JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

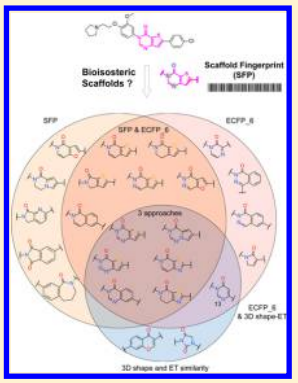
Novel Scaffold Fingerprint (SFP): Applications in Scaffold Hopping and Scaffold-Based Selection of Diverse Compounds

Obdulia Rabal, Fares Ibrahim Amr, and Julen Oyarzabal*

Small Molecule Discovery Platform, Molecular Therapeutics Program, Center for Applied Medical Research (CIMA), University of Navarra, Avda. Pio XII 55, E-31008 Pamplona, Spain

Supporting Information

ABSTRACT: A novel 2D Scaffold FingerPrint (SFP) for mining ring fragments is presented. The rings are described not only by their topology, shape, and pharmacophoric features (hydrogen-bond acceptors and donors, their relative locations, sp3 carbons, and chirality) but also by the position and nature of their growing vectors because they play a critical role from the drug discovery perspective. SFP can be used (i) to identify alternative chemotypes to a reference ring either in a visual mode or by running quantitative similarity searches and (ii) in chemotype-based diversity selections. Two retrospective case studies focused on melanin concentrating hormone 1-receptor antagonists (MCH-R1) and phosphodiesterase-5 inhibitors (PDE5) demonstrate the capability of this method for identifying novel structurally different and synthetically accessible chemotypes. Good enrichment factor (155 and 219) and recall values (46% and 73%) are found within the first 100 ranked hits (0.3% of screened database). Our 2D SFP descriptor outperforms well-validated current gold-standard 2D fingerprints (ECFP_6) and 3D approaches based on shape and electrostatic similarity. Scaffold-based selection of diverse compounds has a critical impact on corporate library design and compound acquisitions; thus, a novel strategy is introduced that uses diverse scaffold selections using this SFP descriptor combined with R-group selection at the different substitution sites. Both approaches are available as part of an interactive web-based application that requires minimal input and no computational knowledge by medicinal chemists.



INTRODUCTION

A common approach in drug design is to focus on the molecular scaffold, the central core component of a molecule,[1] rather than on the entire molecule. The goal of this scaffold analysis is typically either the identification of bioisosteric replacements of the scaffold (scaffold hopping) to overcome chemotype-related adsorption, distribution, metabolism, excretion, and/or toxicity (ADMET) issues and/or to achieve intellectual property or the evaluation of the structural diversity of the compound collections. Although the perception of the molecular scaffold is highly dependent on the molecular set under analysis, ring systems are frequently used to define the molecular scaffold due to their dominant role in medicinal chemistry programs.

Concerning scaffold hopping, the number of possible alternatives for a chemotype is too great to be memorized and requires considerable experience and/or creativity. As an example, the virtual VEHICLe database[2] has 24 847 scaffolds and it is "solely" composed of heteroaromatic mono- and bicycles.

Thus, in order to support medicinal chemists, different pharmaceutical companies have developed their own internal databases of ring systems or fragments for scaffold hopping as well as 2D and 3D approaches to query them. In 2003, investigators from GSK[3] presented an engine for querying among 120 K unique ring systems extracted from the GSK corporate collection and commercial databases. The scaffold descriptors implemented within the program are mostly based on atom counts and geometric descriptors for two- and three-connection rings. This approach is well-suited for identifying geometric replacements, especially if the core merely acts as a geometric linker. For traditional isosteric replacement, the searches were complemented with external programs to account for shape and electronic and lipophilic similarities. More recently, Ertl[4] at Novartis presented a database for scaffold hopping by extracting ring systems with three and four connections from ChEMBL. In this case, properties were chosen to find a good balance between ADME properties (e.g., TPSA, log P) and structural descriptors (e.g., shape, position of R groups, pharmacophore features). Scientists from Pfizer presented the lead transformation tool NEAT,[5] which focuses on identifying aromatic rings extracted from a modified subset of the virtual database GDB consisting of 25 131 unique 5- and 6-membered monocycles and bicycles. Scaffolds are ranked according to their Brood electrostatic potential overlap, using QM ESP charges and a set of high-level QM properties (e.g., relative hydrogen-bond strength, aromaticity, etc.). AstraZeneca scientists[6] extended the methodology of CAVEAT for describing the relative geometry of the attachment vectors and used a fragment-based database extracted from different databases of compounds. Notably, this approach was not restricted to cyclic systems. In previous work, we utilized an annotated database of ring systems for 3D shape- and

Received: July 31, 2014
Published: January 5, 2015

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

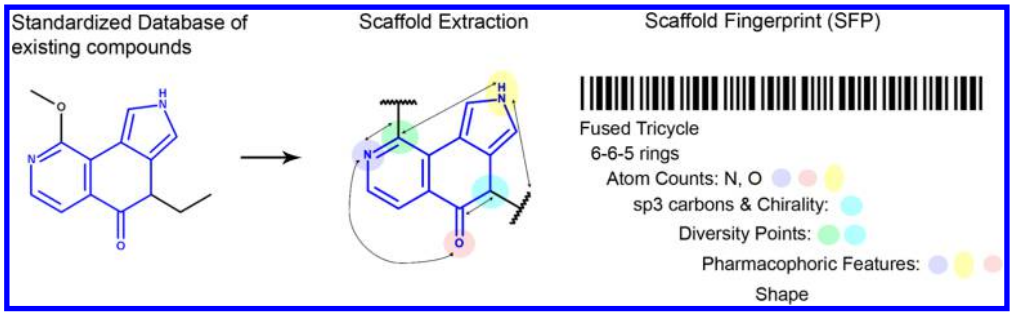


Figure 1. Workflow to generate the Scaffold FingerPrint (SFP).

electrostatic-based similarity scaffold hopping that included labeled growing vectors.[7,8] Recent work in this area includes the Scaffold Keys presented by Ertl,[9] based on topological descriptors.

Here we introduce a novel Scaffold FingerPrint (SFP) that encodes very simple to calculate and easily understandable 2D properties of ring systems in a single fingerprint. SFP accounts for the topology, shape, presence, and relative positioning of pharmacophoric points (hydrogen-bond acceptor and donors), sp3 carbons, and their chirality as well as by the position and nature of the growing vectors of the scaffold. Thus, relevant properties are contemplated whether the ring fragment acts as a mere geometric "scaffold" (just to position the protein-interacting side chains in the correct orientation) or whether it is directly involved in the ligand−receptor interactions. As far as we know, this is the first time a fingerprint combines all these key features, particularly topology of the ring and alignment-free atom pairs between pharmacophoric points and diversity points. SFP capitalizes on the biologically relevant information on the chemical space occupied by ring systems.

Once the corresponding code for SFP calculation was developed, it was implemented within a web-based application that supports visual interactive navigation for molecular design by hand of alternative chemotypes as well as quantitative searches, where the user only needs to input the target scaffold to be replaced and, with just one click, query a compiled database of unique scaffolds. Thus, no special expertise in cheminformatics is required. Moreover, the computational demand of 3D approaches[7,8] is reduced. A study case to identify analogs of Vemurafenib scaffold, a BRAF inhibitor, is shown to demonstrate the utility of the interactive navigator for visually selecting scaffolds having a set of user-required pharmacophoric features and substitution patterns. Besides, we illustrate the impact of similarity-driven searches to bioisostere findings in two retrospective case studies: antagonists of the melanin concentrating hormone 1-receptor (MCH-R1) and phosphodiesterase-5 (PDE5) inhibitors. The results are compared with gold-standard 2D fingerprints and 3D approaches based on shape and electrostatic similarity.

At the same time, considering the impact of an optimal "chemotype/hit" ratio on the next steps of a drug discovery program after the screening campaign,[10] we propose a workflow for scaffold-based diverse selections of commercial compounds based on the combination of the SFP and the selection of the R-groups at the different substitution sites according to our previously reported methodology.[11,12] Commonly, scaffold-based selections of compound libraries are driven by maximizing the number of scaffolds, as applied to the plate-based diversity selection of compounds at Novartis,[13] or its compounds representativeness. In these cases, the structural diversity of the cyclic rings is not considered. Strategies that account for structural information[14−18] are mostly devoted to assessing library diversity, rather than demonstrating practical cases of diverse scaffold selection for library acquisition. Recent reports describe strategies for diverse selections of fragments for fragment-based drug discovery;[19,20] however, they are not based on ring fragments alone.

METHODS

The essential steps of the workflow for defining scaffolds and generating the SFP are depicted in Figure 1 and described below. This process is common to both approaches: scaffold hopping applications and scaffold-based diverse selection of compounds.

1. Scaffold Definition and Scaffolds Source. Among the many different scaffold definitions, we opted to define our scaffolds as the following ring systems: a single ring or a collection of fused or spiro rings, including exocyclic terminal bonds (carbonyls, sulfonyls, imines, sulfinyls, and thiocarbonyls). We did not consider the Bemis-Murcko molecular frameworks[21] because attaching a cyclic R-group to a scaffold changes the molecular framework; therefore, a chemical series decorated with different cyclic substituents would result in different chemotypes. Therefore, our goal is for our tool to be used for scaffold replacement, and because one of the reasons for doing so is to generate patentable analogs, a wider Markush formula is preferable.

Scaffolds are obtained by fragmenting existing chemical structures, which guarantee synthetic feasibility. The information on the position of the R-group substituents is maintained by including the first atom outside of the fragment, with the atom type *Z*, retaining information on the type (single or double) and the stereochemistry of the bond in the original fragment. In order to retain the tautomeric variability caused by −OH, −SH, and −NH2 substituents on rings, these terminal groups are initially kept during fragmentation to be later transformed in the corresponding =O, =S, and =N tautomers (if any). A customized fragmentation procedure was implemented in Pipeline Pilot to generate these ring assemblies that performs mostly as the *Generate Fragments* component but that also retains the terminal single-bonded −OH, −NH2, and −SH groups. After fragmentation, all possible tautomers of each ring system are enumerated with the *Enumerate Tautomers* component (options *Amides Tautomerization* = *Tautomerize Only Diamides*; *MakeAllSp2AtomsAcceptors* = *False*). For each ring system, tautomers with a score lower than 50% of the highest tautomer score are discarded. This threshold value of 50% was established after some tests (*data not shown*). After tautomer enumeration, remaining −OH, −SH, and −NH2 terminal groups on the ring system are

