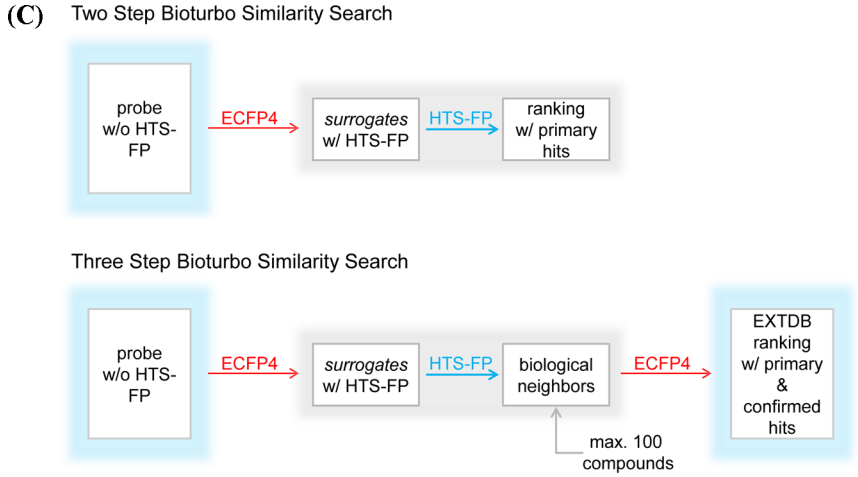


Figure 1. HTS Fingerprints and Bioturbo Similarity Searching. (**A**) A schematic depiction of the Novartis HTS-FP data landscape is shown. Approximately 1.8 million compounds are represented in bioactivity space by the Z-scores that they have historically obtained in 234 HTS assays run at Novartis. Not all compounds have been measured in all assays, and, hence, for many compounds, Z-score profiles are incomplete. For molecular similarity assessment, a pair of compounds is compared by calculating the Pearson correlation coefficient (PCC) for the series of assays that their HTS-FPs have in common. However, using this method, an open question is how to calculate molecular similarities for newly added compounds that have no reported Z-score profile. (**B**) General principles of bioturbo similarity searching are outlined. For a compound without HTS-FP annotation (left), a chemical similarity search is run against the HTS-FP compound deck. Structurally similar compounds are then used as representatives for the starting compound in HTS-FP calculations to find molecules with a similar bioactivity profile. This comparison does not involve chemical features and often leads to scaffold hops, i.e., the identification of compounds with different core structures. Molecules with similar bioactivity profiles can then be used to carry out another chemical similarity search to retrieve structurally similar compounds outside of the HTS-FP compound deck. As shown, mapping compounds via chemical similarity into bioactivity space makes the HTS-FP approach applicable to a much larger fraction of chemical space. (**C**) Steps in the two different BSS approaches are schematically depicted.

effects also showed similar interaction patterns with relevant protein targets. Compounds that were clustered together by bioactivity profiles showed varying degrees of structural similarity,[15] emphasizing the different views taken on molecular similarity by biological and chemical descriptors.[8]

Finally, the concept of bioactivity-based profile comparisons was taken to an unrivaled magnitude by the introduction of *HTS fingerprints* (HTS-FPs) by Petrone et al.[16] HTS-FPs report single-point readouts from about 200 biochemical and cellular HTS assays that have historically been run at Novartis for ∼1.8 million different compounds in the screening deck (Figure 1A). In exhaustive benchmark calculations it was shown that HTS-FP-based virtual screening achieved excellent performance both in the number and structural diversity of recovered active compounds. However, although HTS-FPs enable compound comparisons on an unprecedented scale, they still encounter the same difficulties as any other bioactivity-based method: compounds that do not have an HTS-FP, for example, because they have been newly added to the Novartis screening collection, cannot be included in HTS-FP-based virtual screening. Previous attempts to overcome this limitation of bioactivity-based methods include the invention of virtual bioactivity fingerprints that use predicted values for ligand-target interactions in profile vectors instead of experimental assay data.[17−19] These predicted values were, for example, likelihood scores generated by a multiclass 2D Bayes classifier[17] or docking scores against a reference panel of protein targets.[18,19] However, these approaches face limitations as well; e.g, docking-based fingerprints require the availability of 3D protein structures and a recent critical assessment of docking programs and scoring functions revealed that none of the investigated scoring functions was able to make sufficiently accurate predictions of ligand-target affinity.[20]

Therefore, we asked the question: how can we possibly make use of real assay data-based HTS-FPs for compounds that have not been profiled yet? Herein, we introduce an approach that allows us to exploit the power of HTS-FPs for compounds that have not been previously tested against the Novartis assay panel (Figure 1B). A key aspect of this approach is the combination of chemical and HTS-FP similarity searching in an iterative fashion. As working hypothesis we assume that compounds that are structurally related to a reference compound can be used as its substitute in HTS-FP searching. Because our approach is conceptually related to turbo similarity searching,[21,22] which uses close structural neighbors of a probe compound for a second round of chemical similarity searching, we refer to this approach as *bioturbo similarity searching* (BSS).

As reported in the following, we used 14 benchmark compound data sets to parametrize and evaluate our method. We show that, when using molecules with moderately high chemical similarity to a reference structure in HTS-FP calculations, BSS performs generally better than conventional chemical similarity searching, in particular in retrieving chemically more diverse hits. Furthermore, we report the application of BSS to more than 300 natural products. Starting from complex reference structures, we demonstrate the ability of BSS to find synthetic compounds with desired bioactivity. This has practical applications in drug discovery where a natural product is often used as a proof of concept to answer a well-defined biological question, e.g., whether the inhibition of a certain target will produce a specific phenotype. If the natural product shows the anticipated or desired effect, a follow-up is done with a synthetic compound that can be optimized into a drug candidate.

MATERIALS AND METHODS

HTS Fingerprints. For all compounds in the Novartis screening collection, HTS-FPs are calculated, i.e., each

dx.doi.org/10.1021/ci300607r | J. Chem. Inf. Model. 2013, 53, 692−703

compound is represented by a numeric vector where each position reports its activity in a specific HTS assay.[16] For the HTS-FP calculations reported herein, 234 different biochemical and cellular assays are considered in the generation of HTS-FPs, and activity outcomes are reported in form of Z-scores. As the Novartis screening collection is constantly changing and not all compounds have historically been measured in all HTS assays, HTS-FPs often contain missing values. For the comparison of two molecules via their HTS fingerprints, the subset of assays for which both compounds are annotated with Z-scores is identified, and a Pearson correlation coefficient (PCC) between corresponding activity values is calculated (Figure 1A).[16] The higher the PCC, the better is the linear correlation between activity outcomes for two molecules across the HTS assay panel. However, as we noticed a clear trend toward higher PCCs for smaller numbers of shared assays, a normalization procedure is applied to account for different numbers of assays common to a pair of compounds, converting each PCC into a frequency score that reports how often an equally good or better correlation has been observed for the same number of assays in a reference panel of ∼165 million pairwise HTS-FP comparisons. For example, a PCC of at least 0.7 is observed for 10.9% of all pairs of compounds sharing five assays, whereas this fraction is only 0.0038% when 150 assays are shared. Herein, the frequency score is used as HTS-FP similarity measure: the smaller the score, the higher is the biological similarity of two molecules.