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

Table 1. Calculated Properties for the Set of Unique Ring Assemblies[a]

property	details	q-SFP	Q-SFP	refs
Scaffold Frequency	frequency of the ring system (considering the connection pattern) in the database under analysis	YES	NO	3, 4
Scaffold Topology Class	class of the topology of the scaffold, defined by the number of fused rings, bridges, and spiro atoms (explained in the text)	YES	NO	
Scaffold Topology Size	string that collapses the ring sizes, excluding exocyclic groups, of the assembly (explained in the text)	YES	NO	
Num Bridges	no. of bridges	YES	YES	
Num SpiroAtoms	atom linkages between two rings consisting of a single atom common to both	YES	YES	3[b], 9
Num Free Spiro Atoms	atom linkages that constitute the only union direct or indirect between the two rings	YES	YES	3[b]
Num Aromatic Rings	no. of aromatic rings	YES	YES	3
Num Exocyclic Bonds	no. of exocyclic bonds	YES	YES	4
C count	no. of carbon atoms in the ring system	YES	YES	
N count	no. of nitrogen atoms in the ring system	YES	YES	3, 9
O count	no. of oxygen atoms in the ring system	YES	YES	3, 9
S count	no. of sulfur atoms in the ring system	YES	YES	3, 9
P count	no. of phosphorus atoms in the ring system	YES	YES	3
sp3 carbons count	no. of sp3 hybridized carbons	YES	YES	
Fraction_Csp3	no. of sp3 hybridized carbons/total carbon count	YES	NO	
Num_Csp3_Chiral	no. of sp3 carbons that are true stereocenters (independently of any stereo markings present in the molecule)	YES	YES	
Num_Csp3_Chiral_Rs	no. of sp3 carbons with exit-vectors that are true stereocenters (independently of any stereo markings present in the molecule)	YES	YES	
Num_Csp3_Chiral_Not_Rs	no. of sp3 carbons without exit-vectors that are true stereocenters (independently of any stereo markings present in the molecule)	YES	YES	
Is symmetric	whether the ring system is symmetric or not	YES	YES	4
Num Diversity Points	number of cleavage points in the ring system	YES	YES	3, 4
Diversity point features	vectors characterizing the number of different atom types derivatized with diversity points at certain topological distances between them (explained in the text)	YES	YES	3, 4[c]
Pharmacophore features	vectors characterizing the number of hydrogen bond donors and hydrogen bond acceptors (different types) at certain topological distances between them (explained in the text)	YES	YES	
Diversity point−Pharmacophore features	vectors characterizing number of hydrogen bond donors and/or hydrogen bond acceptors (different types) at certain topological distance from the different types of diversity points (explained in the text)	YES	YES	4[c]
Num Rings	no. of rings in the ring system	NO	YES	3, 9
Shape features	vectors characterizing number of atoms at certain topological distance from the attachment points	NO	YES	4

[a]The column "refs" lists other references in which these properties have already been used to describe ring systems in ring navigation databases or fingerprints designed to characterize rings. [b]Lewell et al.[3] encoded this information as a logical parameter, whether or not the system has spiro atoms. [c]Properties are not exactly as described in ref 4 because we use a correlation vector (CV) to bin the topological distances (see text below).

replaced by attachment point marks. This effort in enumerating tautomers (exemplified in Figure S1) was done to account for lactam−lactim and imidazole-type tautomerism, which is important when looking for new scaffold alternatives. Only the scaffolds with organic atoms (C, N, O, S, P, and H) are retained. Macrocycles are also retained. After fragmentation of the original molecules, the ring systems are merged to obtain unique ring systems.

For scaffold hopping approaches, a database of 157 533 unique scaffolds was assembled by fragmenting standardized libraries[11] of compounds at different stages of pharmaceutical development, as follows: commercially available drug-like molecules from 10 vendor catalogs extracted from the ZINC database[22] (Asinex, ChemDiv, ChemBridge, Enamine, Life Chemicals, Maybridge, Otava, Specs, TimTec and Uorsy), general bioactive compounds (ChEMBL),[23] target-family focused compounds (Kinase SARfari,[24] GPCR SARfari,[25] Kinase Knowledge Database KKB,[26] Oncology Knowledge-BASE OKB),[27] approved drugs (approved drugs subset of DrugBank),[28] or natural products (Traditional Chinese Medicine Database Taiwan).[29] It is also possible to focus the selection of the scaffolds on the scaffolds extracted from individual database. For compound acquisitions, sdfiles of external libraries are processed on-the-fly in a similar fashion.

2. Generation of Scaffold Fingerprint (SFP). The algorithm to calculate SFP was implemented using the Perl API for Molecular toolkit within Pipeline Pilot environment.[30] This toolkit provides functionalities to get and set properties for atoms and bonds, allowing the developer to expand the set of available common ring descriptors (number of rings, aromatic rings, atom counts, number of spiro atoms, ...) already available in standard Pipeline Pilot components. On the other hand, for the classification of the topologies, the determination of the chirality of the sp3 carbons and the correlation vectors to encode shape, pharmacophoric features, and diversity points, a customized novel code was developed, using also this programming environment, as described in detail in the following paragraphs.

Table 1 shows a list of the calculated properties that encompass the different dimensions of this novel SFP, schematized in Figure 1. As introduced, this tool was designed to be used in a visual model (qualitative selections) and to include similarity-based and diversity driven searches (quantitative searches). Therefore, there are properties that are available only for visualization purposes (e.g., the Scaffold Topology Class) or for quantitative measures (e.g., the shape features) or for both purposes. This availability is conveniently indicated in Table 1 under columns q-SFP (qualitative, for visualization) and Q-SFP (quantitative, to apply to similarity/

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

diversity metrics); "YES" (included) or "NO" (not included). The total number of properties for the quantitative measures results in a fingerprint of dimension 1033 (Q-SFP).

Scaffold Frequency (Only for Qualitative Visual Navigation). This descriptor is useful for selecting rings that occur most frequently, providing an indication of the likely synthetic accessibility. This property is calculated when merging the smiles of the fragmented rings to generate the set of unique rings.

Scaffold Topology Class (Only for Qualitative Visual Navigation). Different metrics have been proposed by others to classify the topology of the scaffolds, regarded as either the molecular frameworks defined by Bemis−Murcko and their analogs[31−33] or as ring systems alone.[34,35] Oprea and co-workers[32,33] have developed the ordered return index, derived from the adjacency matrix, to distinguish topologies with up to eight rings (the same or different ring systems) for molecules with atoms of a valence up to four. As we addressed the ring systems alone, for which the number of topologies is significantly more restricted than for molecular frameworks, we declined to implement this approach. However, this study inspired the definition of our hierarchy of the topological classes. Other approaches, such as the classification according to the number of terminal rings and the number of molecular bridges published by Chen et al.[31] and the molecular equivalence indices (MEQI) developed by Xu and Johnson,[36,37] have been devoted to molecular frameworks rather than independent ring systems. Regarding the approaches to ring systems, the complexity measure described by Nilakantan[35] in 1990 is insufficiently amenable to medicinal chemists. Lipkus[34] has presented a method for organizing ring systems based on their topologies. This method relies on the following three simple descriptors: *R* (number of rings), *P*, and *B*, whose corresponding equations are as follows:

$$P = 2E - S \quad (1)$$

$$B = S - E \quad (2)$$

where *E* is the number of edges in the ring system and *S* is the sum of the ring sizes in the Set of Smallest Rings (SSSR). With limitations, *P* could be regarded as the "perimeter" of the ring system, and *B*, as the number of bridge edges. The ring systems are assigned a cell in the three-dimensional space defined by their *PBR* coordinates. To further distinguish between the different topologies within a cell, the ring sizes of the SSSR (in ascending order) are used to split the cell into sub-bins.

Based on the method by Lipkus[34] and the topologies exemplified by Oprea (Figure 2 in ref 32), we established the classification of the ring systems compiled in Table S1 in the Supporting Information for ring systems with up to four rings and coded the corresponding algorithm to perform this classification. Ring systems with more than five rings are assigned to a single class according to their size because there might be many possible topologies, but they are not highly populated.[32] As shown in Table S1, the scaffold topology assignation is obtained for the different combinations of values of the number of rings, the Lipkus B descriptor, the number of bridges, and the number of free spiro atoms (in a few cases).

Scaffold Topology Size (Only for Qualitative, Visual Navigation). The classification of the ring assemblies can be refined using the information in the SSSR. This property stores the information on the ring sizes, sorted in ascending order, disregarding information on how the rings are connected, except for the case of the tricycles with linear topologies (classes 5, 6, and 7 in Table S1). Due to its prevalence in bioactive compounds, the connectivity between the different rings is considered for these A−B−C topologies by identifying the central ring (B) connecting the two other extreme rings (A and C). Then, the ring sizes are sorted starting with the smallest extreme ring (A or C) then the central ring (B) and, finally, the left ring (C or A). A few examples are given in Figure S2.

sp3 Carbons and Chiral sp3 Carbons. Lovering et al.[38] have reported that the fraction of sp3 carbon atoms and the presence of chiral centers correlate with clinical success. Taylor et al.[39] have emphasized the importance of the location of these sp3 carbons in a ring, observing that 40% of drugs do not contain any sp3 carbons in a ring. According to this observation, the number and fraction of sp3 carbons is annotated for each scaffold. From our viewpoint, it is not only important to have an idea of the three dimensionality of the scaffolds by analyzing the fraction of sp3 carbons in the scaffold but also to distinguish between the −CH2− linkers and the true chiral carbons in the scaffold. Thus, we consider the number of chiral carbon atoms, including all of the true sp3 carbon stereocenters (property *Num_Csp3_Chiral* in Table 1), independent of their stereo markings (undefined single bond stereo, unknown bond stereo, up bond stereo or down bond stereo). Carbons bearing two substituents (two exit vectors) are considered as stereocenters, independent of the substitution pattern (i.e., potential equal substituents) in the original compound. Finally, it is also possible to distinguish the location of these chiral sp3 carbons, as follows: chiral carbon atoms generated because of the substitution pattern and located in the position of the exit vectors (*Num_Csp3_Chiral_Rs* in Table 1) or chiral atoms in stereo ring bonds (*Num_Csp3_Chiral_Not_Rs* in Table 1). A new code was written using the Molecular Toolkit to distinguish between the different sp3 carbons. This distinction provides an idea of the stereochemical complexity of the ring fragments and their synthetic accessibility. The meaning of these properties is exemplified in Figure 2.

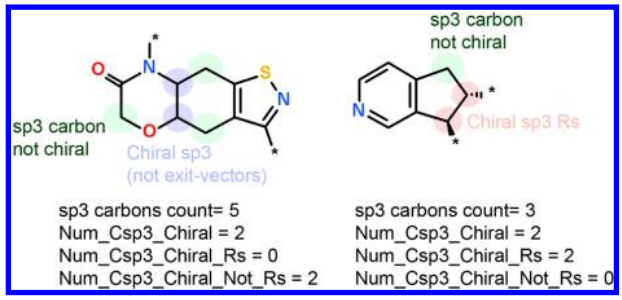
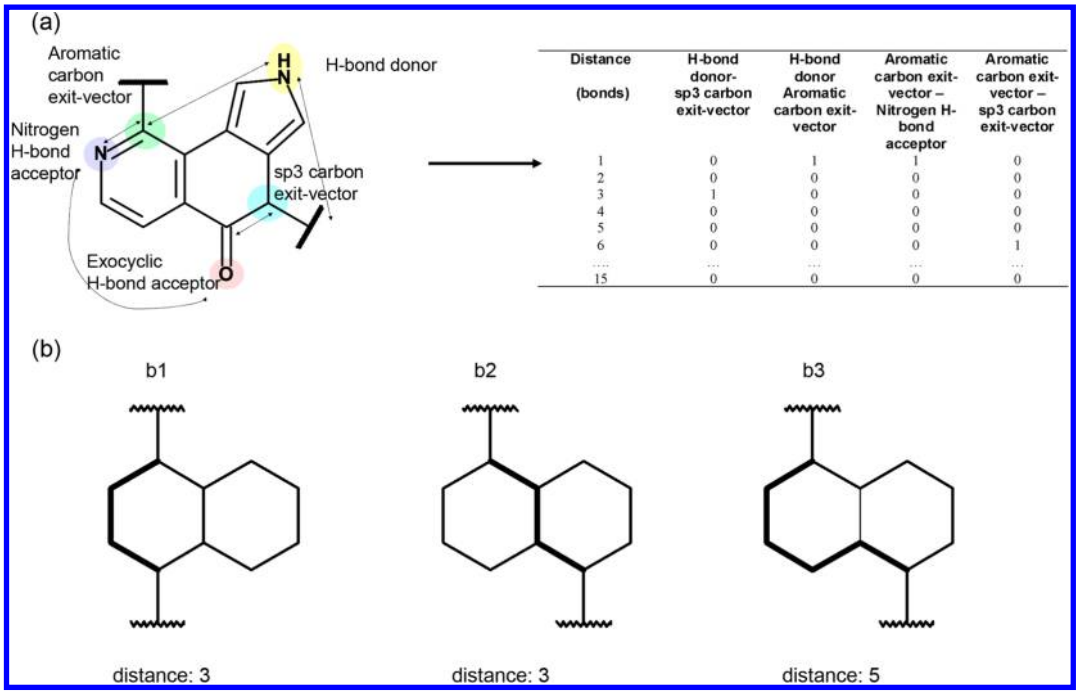


Figure 2. Calculation of the properties accounting for the number of sp3 carbons and chiral sp3 carbons.

Diversity Point Features, Pharmacophoric Features, and Diversity Point−Pharmacophoric Features. The influence of the attachment points, the hydrogen bond donors, and hydrogen bond acceptors in the ring systems is considered by the following: (i) counting the number of different atom types and (ii) calculating the topological distance between the atom pairs for all of the feature pairs using a modified version of the shortest path length and with a maximal correlation distance of 15 bonds (Figure 3a). Similar 2D autocorrelation vectors were initially described by Schneider et al.[40] for scaffold hopping using complete molecules. An adapted version for bioisoteric

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18



Distance (bonds)	H-bond donor-sp3 carbon exit-vector	H-bond donor Aromatic carbon exit-vector	Aromatic carbon exit-vector – Nitrogen H-bond acceptor	Aromatic carbon exit-vector – sp3 carbon exit-vector
1	0	1	1	0
2	0	0	0	0
3	1	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	1
...	...	...	...	...
15	0	0	0	0

Figure 3. (a) Principle of the descriptor for the diversity point features, pharmacophoric features, and diversity point−pharmacophoric features. The atom pairs for all of the feature pairs are counted, and their distance is annotated using a binning scheme. (b) Example of the calculation of the topological distance between two attachment points using a modified version of the shortest path length by removing the bridging bonds between the rings. For the ring system in parts b2 and b3, the calculated distance is five bonds, as in b3, instead of the three bonds obtained using the standard shortest path length (b2).

replacement accounting for the attachment points, although not restricted to ring systems, was later published by Wagener and Lommerse.[41] In our customized SFP code, the attachment point is distinguished according to the atom bearing the connection point into the following six classes: sp3 carbons, sp2 carbons, aromatic carbons, aromatic nitrogens, aliphatic nitrogens, and other atoms (e.g., phosphorus). The hydrogen bond acceptors are classified into the following four groups: exocyclic acceptors (e.g., oxygen in carbonyl atoms), nitrogens, oxygens, and any other acceptor atoms (e.g., sulfur, phosphorus). This classification results in a total of 11 different atom types and 66 possible pairs of atom types. When calculating the shortest path length, the nonbridging bonds between the rings are discarded in an attempt to distinguish between the cases, such as the one exemplified in Figure 3b.

Shape. For the quantitative measures, the number of rings and a property accounting for the shape of the ring system are included. This shape feature is calculated as described by Ertl[4] as the number of atoms at a certain topological distance from the connection points. Again, a maximal correlation distance of 15 bonds was considered.

3. Scaffold Hopping Based on Q-SFP. The Q-SFP for the query scaffold (with exit vectors labeled accordingly) is computed in a similar fashion to that of the unique scaffolds in the precompiled database. The similarity score between the query and the database of scaffolds can be calculated as either Tanimoto similarity or Manhattan distance. It is possible to assign different weights to each descriptor and, for example, upscale the similarity in topological distance between connection points and/or the different pharmacophoric features to optimize the 3D overlap with the query. By default, all dimensions of the 1033-fingerprint are considered with the same weight.

4. Q-SFP-Based Diverse Selection of Scaffolds. The Q-SFP is used to iteratively cluster the scaffolds using a partitioning method. Briefly, a number of representative scaffolds are selected as cluster centers using a maximum dissimilarity method. It begins by randomly choosing a scaffold as the first cluster center. The scaffold maximally distant from the first scaffold is selected as the next cluster center. The scaffold maximally distant from both current points is selected after that. The process repeats itself until there is a sufficient number of cluster centers. The corresponding clusters are found by assigning each remaining scaffold to each representative scaffold, selecting the nearest cluster center. Again, two metrics are available: Tanimoto Similarity and Manhattan distance. For setting the initial clustering, the user may request either an average number of compounds per cluster, a number of clusters, or the user may define a maximum distance a cluster member can be from a cluster center. By default, the resulting set of scaffolds in cluster centers conforms to the diverse Q-SFP based diverse selection. Alternatively, it is possible to select the scaffold within each cluster the highest occurrence frequency (i.e., the number of exemplified molecules bearing the same chemotype).

5. Implementation and Visual Interface for SFP. The whole workflow has been implemented in-house, within Pipeline Pilot.[30] For molecule fragmentation, a modified version of the *Generate fragments* component was used. As mentioned, the Perl API for Molecular toolkit was used to implement the set of properties in Table 1 capitalized in SFP. An internal code was used for the similarity metrics to include user-defined weights. Clustering is carried out by the *Cluster Molecules* component.

A web-based interface was built to enable user-access to the different functionalities: from interactive visual navigation to

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

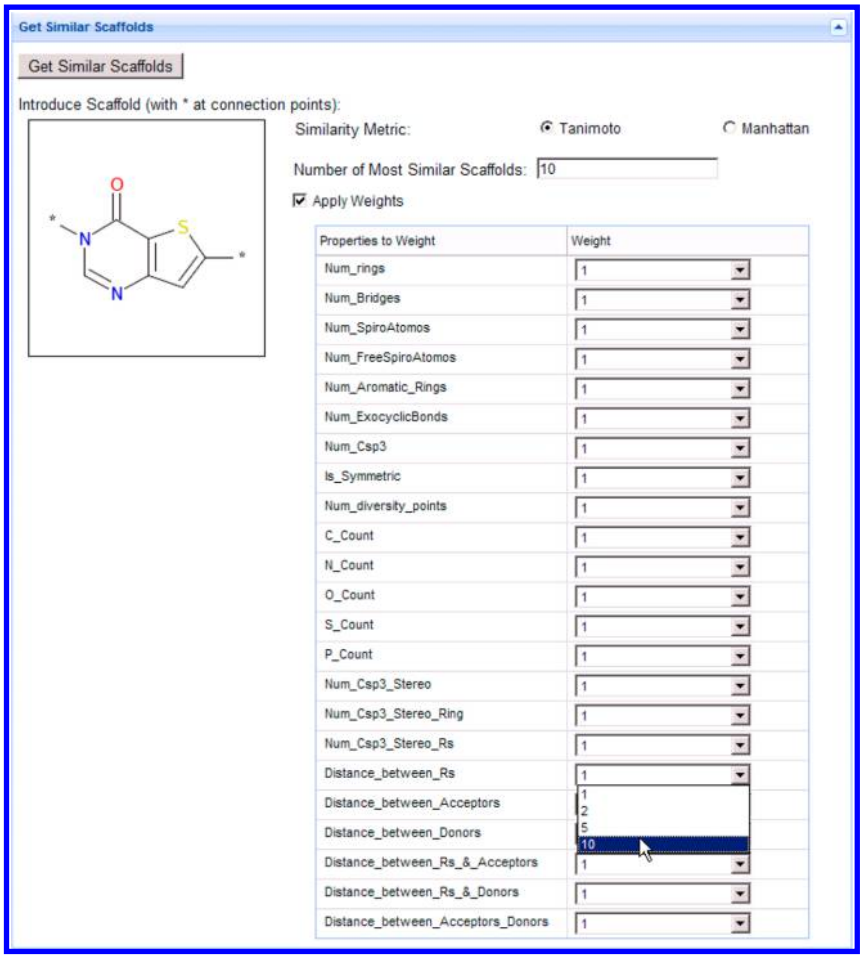


Figure 4. Panel to run Q-SFP-based searches for scaffold hopping.

quantitative similarity assessment. As in our BRCS tool for the R-group analysis,[12] the visualization of scaffolds is based on interactive heat maps; its performance is illustrated in our first study case. On the other hand, as an example, Figure 4 shows the input form to run Q-SFP-based searches for scaffold hopping.

6. Comparison of Q-SFP for Scaffold Hopping with Standard 2D Fingerprints (ECFP_6) and 3D Approaches. ECFP_6 fingerprints, as implemented in Pipeline Pilot,[30,42] were used as 2D gold-standard fingerprint for comparison purposes. Attachment points, defined as dummy atoms (*), were retained for both the reference scaffold and the database. For 3D shape and electrostatics similarity-based scaffold hopping, the same procedure as described in ref 7 was followed. Briefly, the query scaffold is decorated with small fragments (e.g., alkyl, halogens, or phenyl) to mimic the electrostatic impact of the corresponding substitution patterns to chemotypes. Then, scaffolds in the pool database are capped in an analogous manner to the query using the same decorating fragments at each open valence. Then, 3D conformations are generated with Omega.[43−45] For the query scaffold, the lowest energy conformer is chosen and multiple conformers are generated for each scaffold in the pool database using the default settings in Omega. Scaffolds are superposed to the query scaffold with ROCS software[46] and ranked according to matches based on shape and pharmacophoric points. Then, scaffolds with a Tanimoto Shape higher than 0.75[7] are reranked by their electrostatic field similarity (Poisson−Boltzmann) to the query using EON.[47]

7. Performance Metrics. For retrospective validation, the recall (R) or sensitivity and enrichment factor (EF) values are reported at different percentages of ranked database. Recall is the fraction of known active chemotypes that are retrieved. EF is usually expressed as follows (eq 3):

$$\mathrm{EF} = \frac{(LV_A)}{(VL_A)} \tag{3}$$

where L and L_A are the total number of scaffolds and known active scaffolds in the library, respectively, and V and V_A correspond to the number of chemotypes predicted as active and the number of known active chemotypes contained in V, respectively. An enrichment factor of 1 is as good as a random selection.