Bioturbo Similarity Searching. *(a). Two Step BSS.* For a bioactive reference compound that does not have an HTS-FP, a chemical similarity search using ECFP4 fingerprints[23] and the Tanimoto coefficient[2] (Tc) is applied to find structurally similar compounds above a given Tc threshold t in the Novartis screening collection (step 1). ECFP4 fingerprints in combination with the Tanimoto coefficient are a frequently employed ligand-based virtual screening method that showed state-of-the-art performance in many previously reported fingerprint and similarity metric benchmark studies.[24−27] Other fingerprints and similarity metrics could also be employed in this first step; however, a re-evaluation of different fingerprints and similarity coefficients was outside of the scope of our study that rather aimed at the exploration of chemical and biological similarity in a unified virtual screening protocol, as further outlined in the following. The identified structural neighbors from step 1 are then compared to the remainder of the screening collection using their HTS-FPs. Each compound in the collection is compared to all compounds in the reference set in a pairwise manner and annotated with the lowest frequency score, i.e., the maximal biological similarity to one of the reference compounds, corresponding to a 1-nearest neighbor (1-NN) search. Then, a ranking is generated by sorting compounds in order of ascending frequency scores. However, for some compounds from the screening deck that share less than three assays with any identified structural neighbor, frequency scores cannot be calculated in a meaningful way. These compounds are appended to the end of the ranking in order of decreasing chemical similarity to the original reference compound. In the study reported herein, we only considered virtual screening trials where at least half of the screening collection could be scored by HTS-FP similarity. Furthermore, we required that at least one reference compound for HTS-FP comparisons had been tested in ten or more assays.

(b). Three Step BSS. Using the ranking of the screening collection obtained by Two Step BSS, all compounds below a given frequency score f are identified and used as a reference set to rank compounds in an external database according to their chemical similarity, again using ECFP4 fingerprints and the Tanimoto coefficient. Similarly to the HTS-FP calculations, a 1-NN strategy is used to sort compounds, i.e., compounds are ranked in order of their maximal similarity to one of the reference compounds. In cases where more than 100 compounds fall below a given threshold, only the 100 compounds with the best, i.e., smallest, frequency scores are used as reference compounds to not make search calculations computationally too expensive. This decision is further supported by turbo similarity search results reported by Hert et al. who found that building reference sets of more than 100 compounds led to decreased search performance.[21]

Figure 1C shows schematic illustrations of the Two Step and Three Step BSS approaches.

Data Sets and Calculations. To benchmark the performance of bioturbo similarity searching and determine appropriate thresholds for the Tanimoto coefficient t and frequency score f in steps 2 and 3, respectively, 14 different compound data sets were assembled from external (ChEMBL,[28] GVK[29]) or internal sources. Each set consisted of 20 inhibitors active against a given protein target or, in one case, a cellular pathway, as further detailed in Table 1. Following the classification scheme of the UniProt[30] database, protein targets belonged to 11 different protein families, including GPCRs, kinases, and other enzymes. The 280 compounds in these data sets served as starting compounds for BSS. If available, compounds that had not been screened at Novartis and did not have an HTS-FP were used. However, for some data sets, activity data were too sparse to exclusively select compounds that were not part of the Novartis screening collection (∼1.8 million compounds). In these cases, we used compounds from the screening deck as references but removed them from the collection in our simulated virtual screening trials.

Table 1. Benchmark Data Sets[a]

	Two Step BSS		Three Step BSS	
target	#hits	#scaffolds	#hits	#scaffolds
Methyltransferase 1	8068	4991	4788	3170
Methyltransferase 2	24567	11059	15991	8800
Deacetylase	4140	2922	12249	5963
Ser Protease	7682	4779	6672	3893
Cys Protease	37	30	2455	1190
Tyr Kinase	12494	7326	20997	10321
Ser/Thr Kinase 1	18833	10653	26080	12673
Ser/Thr Kinase 2	22996	10749	23096	10839
Ser/Thr Kinase 3	29451	13033	41132	16750
Ser/Thr Kinase 4	11160	5824	13747	6086
GPCR A 1	2896	1917	1882	1127
GPCR A 2	35924	20282	19800	12433
GPCR C	5003	3079	7908	3941
Cellular Pathway	43136	21652		

[a]For the 14 benchmark data sets used in our study, masked target names and the number of hits for Two and Three Step BSS are reported. Furthermore, numbers of different Bemis and Murcko scaffolds calculated from the hit compounds are provided. Since the inhibition of the cellular pathway cannot be traced back to a single protein target, no hits are reported for this benchmark set and Three Step BSS as only defined ligand-protein interactions were extracted from EXTDB.

dx.doi.org/10.1021/ci300607r | J. Chem. Inf. Model. 2013, 53, 692−703

Table 2. Two Step BSS − Enrichment Factors[a]

		Tc threshold t											
		TSS	BSS										
target	SIM	0.70	0.45	0.50	0.55	0.6	0.65	0.70	0.75	0.80	0.85	0.90	0.95
Methyltransferase 1	2.00	2.14	4.75	4.54	5.73	6.65	6.09	7.06	8.17	7.92	8.70	9.17	8.85
Methyltransferase 2	1.80	2.10	5.19	6.22	6.60	7.24	7.61	9.64	11.15	11.32	2.78	2.78	4.13
Deacetylase	1.89		1.27	1.33	1.63	1.89							
Ser Protease	1.80	0.61	0.68	0.68	0.64	0.45	0.56	1.05	0.50				
Cys Protease	0.98	1.75	0.74	0.49	0.26	0.29	0.29	0.00	0.00	0.00	0.00	0.00	0.00
Tyr Kinase	1.32		5.05	4.13	3.87	3.49	2.59						
Ser/Thr Kinase 1	1.95	1.49	4.46	5.23	4.76	4.12	6.56	10.04	10.02	10.02	10.13	10.13	10.13
Ser/Thr Kinase 2	1.74	3.49	3.75	6.00	7.01	9.65	12.01	8.42					
Ser/Thr Kinase 3	3.41		5.56	2.17	2.83	4.13							
Ser/Thr Kinase 4	1.58	1.49	8.49	9.78	10.59	13.84	16.12	18.71	18.85	0.86			
GPCR A 1	0.88	0.83	0.88	0.97	0.93	0.95	0.97	1.02	1.16	0.91	0.44		
GPCR A 2	0.86	0.75	1.65	1.67	1.89	2.04	2.14	1.46	1.48	1.05	1.38	1.38	
GPCR C	2.55	0.78	1.61	1.74	1.17	0.89	0.91	0.64	0.51	0.51	0.51	0.51	0.51
Cellular Pathway	1.61	1.45	2.58	2.94	2.85	3.14	2.80	2.85	4.48	3.84	5.27		
average	1.74	1.54	3.33	3.42	3.63	4.20	4.89	5.54	5.63	4.05	3.65	4.00	4.73

[a]For our 14 benchmark data sets, average enrichment factors obtained for a selection set size of 10,000 compounds using standard similarity searching (SIM), turbo similarity searching (TSS), or BSS with variable Tc thresholds t are provided. To make enrichment factors directly comparable, only probes for which BSS could be carried out were considered in the calculation of average values for standard similarity searching. Average enrichment factors in the bottom row are obtained by averaging mean enrichment factors of all data sets.

For each target or pathway in our benchmark set, an HTS campaign had been run at Novartis. For the evaluation of Two Step BSS, hits were defined as compounds in the Novartis screening collection with a Z-score ≤ -3 for the corresponding assay. As shown in Table 1, hit numbers were very variable for the selected benchmark sets ranging from 37 to 43,136 active compounds. To avoid bias, for each data set, the respective assay was excluded from HTS-FP comparisons in our simulated virtual screening trials.

For the last step in Three Step BSS, we used a database of ~2.7 million compounds containing dose−response bioactivity-annotated structures from public and proprietary sources. This database is in the following referred to as EXTDB (EXTernal DataBase). For performance evaluation, all molecules with a K_i or IC50 value ≤ 5 μM for a given target were considered hits. Furthermore, as molecule structures in the screening collection and EXTDB partly overlapped, compounds that had been assayed and fulfilled the hit criterion for Two Step BSS (Z-score ≤ -3) were also considered active. To assess the chemical diversity of hit compounds, scaffolds were calculated according to Bemis and Murcko[31] (Table 1).

To assess the performance of BSS, enrichment factors of hits over random compound subset selection, receiver operating characteristic (ROC) scores, and numbers of retrieved scaffolds were calculated and compared to results obtained by conventional (one step) chemical similarity searching and turbo similarity searching as introduced by Hert et al.[21]

In addition to our 14 HTS benchmark sets, we assembled 304 natural products, i.e., compounds purified from plant, bacterial, and fungal sources, from the GVK and ChEMBL databases. We only collected natural product compounds that were active against one or more human targets with an activity (IC50, K_i) of ≤ 5 μM. Furthermore, we only considered targets for which at least 500 ligands were available in EXTDB to enable a meaningful performance evaluation of BSS. Overall, the 304 natural products were active against 419 different targets, forming 1,444 different target-ligand pairs in total. For each target-ligand combination, Three Step BSS was carried out using the thresholds t and f that performed best for our benchmark data sets. Moreover, 73 of the 304 probes inhibited targets for which primary assays had been run at Novartis. For these compounds that were overall active against 43 different targets, we also evaluated the performance of Two Step BSS. If a natural product extracted from GVK or ChEMBL was part of the Novartis screening deck, its bioactivity profile was removed from the collection, and the compound was treated as if no HTS-FP was available, as done previously for our benchmark sets.

RESULTS AND DISCUSSION

The major aim of our study has been to determine whether the potential of HTS fingerprints to retrieve structurally diverse bioactive compounds can also be extended to compounds that are not part of the Novartis screening collection and, consequently, do not have an HTS-FP. We evaluated to what extent chemical similarity can be used to find surrogate compounds with HTS-FPs that are subsequently used in bioactivity-based virtual screening. We evaluated BSS using 14 benchmark sets containing tested active and inactive compounds.

Two Step BSS Parameterization and Performance Assessment. For each of our 14 benchmark data sets, we conducted 20 different virtual screening trials that used a single reference compound as a starting point. This probe compound was compared to the Novartis screening collection, and all compounds exceeding an ECFP4 Tc similarity threshold t to the probe compound were included in a reference set for subsequent HTS-FP comparisons. Our goal was to find a good balance between the bioactivity similarity of the HTS-FP annotated compounds to the original reference structure (very close structural analogs are more likely to have a suitable surrogate bioactivity profile) and the number of compounds available for HTS-FP searching (only a few reference compounds have close structural analogs with HTS-FPs). For that, we evaluated 11 different similarity thresholds t ranging from 0.45 to 0.95 for the first chemical similarity step.

dx.doi.org/10.1021/ci300607r | J. Chem. Inf. Model. 2013, 53, 692−703

Enrichment Factors. For all 14 benchmark data sets, we have compared the number of active molecules retrieved using different thresholds t to the number of hits that would be expected for a random selection of 5,000, 10,000, or 20,000 compounds (Table 2). For comparison we also calculated enrichment factors for a conventional similarity search that ranks the screening collection according to its ECFP4 similarity to the probe (Table 2, Figure 2A). Strikingly, BSS generally outperformed conventional similarity searching: even with the worst performing choice of 0.45 for the threshold t, almost twice as many hits were retrieved with BSS compared to similarity searching at a selection set size of 10,000 compounds (Figure 2A). The best average enrichment factor obtained for a data set using standard similarity searching was 3.4 for Serine/Threonine Kinase 3. By contrast, the best BSS enrichment was 18.9-fold (Tc 0.75, Serine/Threonine Kinase 4), highlighting the potential of BSS to substantially enhance search performance over random selection. Best search performance was obtained by setting t to 0.7 or 0.75. It is likely that lower thresholds select too many compounds that are too dissimilar to be a suitable surrogate for the starting compound in HTS-FP calculations, i.e., too many *in silico* false positives are used as references. On the other hand, setting a higher threshold and limiting the HTS-FP reference pool to very few compounds might not fully exploit the biological information that can be gained from additional structural neighbors. At t = 0.75, BSS could only be conducted for 16.1% of the 280 probes because, for most starting compounds, no structural neighbor exceeding that similarity level was available (Figure 2B). Hence, we decided to use a similarity cutoff of 0.7 for future calculations. This cutoff performed second best overall (Figure 2A) and was applicable to 23.9%, i.e., 67 probes in our benchmark system. We noticed that applicability also varied among classes: for example, for each of the targets Deacetylase, Tyrosine Kinase, and Serine/Threonine Kinase 3, none of the 20 probes had a structural neighbor at this similarity threshold, whereas for both the Cellular Pathway and Cysteine Protease data sets, 14 of 20 probes showed a similarity of at least 0.7 to one or more compounds in the Novartis screening collection (Supplementary Table S1). As reported in Table 2, using this threshold and selecting the 10,000 compounds top-ranked by BSS, enrichment factors between 7.1 and 18.7 were obtained for different methyltransferases and kinases, an enrichment factor of 2.9 for the cellular pathway and enrichment factors centering around 1 (i.e., similar performance to random selection) for GPCRs and the serine protease included in our study. For the cysteine protease data set, no hit was retrieved among the top 10,000 compounds. However, this can be attributed to the fact that, for this target, only 37 primary hits were available in the screening collection, making it a rather unusual test case.