RESULTS

1. Scaffold Hopping of BRAF Inhibitors Using the q-SFP-Based Visual Navigator. The 1*H*-pyrrolo[2,3-*b*]pyridine of Vemurafenib,[48] an approved BRAF inhibitor, is our target scaffold (Figure 5a). The merged database of 157 533 scaffolds is our source of scaffolds and the search is initially restricted to ring systems having two exit vectors and up to three rings (the visual procedure is illustrate in Figure S3 in the Supporting Information). Initially, a new html opens with scaffolds assigned to cells in an interactive heat map representing the following q-SFP properties: number of diversity points against the scaffold

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

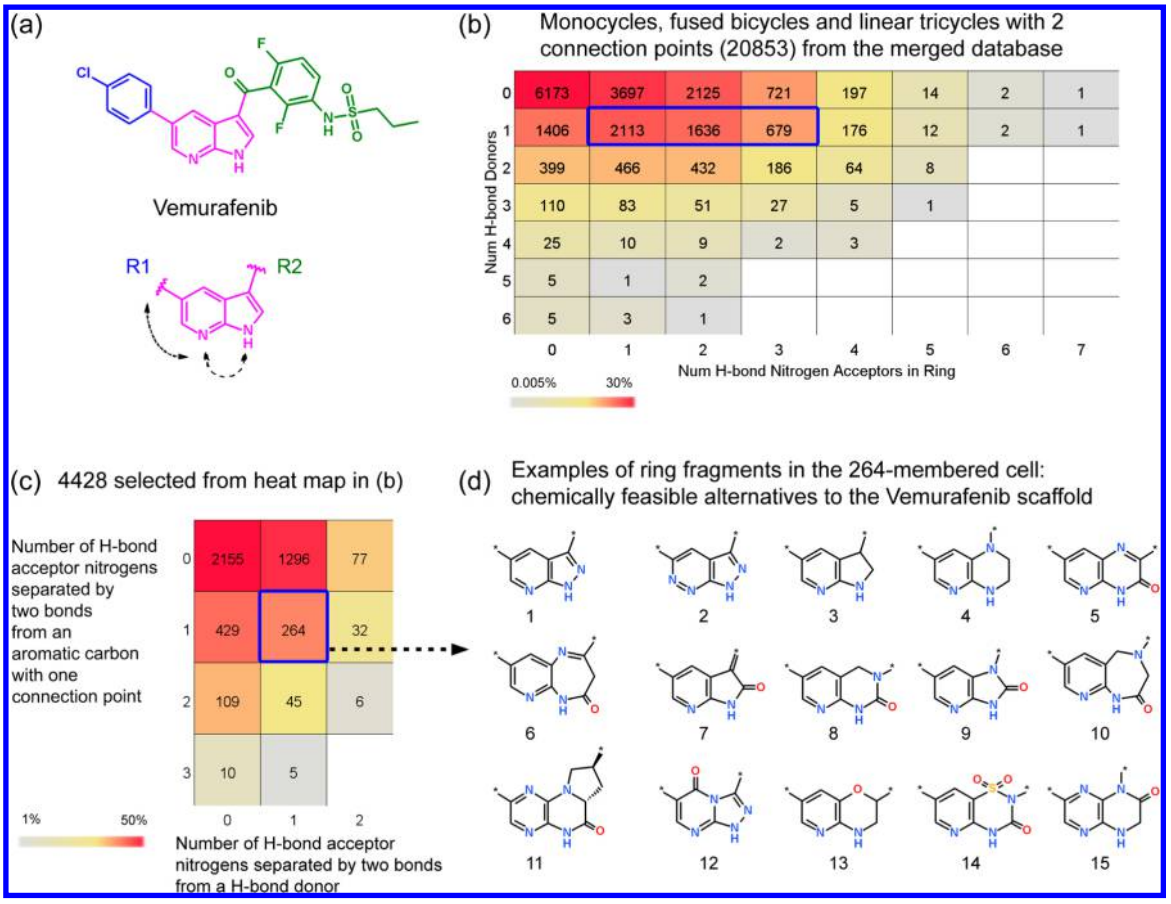


Figure 5. Interactive visual application of the SFP-based tool to search for potential bioisosters of the scaffold of Vemurafenib[48] having a set of required pharmacophoric features and substitution pattern.

topology class (Figure S3B). Then, monocycles, fused bicycles, and linear tricycles are exported into a new tab for the detailed analysis by clicking on the corresponding cells in the heat map. This results in a set of 20 853 scaffolds (Figure S3C and Figure 5b).

Serving as both the hydrogen-bond acceptor and the hydrogen-bond donor, the nitrogens are key pharmacophoric points interacting with the hinge region of BRAF[48] (PDB entry 3OG7) and ring fragments having one hydrogen-bond donor and one to three nitrogen acceptors are selected by exporting the ring fragments in the blue bordered cells in the heat map shown in Figure 5b. In the next step, we select the ring systems having the two key pharmacophoric points at a distance of two bonds (the *x*-axis in the heat map in Figure 5c) and an aromatic carbon bearing a substitution separated by two bonds from the hydrogen-bond acceptor nitrogen (the *y*-axis in the heat map in Figure 5c), thus having the same spatial arrangement of the key features as the 1*H*-pyrrolo[2,3-*b*]pyridine scaffold of Vemurafenib. Examples of 15 feasible alternatives to this scaffold, selected from the 264-membered cell in Figure 5c, are shown in Figure 5d. A literature search was performed to examine how many of the 15 ring systems were found as scaffolds in BRAF inhibitors. Individual substructure searches, constraining the required substitution pattern, were performed in SciFinder.[49] Scaffolds **1** and **3** were found in BRAF inhibitors (refs 50 and 51, respectively). Scaffolds **2**, **4**, **7**, **9**, **10**, **11**, **13**, and **15** have been described in other kinase inhibitors. Compounds having the substructure of scaffolds **5**, **6**, **8**, **12**, and **14** either have not been described as kinase inhibitors or do not have any reported biological activity (e.g., scaffold **12**). Thus, with just a few clicks (the whole visual procedure is illustrated in Figure S3 in the Supporting Information), it is possible to obtain new ideas for designing novel molecules that might have potential as Vemurafenib bioisosters. This scaffold hopping strategy can be further followed by a selection of R-groups at the R1 and R2 positions (Figure 5a) using our previously reported web-based tool for R-group navigation[12] to identify the optimal substitution pattern extracted from the reported close analogues in the patents (SAR transfer) and to synthesize the most promising new compound bearing an alternative scaffold.

2. Scaffold Hopping of Melanin Concentrating Hormone 1-Receptor Antagonists (MCH-R1) Running SFP-Based Similarity Searches. Twenty-three pharmaceutical companies have investigated the use of melanin-concentrating hormone receptor 1 (MCH-R1) antagonists for the treatment of obesity, depression, and anxiety since the discovery of the pioneering nonpeptide MCH-R1 antagonist T226296 by Takeda in 2001.[52] Five compounds have proceeded into clinical testing, although none has advanced to long-term efficacy and safety studies.[53] Thus, there is a great number of structurally diverse chemotypes available to validate retrospectively our strategy for chemotype hopping. A set of 35 unique central scaffolds extracted from 53 representative MCH-R1 antagonists covering the patent literature to 2014 was compiled. A subset of 15 diverse cores selected among the 35 scaffolds is shown in Figure 6. The chemical structures of the 53 MCH-R1 representatives, CAS registration number, patent code source, and proprietary pharmaceutical company are

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

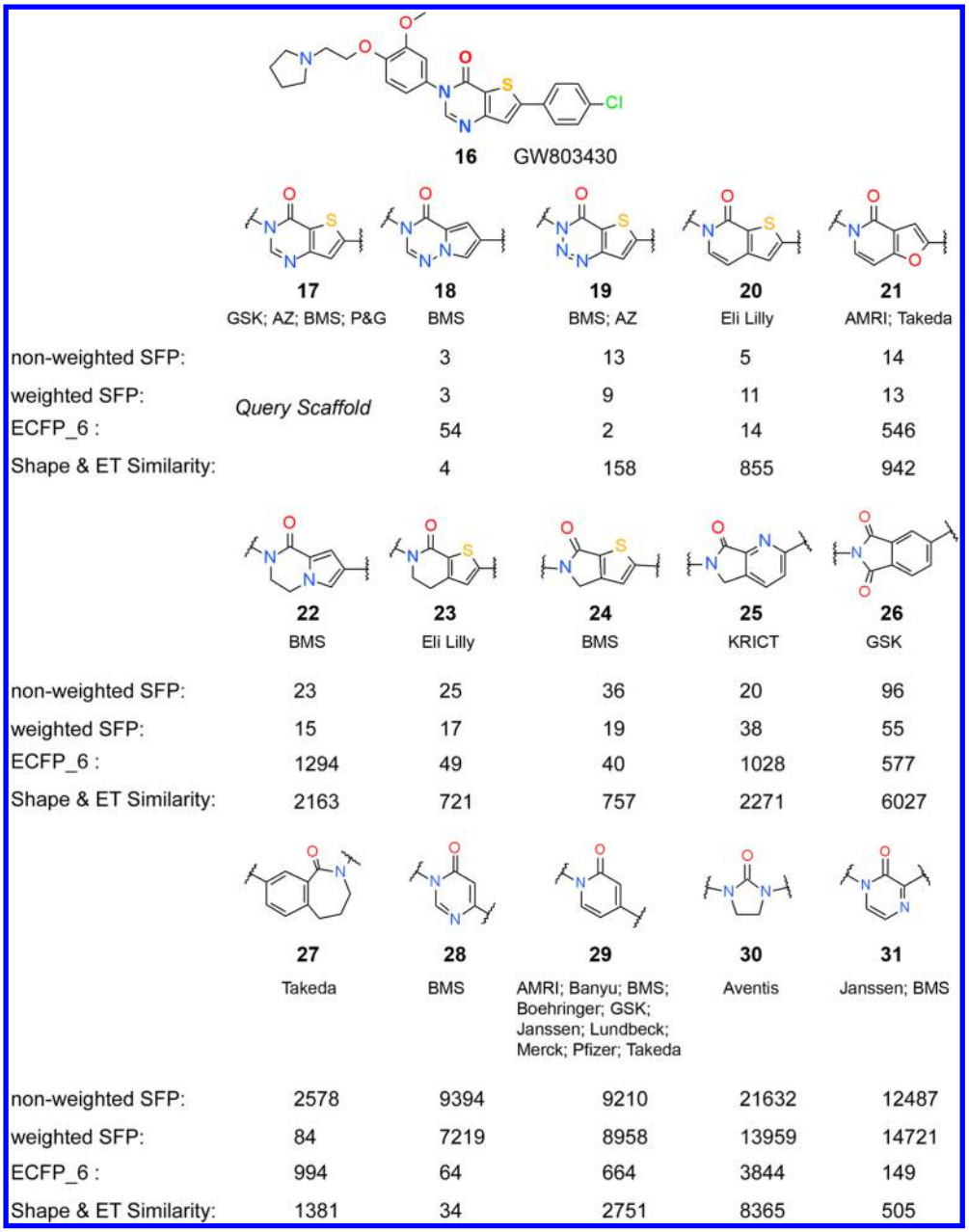


Figure 6. Clinical candidate, reference compound **16** GW803430.[54] Chemical structures, labeled with their connection points, of 15 representative diverse scaffolds among the 35 scaffolds extracted from known MCH-R1 antagonists that conform to our validation set. Information on the proprietary companies of each chemical series is shown. Numbers below each scaffold correspond to the ranking position at which each scaffold was detected using the indicated approach.

provided as Supporting Information (sheet MCH-R1 Antagonists 53 reference).

As in our previous 3D scaffold hopping approach,[7] the chemotype of the GlaxoSmithKline clinical candidate GW803430,[54] compound **16** in Figure 6, was chosen as the query scaffold (core **17**). As this scaffold has two substitutions, we focused the selection on the 33 925 ring systems bearing two connection points from the compiled database of 157 533 unique synthetically accessible scaffolds. Scaffolds were ranked by decreasing Tanimoto similarity, with either (i) the default uniformly weighted version of SFP (nonweighted SFP) and (ii) with the dimensions of the fingerprint accounting for the similarity in topological distances between connection points and also between connection points and acceptors up-scaled by ten (weighted SFP).