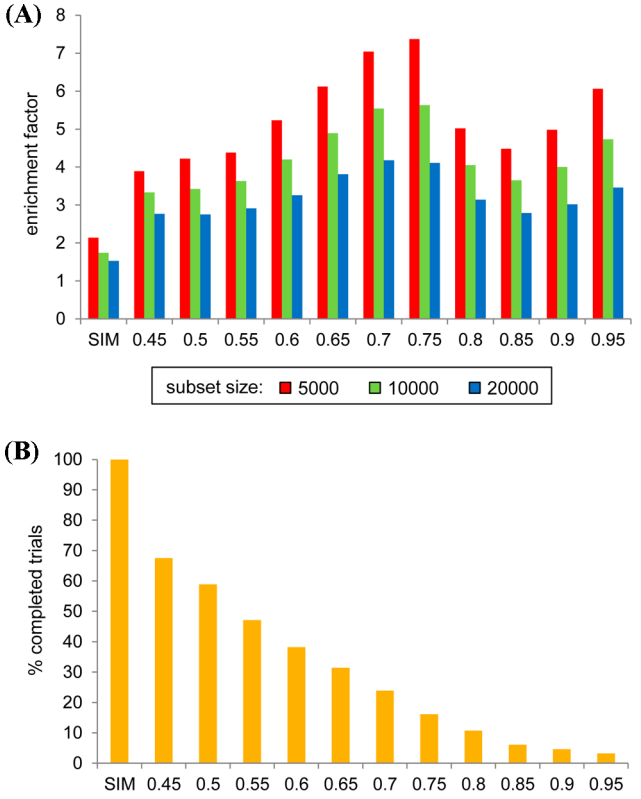


Figure 2. Two Step BSS – Performance Evaluation. (**A**) Enrichment factors averaged over all 14 benchmark data sets are displayed for selection set sizes of 5,000, 10,000, and 20,000 compounds. Two Step BSS is carried out with the Tc threshold t incrementally increased in steps of 0.05 from 0.45 to 0.95 and is compared to standard chemical similarity searching (SIM). Shown enrichment factors are calculated by averaging mean enrichment factors of all data sets. (**B**) For each Tc threshold t, we report the percentage of completed virtual screening trials, i.e., the percentage of probes for which a structural neighbor above the similarity cutoff was available in the Novartis screening collection that could be used as a reference in HTS-FP calculations. Standard similarity searching (SIM) that does not require the presence of structurally similar compounds in the screening collection can always be carried out.

ROC Scores. We also calculated the area under the ROC curve (AUC) to evaluate the performance of BSS over the entire range of the screening collection and not only for a subset of top-ranked compounds (Supplementary Table 2). Setting t to 0.7 achieved overall best ROC scores, although ROC scores were generally modest (~0.57). However, standard similarity searching using chemical information only yielded average ROC scores of ~0.5, which essentially corresponds to a random ranking of the screening collection. This poor performance of standard ligand-based similarity searching may seem rather surprising. We believe that the low ROC scores can be attributed to our use of a chemically diverse screening library that is not biased toward chemical series that are often responsible for the (apparently) good performance of 2D methods.[32,33] Hence, our benchmark scenario might give a more realistic expectation for search performance.

Comparison to Turbo Similarity Searching. In BSS, we use iterative virtual screening, i.e., we expand from a single reference structure to multiple compounds that are then used as probes in a second virtual screening step. To show that for BSS not only the inclusion of multiple reference compounds but also especially their biological annotation is critical to search performance, we carried out turbo similarity searches for comparison. At our selected threshold t = 0.7, we identified all structural neighbors for a probe compound and then combined them into a reference set for a standard 1-NN chemical similarity search. As shown in Table 2, turbo similarity searching performed worse than a standard one-step similarity search on our benchmark system, and only moderate enrichments of hits (up to 3-fold for single data sets) were obtained for sets of 10,000 compounds. This observation clearly emphasizes the important contribution that HTS-FPs made to search performance.

Scaffold Retrieval. The structural diversity of identified hits is often more important than their actual number as the discovery

dx.doi.org/10.1021/ci300607r | J. Chem. Inf. Model. 2013, 53, 692–703

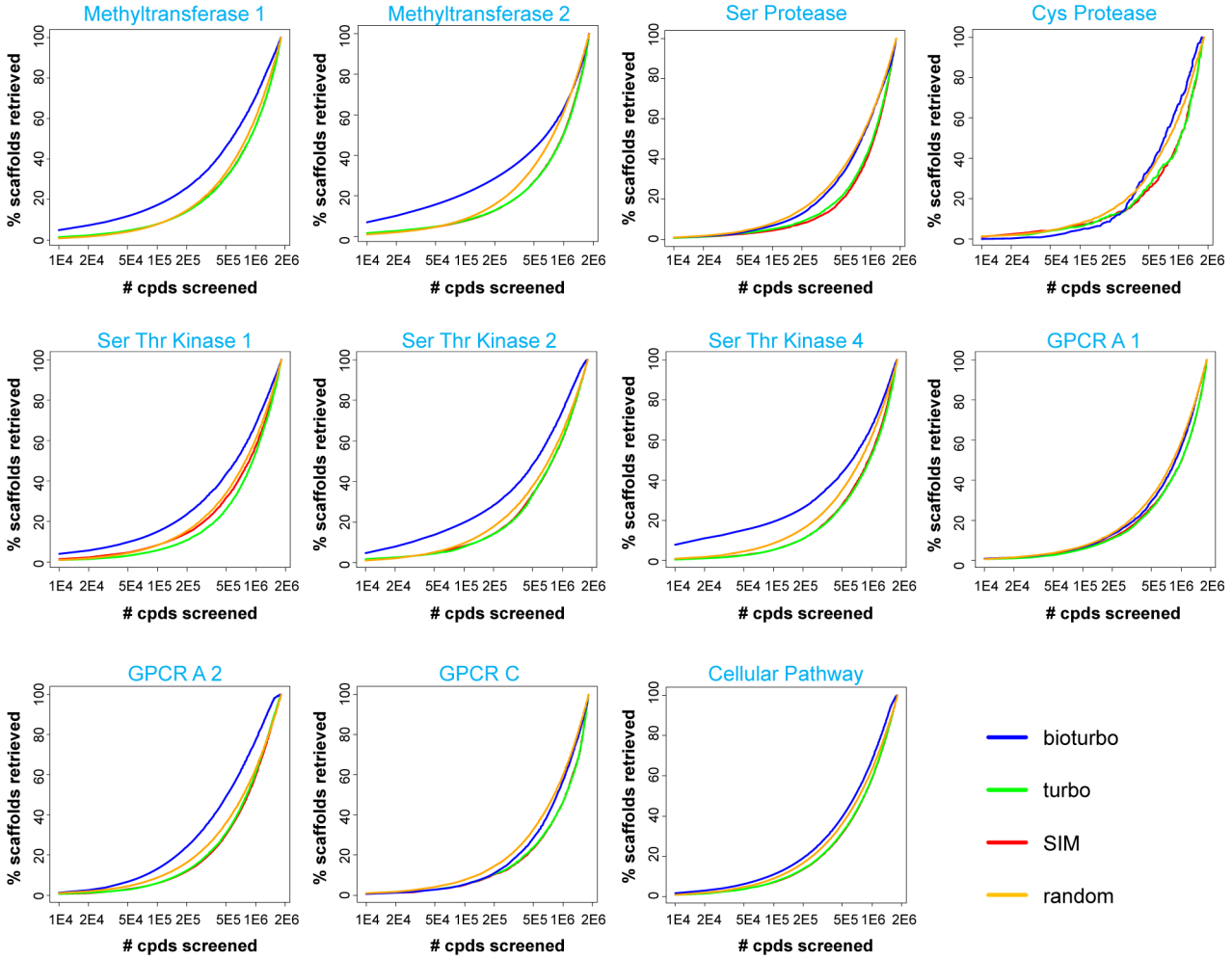


Figure 3. Two Step BSS – Scaffold Retrieval. For the 14 benchmark sets in our study, scaffold retrieval curves showing the average percentage of recovered scaffolds at increasing selection set sizes are shown. BSS (blue) is compared to turbo similarity searching (green), conventional similarity searching (red), and a random ranking of the screening database (orange). For BSS and turbo similarity searching, the ECFP4 Tc threshold t is set to 0.7.

of novel active chemotypes might result in new starting points for exploratory chemistry and lead optimization.[34] Thus, in addition to the number of active compounds that were retrieved, we also analyzed the number of different scaffolds contained in the hits found by the investigated VS methods. First, we noticed that the methods were complementary in the active compounds and scaffolds they recovered. On average, 93.6% (95.5%) of all scaffolds (compounds) recovered at a selection set size of 10,000 compounds were unique to BSS and not found by chemical similarity searching. On the other hand, also 85.3% (88.5%) of all scaffolds (compounds) identified by chemical similarity searching were not found in the hit list of BSS. These findings clearly emphasize the orthogonal definitions of molecular similarity that are taken on by the two approaches and highlight the benefit that could be gained from their parallel application in hit finding efforts. Second, in comparison to turbo and standard similarity searching, BSS consistently led to an earlier detection of more scaffolds (Figure 3), in accordance with previously reported results that highlighted the strong scaffold hopping potential of HTS-FPs.[16] Furthermore, scaffold retrieval rates expected by arbitrary compound picking were simulated by generating five random database rankings for each reference structure and averaging the results. Figure 3 shows that HTS-FPs led to a substantial increase in recovered active scaffolds for six targets. For the remaining five activity classes, the number of distinct active scaffolds was similar to that expected by random compound subset selection.

Additional Iterations. To explore whether the diversity of hits would improve further when adding another round of HTS-FP searching to Two Step BSS, we selected the top ranked compounds from this approach, included them into the reference set, and carried out another 1-NN HTS-FP search against the screening collection. However, for most data sets, adding another round of HTS-FP searching yielded very similar numbers of retrieved scaffolds and compounds. Hence, as no immediate benefit from incorporating more computational complexity into the approach was noticeable, we recommend the application of the simpler Two Step BSS approach.

Three Step BSS Performance Evaluation. So far, we have shown that, starting from a single reference structure without HTS-FP annotation, BSS can be used to enrich selected subsets of the Novartis screening collection with structurally diverse hits. However, it would also be desirable to find novel active molecules that are not part of the HTS-FP annotated compound pool. Therefore, we asked the question whether it would be possible to add another step to our bioturbo similarity search by using the top-ranked compounds from the HTS-FP comparison (step 2) to carry out another chemical similarity search against an external database. Of the 67 probes for which BSS with $t = 0.7$ could be applied, 53 compounds inhibited a protein target for

dx.doi.org/10.1021/ci300607r | J. Chem. Inf. Model. 2013, 53, 692–703

Table 3. Three Step BSS − Enrichment Factors[a]

		frequency score threshold							
target	SIM	2.5×10^{-4}	2×10^{-4}	1.5×10^{-4}	1×10^{-4}	5×10^{-5}	1×10^{-5}	5×10^{-6}	1×10^{-6}
Methyltransferase 1	4.12	6.67	6.70	6.67	6.56	6.15	3.69	5.06	6.07
Methyltransferase 2	2.86	8.25	8.20	8.14	8.87	9.15	7.55	6.90	7.43
Ser Protease	1.66	1.26	1.34	1.38	1.50	1.85			
Cys Protease	1.96	2.25	2.27	2.46	2.70	2.26	1.44	1.41	1.74
Ser/Thr Kinase 1	3.35	15.27	15.27	15.27	15.27	15.27	15.27	13.15	7.20
Ser/Thr Kinase 2	1.11	12.65	12.65	12.65	12.65	13.95	0.56	1.35	
Ser/Thr Kinase 4	0.84	9.71	9.71	9.71	8.50	5.98	2.96	2.14	0.32
GPCR A 1	1.15	0.86	0.68	0.76	0.97	0.68	0.29	0.29	0.86
GPCR A 2	1.88	2.69	2.79	2.40	2.17	2.14	1.67	0.88	0.73
GPCR C	1.39	0.77	0.77	0.57	0.50	0.33	0.89	1.03	
average	2.03	6.04	6.04	6.00	5.97	5.78	3.81	3.58	3.48

[a]For the ten benchmark sets subjected to Three Step BSS, enrichment factors for sets of 10,000 top-ranked compounds are reported. Three Step BSS is applied with different values for the frequency score threshold f, ranging from 2.5×10^{-4} to 10^{-6}. Average enrichment factors in the bottom row are obtained by averaging mean enrichment factors of all data sets.

which hit sets were available in our external database. For this probe subset, we performed another chemical similarity search with the frequency score threshold f being varied between 10^{-6} and 2.5×10^{-4}.

Enrichment Factors. Enrichment factors were calculated analogously to Two Step BSS, but this time using both primary activity data and dose−response measurements from EXTDB for hit definition (see Methods). Overall, enrichment factors were slightly higher than for Two Step BSS (Table 3). Perhaps not surprisingly, Three Step BSS worked best for those targets for which Two Step BSS achieved highest hit enrichments, i.e., methyltransferases and kinases. Best results were obtained for frequency score cutoffs between 5×10^{-5} and 2.5×10^{-4}, yielding on average a 6-fold enrichment for sets of 10,000 compounds (Figure 4). By contrast, chemical similarity searching resulted

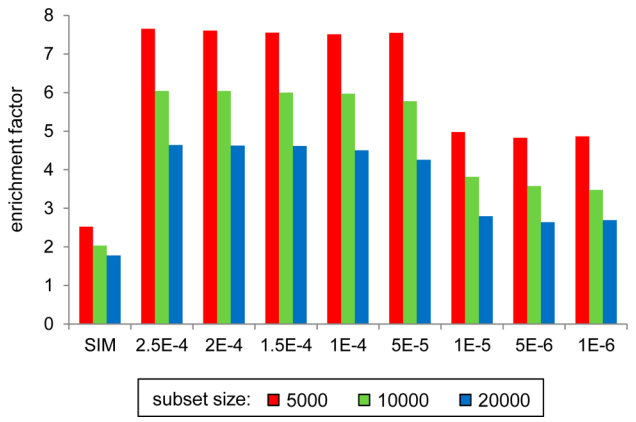


Figure 4. Three Step BSS − Enrichment Factors. Average enrichment factors are displayed for selection set sizes of 5,000, 10,000, and 20,000 compounds. Three Step BSS is carried out with the frequency score threshold f incrementally decreased from 2.5×10^{-4} to 10^{-6} and is compared to standard similarity searching (SIM).

only in a 2-fold enrichment of active compounds and was outperformed by BSS for all targets but GPCRs A1 and C. In accordance with our findings for Two Step BSS, the maximal average enrichment for a data set achieved by standard similarity searching was 4-fold (Methyltransferase 1), whereas BSS led to a more than 15-fold enrichment of hits for the Serine/Threonine Kinase 1 set.

For the BSS approach, similar enrichment factors for thresholds between 5×10^{-5} and 2.5×10^{-4} can be traced back to largely overlapping or identical reference sets that result from our decision to limit these sets to 100 compounds. For example, for approximately 50% of all probes, more than 100 compounds fell below a cutoff f of 5×10^{-5}. Consequently, for these probes, the same reference set was repeatedly generated for this and all higher cutoffs.

For all of our 53 probes, reference sets can be assembled when setting f to 5×10^{-5} or greater, hence ensuring a high applicability of the approach (Supplementary Table 3). For the test calculations reported in the following, we decided to use 5×10^{-5} as cutoff that had the lowest computational cost among the well-performing thresholds.