In a real scenario, a few hundred alternative chemotypes are typically analyzed in close detail from the viewpoint of synthetic strategy and IP freedom. Therefore, we set a maximum value of 500 retrieved potential hits (~1.5% of the whole set of scaffolds with two connection points). The corresponding enrichment factor curve is shown in Figure 7. For comparison, we also performed a search with the same query and the same ring database using the ECFP_6 fingerprints and the 3D shape and electrostatics similarity approach. Concerning the 3D approach, the query scaffold **17** was decorated with methyl and phenyl groups according to the substitution pattern in GW803430.[7] Both 2D fingerprint methods (with either the weighted and

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

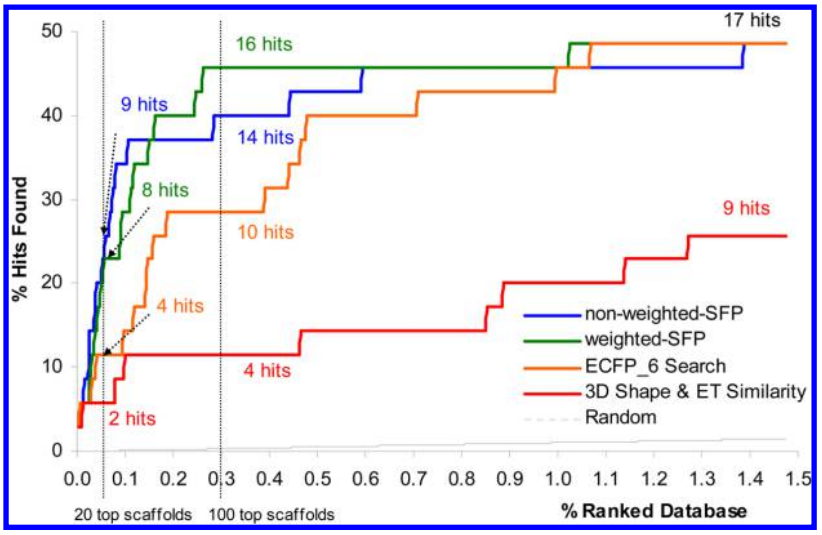


Figure 7. Boost plot of the percentage of GW803430 bioisosteres identified versus percentage of the screened database of scaffolds by four different approaches: SFP (nonweighted and weighted), ECFP_6, and 3D shape and electrostatic similarity.

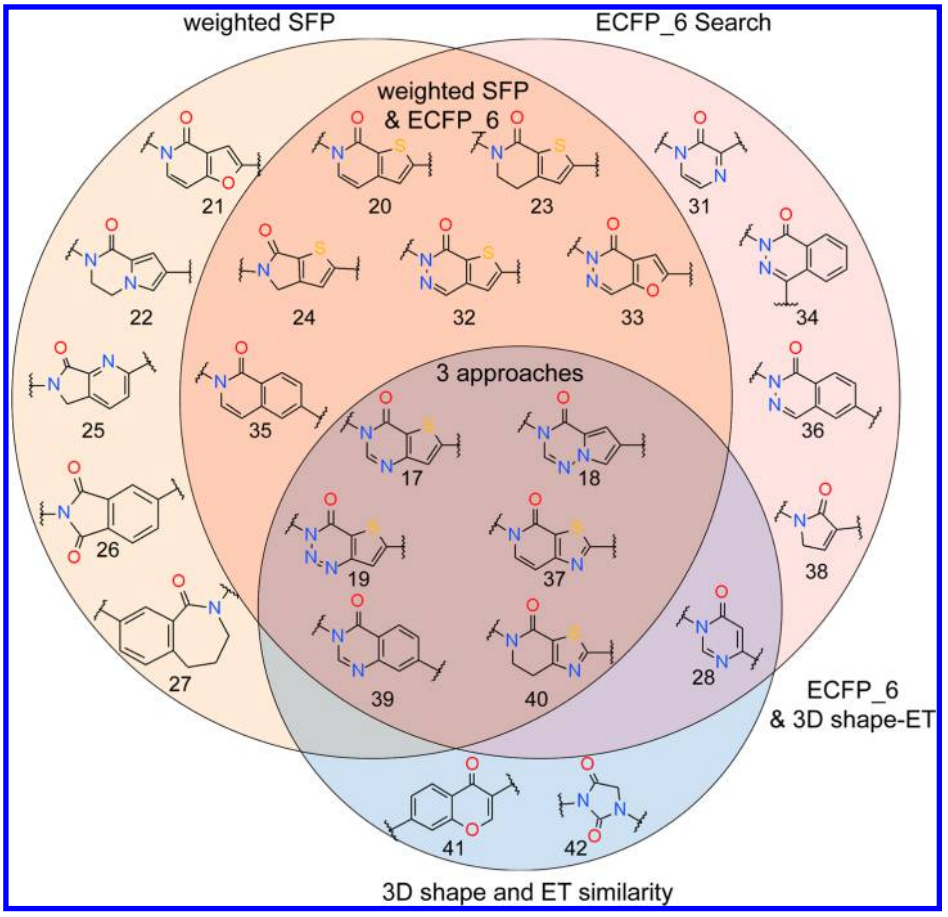


Figure 8. Overlap between the three different methods (weighted SFP, ECFP_6, and 3D shape and electrostatics similarity) of MCH-R1 hit scaffolds found among the 500 top ranked scaffolds (1.5% database).

nonweighted SFP and ECFP_6) identified 17 of the 35 scaffold hits (recall = 48.6%, EF = 33) within the top ranked 500 scaffolds, whereas only 9 of the 35 hits were retrieved by the 3D approach (recall = 25.7%, EF = 17). As shown in Figure 7, our method outperformed the ECFP_6 search at lower percentages of screened database. Thus, within the top 100 scaffolds (~0.3% of the ranked library), 16, 14, and 10 hits are retrieved by our weighted SFP, nonweighted SFP, and the ECFP_6 search, respectively (corresponding to recall values of 45.7%, 40%, and 28.7%, respectively). Thus, at this 0.3% of screened database, enrichment factor (EF) for our approach is 155 (weighted SFP) or 136 (nonweighted SFP) compared to 97 for the ECFP_6 approach. In contrast, the 3D approach identified 4 out of the 35 scaffolds within the top 100 scaffolds (recall =

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

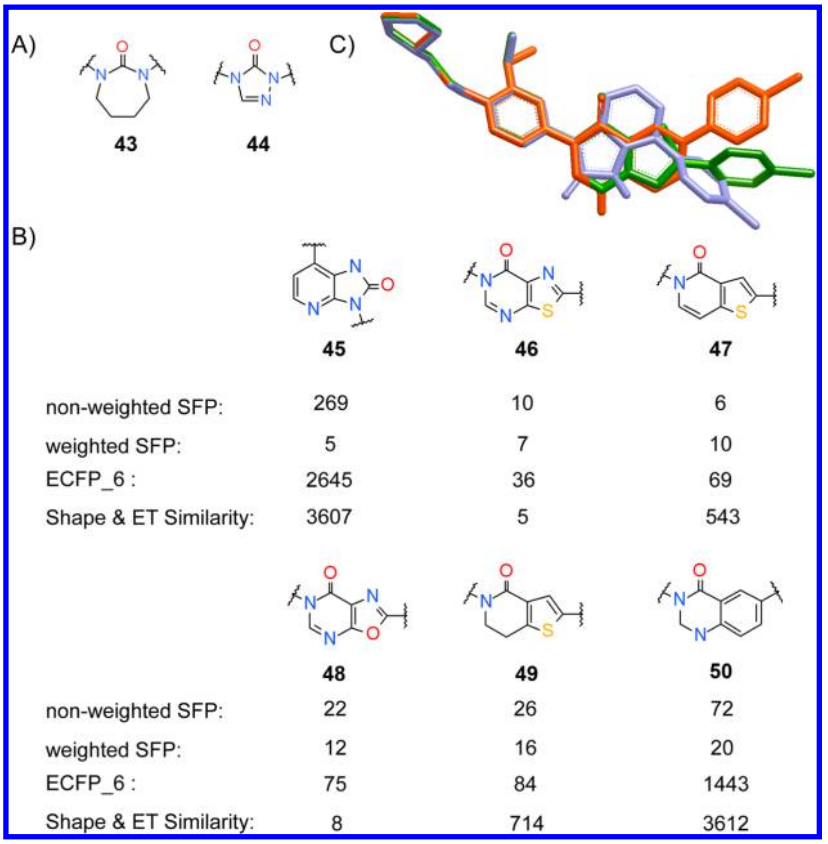


Figure 9. (A) MCH-R1 additional scaffolds found when focusing on monocycles. (B) Six novel potential bioisostere scaffolds of GW803430 ranked among the top 20 hits using weighted SFP. Numbers below each scaffold correspond to the ranking position at which each scaffold was detected using the indicated approach. (C) 3D superimposition of GW803430 (**16**, green) with the derivatized analogs of the proposed scaffolds **45** (violet) and **50** (orange) bearing the same R-groups as GW803430—SAR transfer.

11.4%, EF = 39). Please note that differences with previously reported values in ref 7 are due to the different composition of the scaffold pool and the validation set, which has been extended in this work. The rank position and Tanimoto similarity retrieved by the nonweighted and weighted SFP, ECFP_6, and shape and electrostatics-based searches for each of the 35 target scaffolds are provided in the Supporting Information (sheet MCH-R1 Antagonists 53 reference). Also, for comparison, the ranking position obtained by each method for each scaffold in Figure 6 is indicated. In terms of overlapping chemistry, the ECFP_6 search identified scaffold **28** as the only new scaffold within the top 100 scaffolds. By comparison, this scaffold was retrieved by our methodology at rank 7219 (weighted version) or rank 9394 (nonweighted run). The weighted SFP version performed slightly better than the nonweighted search, especially at 0.3% or ranked library (Figures 6 and 7). The overlap of the 24 total unique hits identified by the weighted SFP, ECFP_6, and shape and electrostatics similarity methods at 1.5% of the ranked library is shown in Figure 8 and the corresponding overlap of the 19 unique his retrieved by the nonweighted and weighted SFP searches is given in Figure S4 of the Supporting Information. Six scaffolds are found by the three approaches (also detected by the uniformly weighted SFP, Figure S4) and another six scaffolds are common to the SFP and ECFP_6 fingerprints (all of them, except for scaffold **35**, are also detected by the uniformly weighted SFP).

The common pools in Figure 8 have both close structural analogues to the query scaffold **17** with the same 6,5-fused bicycle topology (i.e., **18**, **19**, **20**, **23**, **37**, **40**) and less obvious bioisosteres with different topologies (i.e., 5,5-fused bicycles (**24**) and 6,6-fused bicycles (**35**, **39**). A notable degree of structural diversity is also found in the noncommon pool: 5,6-fused bicycles (**25**, **26**); 6,7-fused bicycles (**27**) detected using weighted SFP; and monocycles retrieved by the ECFP_6 search alone (**31**, **38**) or in common with the 3D approach (**28**). Notably, the 3D approach provides a reduced (lower EF) but diverse set of hits: scaffolds **41** and **42** are only retrieved with this approach within the top 500 hits. The ECFP_6 search and 3D scaffold hopping approach rank monocycles before other bicycles because no 2D overlap in the exit vectors is imposed. In this sense, the influence of the weights on the exit vectors is observed when comparing hits retrieved by either the uniformly weighted SFP search (e.g., scaffold **34**) or the weighted SFP search (scaffolds **27** and **35**), as shown in Figure S4. Because of this restriction, monocycles are ranked very low by the weighted SFP search. As an alternative, the user could concentrate the selection on the monocycles using the visual interface and remove the weights assigned to the distance between the connection points. If this is done, scaffolds **28** and **31** are ranked at the 30th and 40th position, respectively. Moreover, two additional monocycles (**43** and **44**) not contained in our validation set, found as chemotypes in MCH-R1 antagonists as reported in Scifinder,[49] are ranked at the fourth and seventh position, respectively (Figure 9a and sheet MCH-R1 100 best monocycles in the Supporting Information).