Application of Two and Three Step BSS to Natural Products. To test the selected cutoffs $t = 0.7$ and $f = 5 \times 10^{-5}$ on an external test set, we extracted 304 activity-annotated natural products from ChEMBL and GVK that had a structural neighbor with an ECFP4 Tc similarity of at least 0.7 in the Novartis screening collection. For the 73 probes that inhibited targets for which primary assays had been run at Novartis (see Methods), we evaluated the performance of Two Step BSS. For selection set sizes of 5,000, 10,000, and 20,000 compounds, average enrichment factors of 9.9, 7.9, and 6.1, respectively, were obtained. Hence, the performance of Two Step BSS on the external test set slightly exceeded the performance that we had seen for our benchmark system. Figure 5 (left) shows the structure of trapoxin A (TPX-A), an irreversible histone deacetylase 2 (HDAC2) inhibitor that binds covalently to the enzyme via the epoxide. A chemical similarity search on the Novartis screening collection retrieves compound NV-A (Figure 5 center, Tc = 0.77). We found 140 hits among the top 10,000 compounds selected by HTS-FP similarity to NV-A, which corresponds to a 11.3-fold enrichment. Figure 5 (right) shows nine of these 140 hits. The compounds are annotated with their ECFP4 Tc similarity to TPX-A and the rank that they obtained in a conventional similarity search. It becomes apparent that all of these hits that are structurally distinct from the probe would not have been found by ligand-based approaches. Hit compounds are chemically much simpler and synthetically more accessible than the probe, making them attractive from a medicinal chemistry perspective. All compounds are void of the reactive epoxide group that is present in TPX-A, i.e., their mode of inhibition − chelation of the zinc by the hydroxamic acid − is different from

dx.doi.org/10.1021/ci300607r | J. Chem. Inf. Model. 2013, 53, 692−703

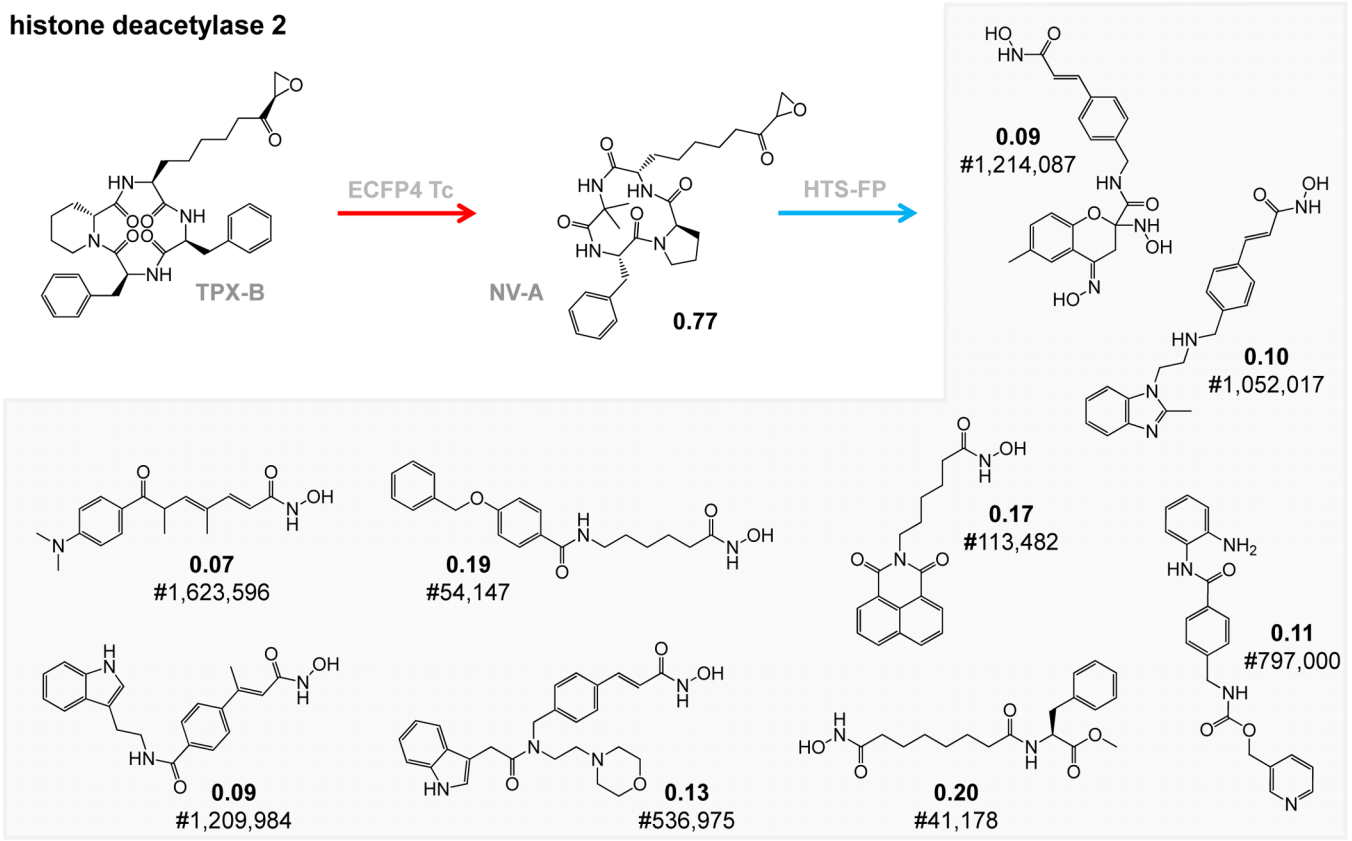


Figure 5. HDAC inhibitors. Two Step BSS is applied to TPX-A, an HDAC2 inhibitor. Chemical similarity searching retrieves NV-A as an HTS-FP annotated structural neighbor of TPX-A (ECFP4 Tc = 0.77). Nine HDAC2 inhibitors that have high HTS-FP similarity to NV-A and rank among the top 10,000 compounds are shown on the right. All nine hit compounds are annotated with their ECFP4 Tc similarity to TPX-A (bold) and the rank that they obtained in conventional similarity searching.

the mode of inhibition of the probe (covalent binding of the epoxide). Moreover, the hit compound in the lower right of Figure 5 lacks the hydroxamic acid moiety which sets it apart from the other HDAC inhibitors. Although it is only moderately active against HDAC2 (~4.9 μM), its unique structural composition makes it an interesting starting point for compound optimization efforts.

For 296 of our 304 probe compounds, Two Step BSS yielded one or multiple biologically similar neighbors below the selected frequency score threshold of 5×10^{-5}, enabling us to carry out a third step and search for hits in EXTDB. Many of the compounds showed inhibitory activity against more than one target (see Methods). Overall, Three Step BSS was carried out for 419 different targets and 1,422 different target-ligand pairs. Similar to our observations for Two Step BSS, we found that average enrichment factors obtained for our natural product set were slightly higher than for our benchmark system: 14.5, 10.2, and 7.1 for subsets of 5,000, 10,000, and 20,000 compounds, respectively. This performance increase can partly be attributed to a large fraction of kinase inhibitors in the natural product set for which bioturbo similarity searching works particularly well. However, also for GPCRs improved performance is obtained. Whereas bioturbo similarity searching largely failed on this target class in the benchmarking, we now obtained on average an almost 3-fold enrichment of hits among the 10,000 top-ranked compounds for proteins from this family.

Figure 6 shows exemplary compound sequences that lead from a natural product probe to a hit compound in EXTDB. In Figure 6A, the starting compound is tetrandrine, an alkaloid purified from a Chinese medicinal herb and known to act strongly on alpha-adrenergic receptors. Tetrandrine is structurally related to oxyacanthine that, in turn, is similar in bioactivity space to the third compound in the sequence that represents a structurally very simple three ring-containing scaffold. Finally, small structural modifications lead to the final hit compound detected in EXTDB that, as the starting compound, binds to adrenergic receptors in the nanomolar range. Figure 6B shows how plerixafor, a macrocyclic compound that antagonizes allosterically the binding of ligands at the chemokine receptor 2 (CCR2), is linked to a dihydrobenzodioxine-containing structure. The hit compound exhibits only micromolar activity against CCR2. However, it is certainly widely accepted that the discovery of a novel active chemotype outweighs an observed potency loss that could be improved through lead optimization. By contrast, Figure 6C displays the steps from a linear chain structure with micromolar activity against the substance P receptor to a 4-ring containing compound with nanomolar affinity; i.e., in this case, a chemically more attractive chemotype is found that, in addition, shows higher potency than the probe compound.

It should be mentioned that, in general, little is known about protein targets of the bridging compounds shown in Figures 6A–C. Only the linear chain compound marked with an asterisk in Figure 6C is known to have inhibitory activity at the substance P receptor. However, as the displayed compounds were able to establish a connection between compounds sharing the same bioactivity, it seems straightforward to suggest that they possess the same bioactivity as the starting and end compounds.

dx.doi.org/10.1021/ci300607r | J. Chem. Inf. Model. 2013, 53, 692–703

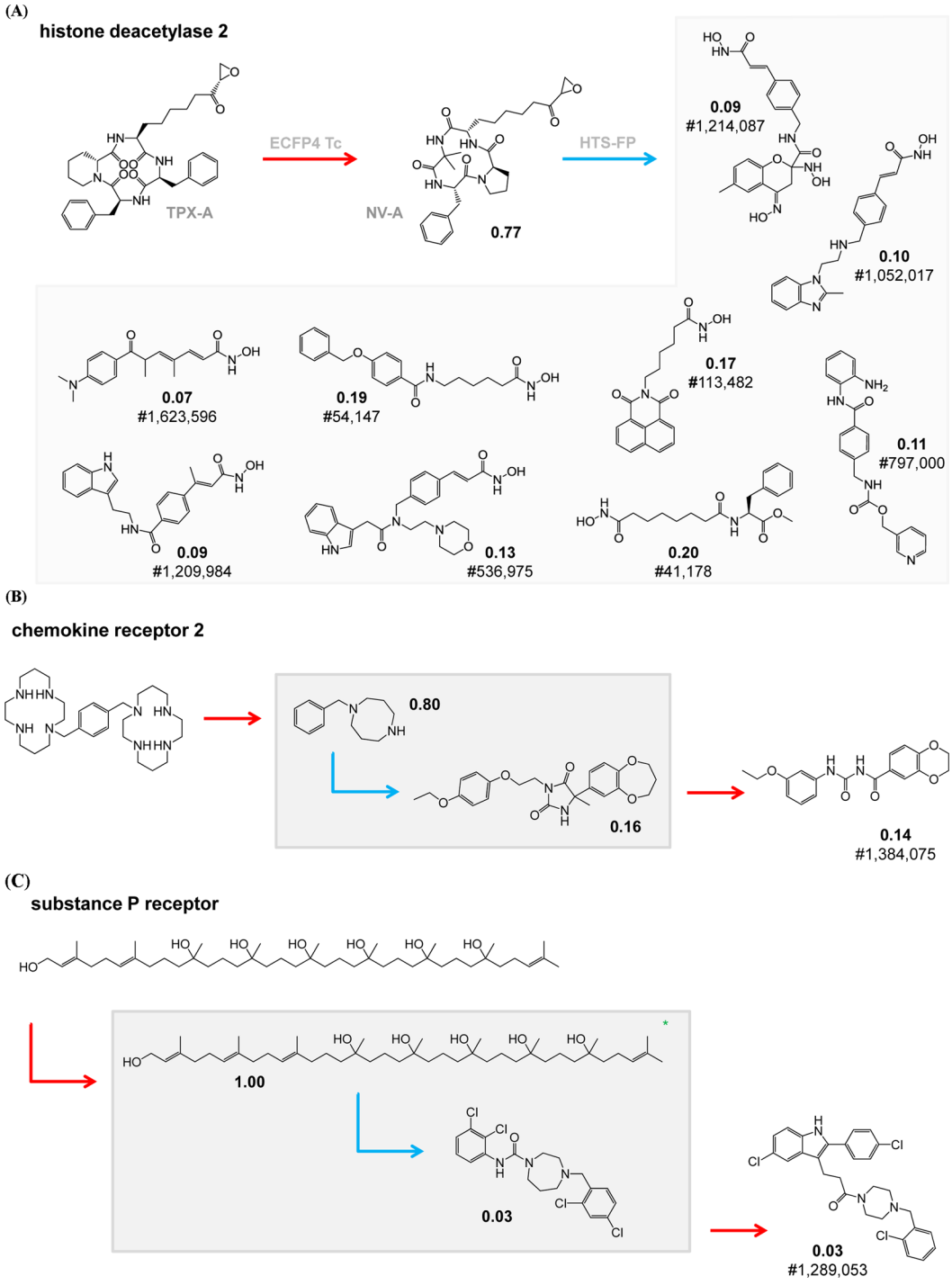


Figure 6. Exemplary compound sequences. For three Three Step BSS trials, the steps leading from a natural product probe (left) to a final hit compound (right) are displayed. All hit compounds were found in selected subsets of 10,000 compounds. Structures are annotated with their ECFP4 Tc similarity to the starting compound (bold) and the rank that they obtained in standard similarity searching. The shown start and end compounds act on (**A**) alpha-adrenergic receptors, (**B**) chemokine receptor 2, and (**C**) substance P receptor. The compound marked with an asterisk in (**C**) is a known substance P receptor modulator.

Taken together, a common feature of all shown examples is the successful retrieval of structurally simpler, synthetically accessible molecules that are chemically very distinct from the original natural product. In all cases, the scaffold hopping ability of the HTS-FP approach was exploited and successfully extended to molecules without HTS-FP annotation by using chemical similarity as additional link.

CONCLUSIONS

We have introduced the concept of bioturbo similarity searching that overcomes general limitations of bioactivity-based virtual screening. Whereas traditional bioactivity-based virtual screening is restricted to compounds that have been biologically profiled, BSS uses chemical similarity to map molecules without a priori biological annotations into bioactivity space. It should be noted that this step faces the same limitations as other ligand structure-

dx.doi.org/10.1021/ci300607r | J. Chem. Inf. Model. 2013, 53, 692−703

based approaches, i.e., it is well-acknowledged by the authors that the presence of activity cliffs might lead to the identification of structurally similar but biologically differently acting compounds. However, our benchmark calculations revealed that, by using an ECFP4 threshold of 0.7, chemical neighbors that were suitable biological surrogates for the probe compound could often be identified. These biologically annotated compounds were then used to hop in chemical space by biological profile comparisons. In our study, BSS outperformed conventional similarity searching (using chemical similarity only) in many cases both in terms of the number and diversity of recovered hit structures.