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

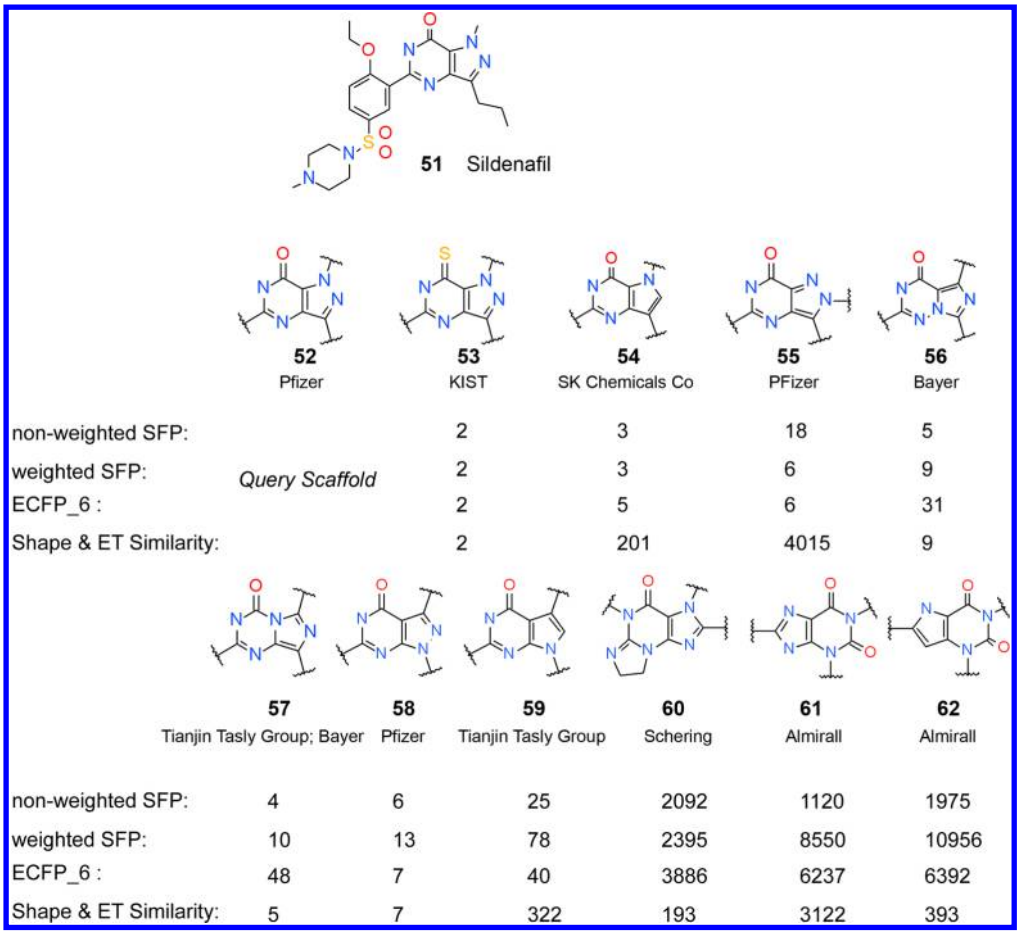


Figure 10. Sildenafil, reference compound **51** with query scaffold **52**, and 10 chemical structures of sildenafil bioisosteres, labeled with their connection points. Information on the proprietary companies of each chemical series is shown. Numbers below each scaffold correspond to the ranking position at which each scaffold was detected using the indicated approach.

A careful search of the literature was performed to determine how many of the top 100 ring systems ranked by each methodology were in fact found as scaffolds in MCH-R1 antagonists (provided in the Supporting Information, sheet MCH-R1 100 best scaffolds_SFP). This search was performed because our validation set consists of scaffolds present in advanced or highly exemplified compounds in patent applications, most of which correspond to the main chemical series in the claimed Markush. However, patents usually contain a few exemplary compounds of less preferred chemotypes, and moreover, singleton compounds might be found in the literature. Thus, individual substructure searches where the required substitution pattern was constrained were performed using SciFinder[49] and successful matches were verified as reported MCH-R1 antagonists. Eighteen additional scaffolds (different from the 16 hits in the initial 35-membered validation set) are within the 100 top ranked compounds by the weighted SFP, which means that 34 out of the 100 top ranked bioisosteres are actually true positives. Moreover, 14 out of the 20 top ranked scaffolds have already been reported in known MCH-R1 antagonists (see sheet MCH-R1 100 best scaffolds_SFP in the Supporting Information). The other top six hits, which correspond to scaffolds **45**−**50** in Figure 9b, have not been reported in the literature as MCH-R1 antagonists; therefore, they might have potential as novel MCH-R1 antagonists. As it can be seen in Figure 9b, these six scaffolds are mostly ranked lower in the ECFP_6 and ET-ranked sets, specially scaffolds **45** and **50**, which are most likely, the most novel and less obvious analogs. The 3D overlap calculated with ROCS[46] of these two potential bioisosteres derivatized with the same R-groups as GW803430 with the query compound GW803430 is shown in Figure 9c. The three compounds have similar shape and the position of the R-groups points reasonably to the same directions.

Repeating the Scifinder searches for the 100 top ranked scaffolds by the other approaches yielded 14 (nonweighted SFP), 7 (ECFP_6), and 7 (shape and ET similarity) scaffolds already described in MCH-R1 antagonists that were not present in the initial validation set. Except for three out of the seven true positives identified by the shape and ET similarity searches, the rest of the scaffolds identified by this approach, ECFP_6, and the nonweighted SFP searches overlap with the scaffolds retrieved by the weighted SFP.

3. Scaffold Hopping of Phosphodiesterase-5 Inhibitors (PDE5) Running SFP-Based Similarity Searches. The pyrazolopyrimidine core **52** in Sildenafil **51** was used as query for our second validation of Q-SFP for scaffold hopping. The bioisosteric replacement of this chemotype is a well-known example in medicinal chemistry[55] that has also been used to validate the NEAT methodology developed at Pfizer.[5] Here, we compiled a set of 11 heteroaryl bioisosteres of sildenafil (Figure 10, available as Supporting Information PDE5 11 Inhibitors Reference) and focused the selection on the 30 170 ring systems with three exit vectors from the 157 533-membered

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

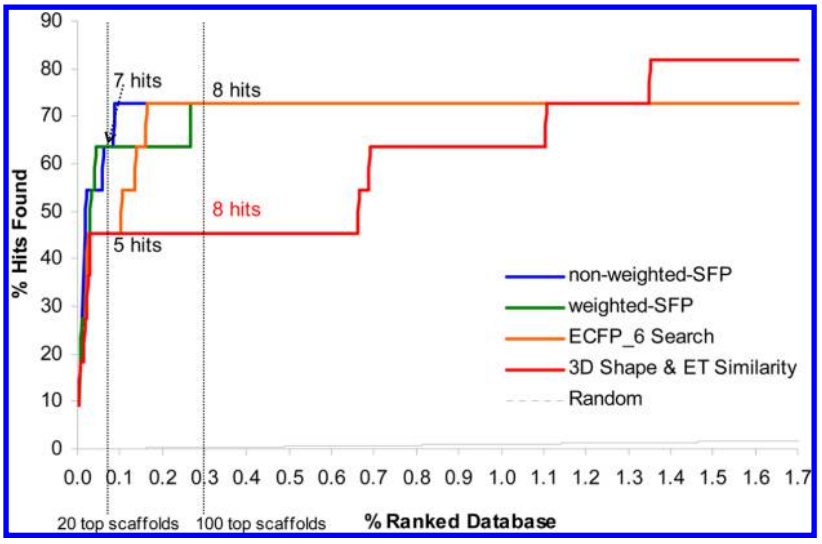


Figure 11. Boost plot of the percentage of sildenafil bioisosteres versus percentage of the screened database of scaffolds by four approaches: nonweighted SFP, weighted SFP, ECFP_6 and 3D shape and electrostatic similarity.

database. Tanimoto similarity was chosen as the classifying metric and two searches were run: (i) a uniformly weighted SFP search (nonweighted SFP) and (ii) a weighted SFP search with the dimensions up-scaled by ten of the fingerprint accounting for the similarity in topological distances between connection points, connection points, and acceptors and between hydrogen-bond donor and acceptors (weighted SFP). The corresponding enrichment factor curve is shown in Figure 11 for the top 500 ranked chemotypes (~1.7% of the ring systems with three exit vectors). Again, ECFP_6-based and 3D shape and electrostatics similarity searches were carried out for comparison purposes. For this last approach, scaffold **52** was conveniently decorated with methyl, propyl and phenyl moieties. Both 2D fingerprint-based methods (with either the weighted and nonweighted SFP and ECFP_6) identify the same number and the same hits within the 100 top ranked chemotypes (0.33% of ranked database): 8 out of the 11 bioisosteres (recall = 72.7%, EF = 219.4), whereas five hits (also detected by the 2D fingerprint searches) are retrieved by the 3D approach (recall = 45%, EF = 137). As in the MCH-R1 case, our SFP approach outperforms the ECFP_6 search at lower percentages of screened database: 7 of these 8 hits are within the 20 top ranked chemotypes (0.06% of ranked database, the same by either the weighted or nonweighted SFP run), in comparison with the 5 hits (also detected by SFP) retrieved by the ECFP_6 search and the 3D approach. For the ECFP_6-based search, the vardenafil scaffold **56** and its close analogue **57** are found at rank 31 and 48, respectively. In this case, imposing weights to the distance between the different pharmacophoric points and exit vectors has, on average, a slightly negative impact on the ranking position at which each scaffold is detected (Figures 10 and 11), compared to the uniformly weighted SFP search. This is probably due to the fact that the weights do not differentiate between the different nitrogen acceptors in the heterocycle, whereas, as demonstrated by the reported scaffolds, nitrogen swapping is widely accepted for this particular target (even for the loss of a hydrogen-bond of the N2 of the pyrazole ring with a crystallographic water in scaffold **59**).[55]

With the NEAT approach,[5] scaffolds **52**, **56**, **57**, and **58** were found within the 12 ranked hits and no data were given for the remaining 7 other scaffolds in Figure 10. Again, a careful search of the literature for the top 20 SFP hits identified the pyrazolo[4-3-*e*][1,4]diazepin-8-one ring system at rank 4 (weighted SFP) and rank 13 (uniformly weighted SFP), a scaffold not present in the validation set but one that has already been described in selective PDE2 and PDE4 inhibitors.[56] This scaffold was also detected by the ECFP_6 search at rank 3 but not retrieved by the shape and ET similarity method. Interestingly, ECFP_6 was the only method that also retrieved the 1*H*-quinazolin-4-one scaffold (rank 14) as a new true positive[57] (not present in the validation set) and shape and ET similarity was the only approach that detected 3,7-dihydropurine-2,6-diones (rank 19), a scaffold also reported in PDE5 inhibitors.[57] These results highlight the complementarity of the different approaches for scaffold hopping.

4. Scaffold-Based Selection of Diverse Compounds from Commercial Libraries Using Q-SFP. As introduced, SFP can also be used to select diverse compounds in commercial libraries. The underlying goal of the proposed scaffold-based selection of commercial compounds is to add differential value to corporate libraries: diversity in chemotypes (IP), diversity in biologically relevant chemical space (diversity in biological response as well as in primary activities to ADME), and to generate a rapid preliminary structure activity relationship (SAR) when postprocessing raw HTS data. Compound-based selections might be populated with single exemplars of a single scaffold, which may complicate hit confirmation and selection of a candidate series for further development. Thus, we propose the following workflow when acquiring external compounds for library enrichment: first, collect a set of diverse scaffolds using the approach presented in this paper (see below), and second, select a set of representative compounds for each scaffold, covering as many as possible of the 17 ligand−receptor interaction classes for the R-groups at different substitution sites, using our previously published methodology.[11,12] Although there is a lack of consensus on the number of members of a scaffold class that are needed to fully represent a scaffold for HTS screening,[58] the minimal reference value of five representatives proposed by McFayden et al.[59] may be considered a reasonable alternative to cover the 17 ligand−receptor interaction classes for each attachment point class.[12]

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

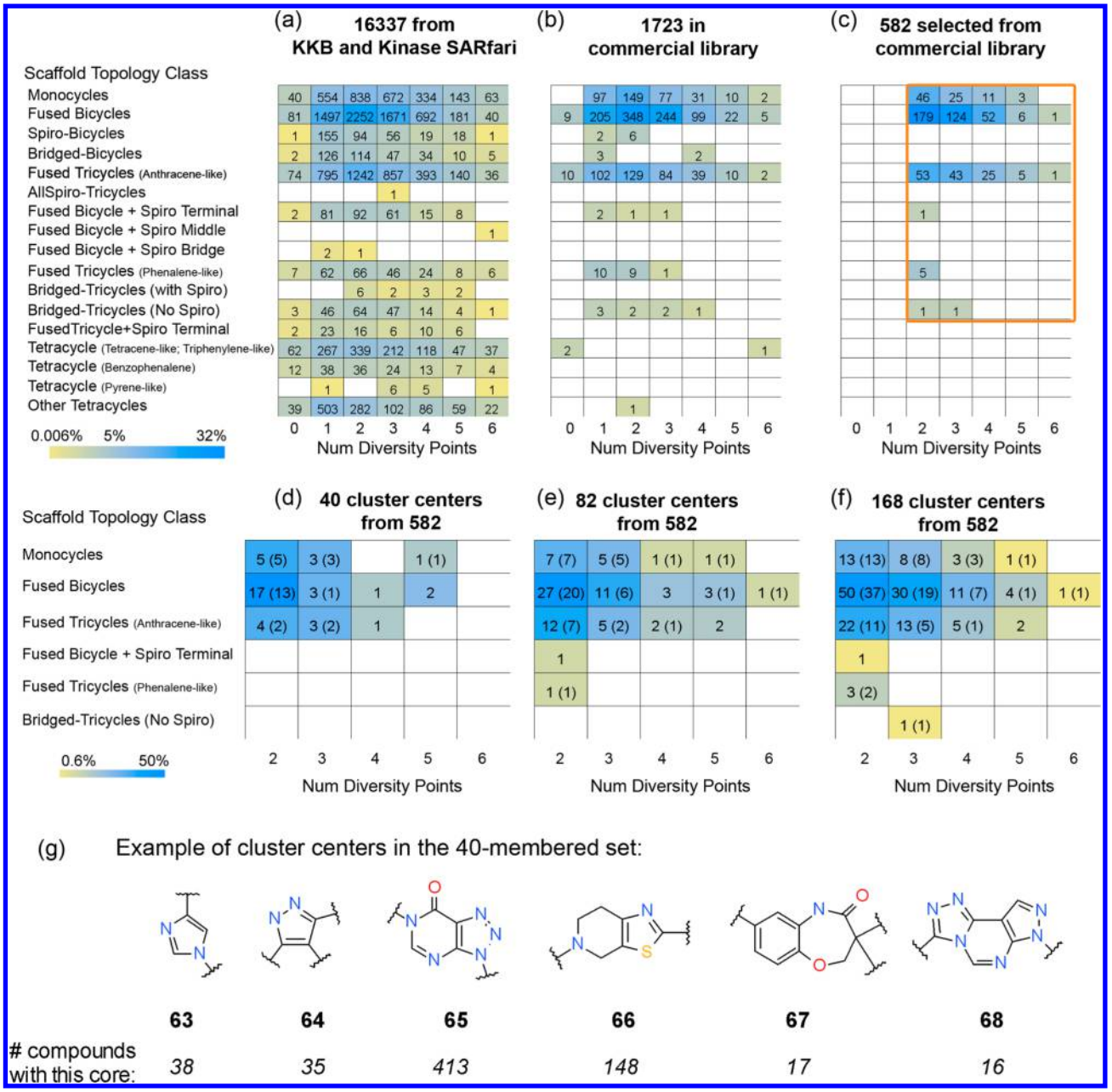


Figure 12. Heat map showing the distribution of ring fragments extracted from (a) the set of reported kinase inhibitors from KKB and Kinase SARfari; (b) the commercial kinase-focused library from Life Chemicals; and (c) a subset of the 582 most interesting fragments from these 1723 rings systems in part b according to their number of diversity points (*x*-axis) and scaffold topology class (*y*-axis). Heat maps are color coded by cluster coverage (in percentage, logarithmic scale) and labeled with the total number of ring fragments. Parts d–e show the distribution of the scaffolds in the cluster centers obtained after clustering the set of 582 scaffolds in part c with different values of maximum distance within each cluster. Labels correspond to the number of scaffolds and, in parentheses, the number of overlapping scaffolds within the set of scaffolds extracted from KKB and Kinase SARfari sets. (g) Example of six cluster centers from the 40-membered diverse set, with the number of compounds in the commercial library having this scaffold.

To illustrate this proposal for library enrichment, a commercial kinase-focused library was subjected to the workflow described above. The combination of two described data sets of reported kinase inhibitors (i.e., Kinase Knowledge Database (KKB)[26] and Kinase SARfari)[24] consisting of a total of 658 721 compounds and 21 144 unique scaffolds served as the frame reference to test whether selected scaffolds from the commercial library have already been reported in kinase inhibitors.

Three kinase-focused libraries were downloaded from the Life Chemical vendor[60] and compounds were standardized as described previously,[11] resulting in a collection of 75 374 unique compounds. With one exception, all of the compounds have at least one ring system. The fragmentation procedure yielded a total of 1723 unique ring systems. The compound overlap of this commercial library with the combination of KKB and Kinase SARfari is 0.88%: only 665 out of the 75 374 commercial compounds have already been described in these kinase inhibitor sets. In terms of scaffolds, the overlap is much higher (at 67.7%): 1166 out of the 1723 ring systems are present within the known kinase inhibitors. The distribution of the 1723 ring systems according to its number of diversity

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1–18

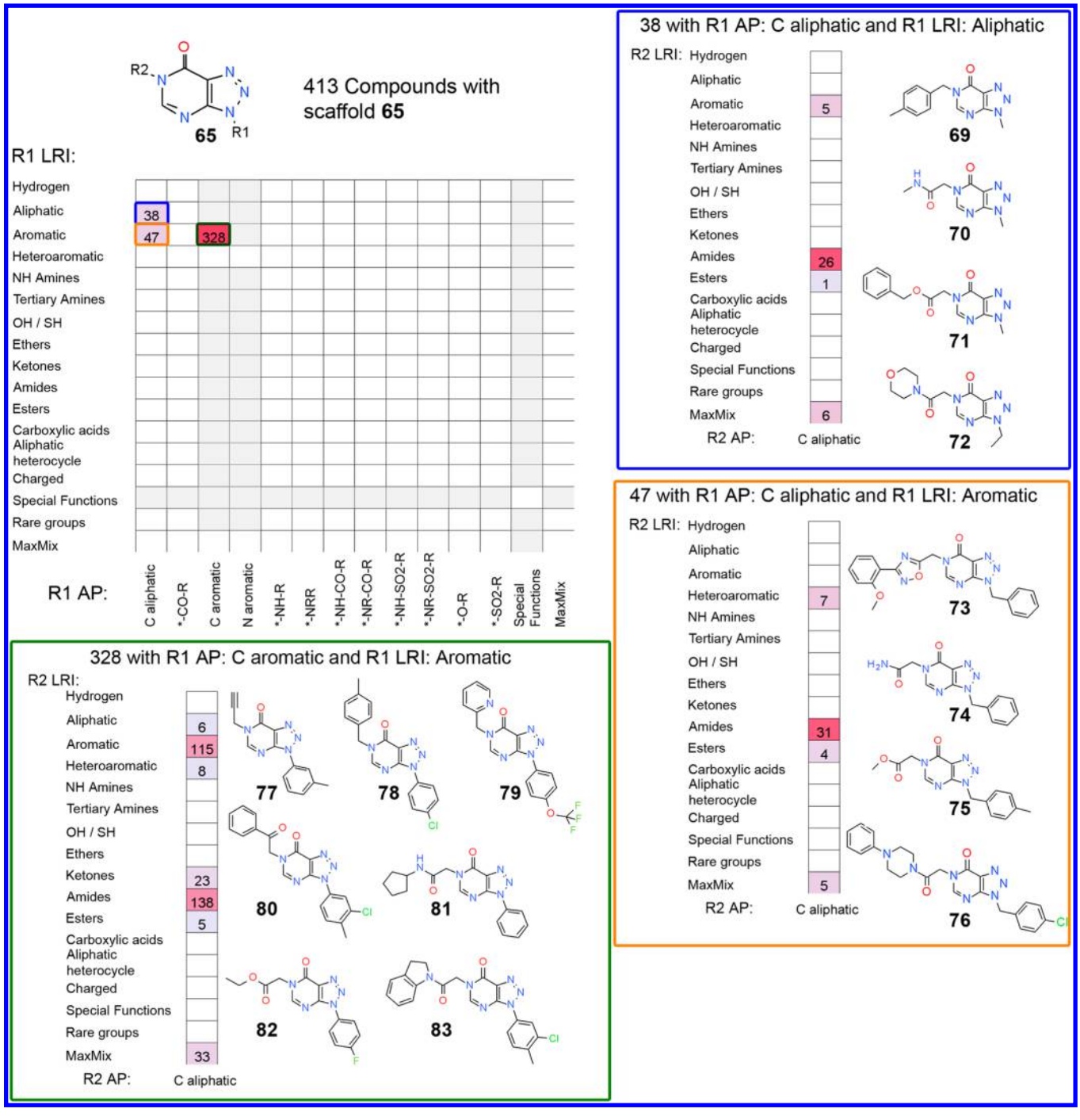


Figure 13. Example of representative selection of 15 compounds from the 413-membered set of commercial compounds with scaffold **65** using our interactive tool based on LiRIF.[11,12] The left top heat map displays the functionality of the attachment point (AP, *x*-axis) against the substitution pattern beyond this linker (LRI, *y*-axis) at the R1 position. The total number of compounds within each cell is shown. By clicking on each of the three sampled cells in this map, a new tab opens that allows the exploration of R-groups at the R2 position. Each of the three color-bordered panels corresponds to the same colored cell in the left top heat map. Each of the 15 selected compounds is a representative compound of its closest cell in these heat maps.

points versus Scaffold Topology Class is shown in Figure 12. The commercial library consists of ring assemblies with a maximum of up to four rings and with up to six substitution sites. The same representation is shown in Figure 12a for the subset of 16 337 scaffolds extracted from the set of reported kinase inhibitors having these property ranges. Ring systems in the commercial library are simpler in terms of topological class and substitution patterns: most rings are in fact monocycles, fused bicycles, and fused tricycles with 2-, 3-, and 4-growing vectors, whereas other topological classes with less common features (e.g., spiro centers, nonclassical fusing, or tetracycles) are under sampled or missing.