Furthermore, using BSS we identified chemically attractive small molecules on the basis of natural product reference compounds. This application emphasizes the high practical utility of the approach and its ability to depart from complex compound structures and find chemically much more tractable matter. We demonstrated the ability of BSS to discover hits with novel modes of inhibition that traditional 2D and 3D similarity approaches are unlikely to discover.

BSS is currently actively used in lead discovery efforts at Novartis, and we expect it to make a substantial contribution to hit finding in the future.

Furthermore, to enable all interested readers to explore bioactivity-based compound relationships as part of their own research projects, we provide a matrix of more than 170 million HTS-FP similarities between ∼18,600 public compounds taken from the ChEMBL database. Upon publication, this matrix is made freely available and can be obtained via http://oak.novartis.com.

■ ASSOCIATED CONTENT

Supporting Information

Supplementary Table S1 provides detailed information upon the applicability of Two Step BSS for each of our 14 benchmark sets. **Supplementary Table S2** lists average ROC scores obtained for Two Step BSS using different values for the Tc threshold *t*. **Supplementary Table S3** reports the applicability of Three Step BSS for our benchmark system. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +1 617 8715256. E-mail: anne_mai.wassermann@novartis.com.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.M.W. is a Presidential Postdoctoral Fellow supported by the Education Office of the Novartis Institutes for Biomedical Research. The authors wish to thank Paula Petrone, Iain Wallace, Peter Kutchukian, and Jeremy Jenkins for helpful discussions.

■ REFERENCES

(1) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening − an overview. *Drug Discovery Today* **1998**, *3*, 160−178.

(2) Willett, P. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.

(3) *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990.

(4) Bender, A.; Glen, R. C. Molecular similarity: a key technique in molecular informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204−3218.

(5) Maggiora, G. M. On outliers and activity cliffs − why QSAR often disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535.

(6) Stumpfe, D.; Bajorath, J. Exploring activity cliffs in medicinal chemistry. *J. Med. Chem.* **2012**, *55*, 2932−2942.

(7) Beroza, P.; Villar, H. O.; Wick, M. M.; Martin, G. R. Chemoproteomics as a basis for post-genomic drug discovery. *Drug Discovery Today* **2002**, *7*, 807−814.

(8) Bender, A.; Young, D. W.; Jenkins, J. L.; Serrano, M.; Mikhailov, D.; Clemons, P. A.; Davies, J. W. Chemogenomic data analysis: prediction of small-molecule targets and the advent of biological fingerprints. *Comb. Chem. High Throughput Screening* **2007**, *10*, 719−731.

(9) Paul, K. D.; Shoemaker, R. H.; Hodes, L.; Monks, A.; Scudiero, D. A.; Rubinstein, L.; Plowman, J.; Boyd, M. R. Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. *J. Natl. Cancer Inst.* **1989**, *81*, 1088−1092.

(10) Weinstein, J. N.; Kohn, K. W.; Grever, M. R.; Viswanadhan, V. N.; Rubinstein, L. V.; Monks, A. P.; Scudiero, D. A.; Welch, L.; Koutsoukos, A. D.; Chiausa, A. J.; Paul, K. D. Neural computing in cancer drug development: predicting mechanism of action. *Science* **1992**, *258*, 447−457.

(11) Kauvar, L. M.; Higgins, D. L.; Villar, H. O.; Sportsman, J. R.; Engqvist-Goldstein, A.; Bukar, R.; Bauer, K. E.; Dilley, H.; Rocke, D. M. Predicting ligand binding to proteins by affinity fingerprinting. *Chem. Biol.* **1995**, *2*, 107−118.

(12) Krejsa, C. M.; Horvath, D.; Rogalski, S. L.; Penzotti, J. E.; Mao, B.; Barbosa, F.; Migeon, J. C. Predicting ADME properties and side effects: the BioPrint approach. *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 470−480.

(13) Fliri, A. F.; Loging, W. T.; Thadeio, P. F.; Volkmann, R. A. Biological spectra analysis: linking biological activity profiles to molecular structure. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 261−266.

(14) Fliri, A. F.; Loging, W. T.; Thadeio, P. F.; Volkmann, R. A. Biospectra analysis: model proteome characterizations for linking molecular structure and biological response. *J. Med. Chem.* **2005**, *48*, 6918−6925.

(15) Cheng, T.; Wang, Y.; Bryant, S. H. Investigating the correlations among the chemical structures, bioactivity profiles and molecular targets of small molecules. *Bioinformatics* **2010**, *26*, 2881−2888.

(16) Petrone, P.; Simms, B.; Nigsch, F.; Lounkine, E.; Kutchukian, P.; Deng, Z.; Davies, J. W.; Jenkins, J. L.; Glick, M. Rethinking molecular similarity: comparing compounds on the basis of biological activity. *ACS Chem. Biol.* **2012**, *17*, 1399−1409.

(17) Bender, A.; Jenkins, J. L.; Glick, M.; Deng, Z.; Nettles, J. H.; Davies, J. W. "Bayes Affinity Fingerprints" improve retrieval rates in virtual screening and define orthogonal bioactivity space: when are multitarget drugs a feasible concept? *J. Chem. Inf. Model.* **2006**, *46*, 2445−2456.

(18) Briem, H.; Kuntz, I. D. Molecular similarity based on DOCK-generated fingerprints. *J. Med. Chem.* **1996**, *39*, 3401−3408.

(19) Simon, Z.; Peragovics, A.; Vigh-Smeller, M.; Csukly, G.; Tombor, L.; Yang, Z.; Zahoránsky-Kohalmi, G.; Végner, L.; Jelinek, B.; Hári, P.; Hetényi, C.; Bitter, I.; Czobor, P.; Málnási-Csizmadia, A. Drug effect prediction by polypharmacology-based interaction profiling. *J. Chem. Inf. Model.* **2012**, *52*, 134−145.

(20) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912−5931.

(21) 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.

(22) 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.

(23) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(24) 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.
(25) Bender, A.; Jenkins, J. L.; Scheiber, J.; Sukuru, S. C. K.; 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.
(26) Gardiner, E. J.; Holliday, J. D.; O'Dowd, C.; Willett, P. Effectiveness of 2D fingerprints for scaffold hopping. *Future Med. Chem.* **2011**, *3*, 405−414.
(27) 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 data sets. *J. Chem. Inf. Model.* **2012**, *52*, 2884−2901.
(28) 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.
(29) *GOSTAR Online Structure-Activity Relationship Database*; GVK Biosciences Private Limited: Hyderabad, India. http://gostardb.com/gostar/ (accessed October 10, 2012).
(30) UniProt Consortium. Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.* **2011**, *39*, D214-D219.
(31) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.
(32) Jain, A. E.; Cleves, A. E. Does your model weigh the same as a duck? *J. Comput.-Aided. Mol. Des.* **2012**, *26*, 57−67.
(33) Yera, E. R.; Cleves, A. E.; Jain, A. N. Chemical structural novelty: on-targets and off-targets. *J. Med. Chem.* **2011**, *54*, 6771−6785.
(34) Glick, M.; Jacoby, E. The role of computational methods in the identification of bioactive compounds. *Curr. Opin. Chem. Biol.* **2011**, *15*, 540−546.