From a practical perspective, the most interesting ring systems among the 1723 are those having: (i) at least two substitution patterns; (ii) an occurrence frequency ranging between 5 (according to the McFayden reference value)[59] and

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

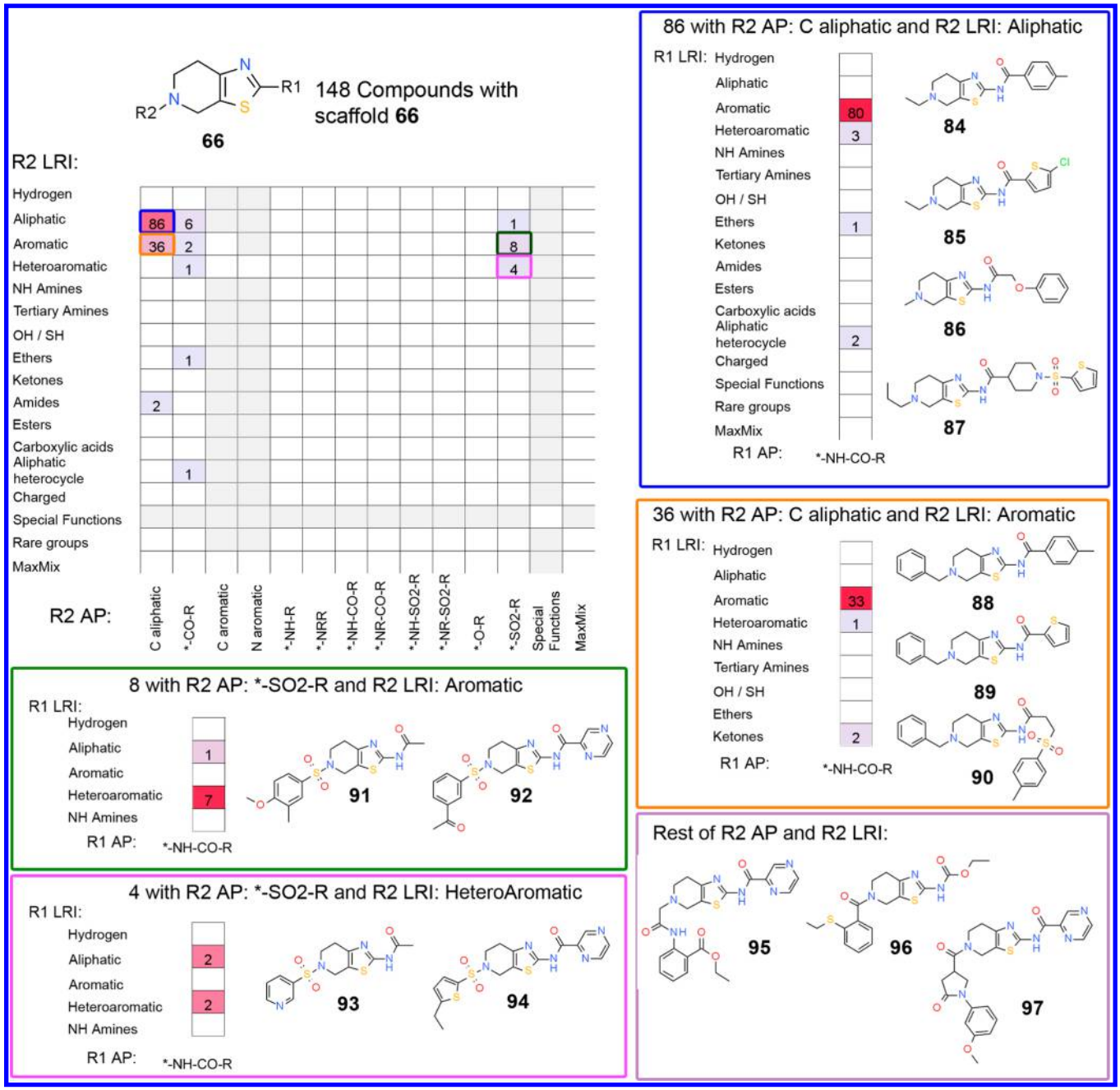


Figure 14. Example of representative selection of 14 (~10%) compounds from the 148-membered set of commercial compounds with scaffold **66** using the automated code to select a representative from each combination of LiRIF properties[11,12] for each substitution site. The left top heat map displays the functionality of the attachment point (AP, *x*-axis) against the substitution pattern beyond this linker (LRI, *y*-axis) at the R2 position. The total number of compounds within each cell is shown. For display purposes, the corresponding map for the R1 position is shown with some selected examples.

5000 (to discard highly occurring rings systems without true chemotype interest such as different substitution patterns of phenyl rings); and (iii) at least one aromatic ring because we are considering kinase inhibitors and its preference for aromatic planar rings is well-known.[61] A set of 582 ring systems remained after filtering out rings that did not satisfy these requirements (Figure 12c). Of these, 435 of them (74.7%) are already present in the set of scaffolds from known kinase inhibitors. The 582 ring systems were clustered such that different values of maximum distance between cluster members and cluster centroid were imposed; that is, with maximal *1-Tanimoto similarity coefficient* values of 0.1, 0.15, and 0.2, which resulted in 168, 82, and 40 clusters, respectively. Scaffolds in cluster centers conform the diverse final set of scaffolds. As shown in Figures 12d–f, these three sets cover reasonably well the initial space of the 582 scaffolds (bordered in orange in Figure 12c) in terms of the number of diversity points and topology class. Additionally, as exemplified in Figure 12g, scaffolds in cluster centers are structurally diverse in terms of pharmacophoric features, ring sizes, aromatic nature and orientation of exit vectors. Moreover, the rate of previously described scaffolds in kinase inhibitors remains almost constant: 27 out of 40 (67.5%), 53 out of 82 (64.6%), and 110 out of 168 (65.5%) are similar to the initial set of 1723 scaffolds (67.7%).

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1–18

Following this proposal, one out of three acquired scaffolds would be novel in our library (note that different substitution patterns are regarded as different scaffolds). If the substitution patterns are ignored, the initial set of unique scaffolds is reduced to 784 (commercial library) and 7942 (reported kinase inhibitors). 595 out of these 784 scaffolds in the commercial library have already been reported in kinase inhibitors (75.9% of overlap). This analysis shows that the novelty in chemistry is not only focused on generating novel ring systems (one out of four scaffolds would be chemically novel) but also on exploring different substitution patterns around reported scaffolds. This overlap between the selected diverse sets of ring systems and scaffolds in reported kinase inhibitors is also increased when discarding the substitution patterns: 32 out of 40 (80%), 63 out 82 (76.2%), and 117 out of 168 (69.6%) chemotypes in the SFP-clustered diverse sets have already been described in kinase inhibitors.

As mentioned, in a second step one would select representative compounds for each scaffold having different side chains at each substitution site according to our previously described LiRIF descriptor,[11,12] which has been implemented as an interactive tool. Briefly, for each diversity site, the user can represent the functionality of the attachment point of R-groups (*x*-axis of heat maps in Figures 13 and 14) against the nature of the substitution pattern beyond the attachment point (*y*-axis of heat maps in Figures 13 and 14). In Figure 13, the left top map represents the R-groups at position R1 for the 413 representative compounds in the commercial library bearing scaffold **65** in Figure 12g. The user identifies three available substitution patterns at R1: aliphatic carbons, aromatic carbons directly bonded to the scaffold, or aromatic carbons attached via an aliphatic linker. From our perspective, it would be desirable to acquire at least one representative from each of these three classes to fully sample the commercially available options for scaffold **65** in terms of potentially different ligand−receptor interactions at R1. In a second round, by clicking on each cell, the user may then inspect the R-groups at R2 in an analogous way: heat maps bordered in blue, orange, and green in Figure 13, with each one originating from the corresponding colored-bordered cell in the left top heat map. A selected set of representative compounds is shown next to each cell. In total, we propose that these 15 compounds would properly represent the 413 compounds with scaffold **65**. This is an example of our proposed strategy for acquiring external compounds in commercial libraries. Neither these 15 compounds nor the scaffold **65** have been reported as kinase inhibitors; therefore, they might have potential as novel kinase inhibitors.

To avoid manual visual selection of compound representatives with the graphical user interface, a very simple code was written to choose a number of representative compounds from the most populated combinations of all substitution sites of attachment points (AP) and ligand−receptor interactions (LRI) classes until the number of user-required representatives is reached or until at least one representative for each combination is identified. This is illustrated in Figure 14 for scaffold **66** with 148 representatives in the commercial library. For illustrative purposes, the corresponding maps of each substitution site, especially at R2, are shown. None of these 14 representatives have been described as kinase inhibitors. However, compounds bearing heteroaromatic rings directly bonded at R2 of scaffold **66**, a position not covered by any compound in the 148 commercial set, have been claimed as PI3K inhibitors by Vertex.[62]

CONCLUSIONS

SFP capitalizes on the biologically relevant information on the chemical space occupied by ring systems; therefore, SFP provides a promising underlying metric to perform bioisosteric identification as well as to select chemotypes with diverse biological profiles.

A methodology for the fast and simple identification of bioisosteric scaffolds is presented and validated using two well-established study cases. High enrichment factors are obtained in both cases using SFP. This performance surpasses that of classical 2D similarity searches using ECFP_6 fingerprints and 3D shape and electrostatic similarity searches at low percentages of the screened database. Scaffold hopping involves a subsequent synthetic effort, which requires the development of customized molecules according to SAR transfer; thus, we need highly efficient predictive methods that only focus on top ranked chemotypes. From a pragmatic perspective, top ranking area coverage is the most valuable scenario when validating scaffold hopping approaches and, indeed, prospective candidates. At higher percentages, both 2D fingerprint methodologies are comparable in terms of enrichment factor curves but complementary from the viewpoint of chemical diversity. Moreover, our approach has the potential to identify novel, structurally differentiated, chemotypes. As it relies on scaffolds derived from existing compounds, synthetic accessibility is also contemplated. The underlying scaffold description is easy to code and searches are fast, taking only a few seconds on a standard PC computer. This is a clear advantage over 3D approaches; indeed, the conformer enumeration of the decorated scaffolds in the MCH-R1 and PDE5 case studies took approximately 7 and 13 h on an Intel Core i5 CPU at 2.67 GHz, respectively. Moreover, this web-based tool does not require any specific knowledge of computational techniques, which makes it amenable for medicinal chemists. As shown in the test case of the Vemurafenib scaffold, this tool provides an intuitive guide as an idea generator for design by hand of potential bioisosters.

A disadvantage of this and other heterocycle-based approaches is that other traditional transformations in medicinal chemistry for finding bioisosteres (e.g., ring opening/ring closure) are not considered. Also, this methodology is not suitable for dealing with acyclic scaffolds.

Finally, a strategy for scaffold-based selection of compounds in commercial libraries is presented that combines our methodology for selecting diverse scaffolds and our previously described approach to navigate the R-groups at different substitution sites.[11,12] This is illustrated for two chemical series (chemotypes) extracted from a commercial library of potential kinase inhibitors. One of them is a novel scaffold nonreported in kinase inhibitors, whereas the other has already been patented against PI3K[62] with a different substitution pattern from that covered in the commercial library. We believe this SFP scaffold-based approach to be very intuitive for a medicinal chemist and will be very useful when acquiring external libraries for HTS because it adds a differential value to corporate libraries: diversity in chemotypes (IP), diversity in biologically relevant chemical space (diversity in biological response), and because it ensures a number of initial representatives for each chemical series, which is a quick assessment of its validity for further development. In summary, this will positively impact the chemotype/hit ratio that, together with the classical hit rate, is a key metrics for any HTS campaign. The case presented

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

here is a hypothetical situation. In a real scenario, the number of acquired scaffolds (i.e., number of clusters) and representative compounds for each scaffold will primarily depend on the availability, budget, logistics, and screening capabilities.

Readers interested in this interactive web-based application (Biologically Relevant Chemical Space navigator, BRCS), not only for ring systems analysis but also for R-groups, may contact the authors to obtain access.

■ ASSOCIATED CONTENT

Supporting Information

Table S1 lists the different scaffold topology classes. Figure S1 shows the process to deal with ring tautomerism. Figure S2 exemplifies the calculation of the property Scaffold Topology Size. Figure S3 provides details of the process to visually select bioisosteric replacements of the scaffold of Vemurafenib. Figure S4 shows the overlap of the 19 scaffolds found in MCH-R1 antagonists detected by the uniformly weighted version and weighted versions of SFP. An Excel file is included that contains the retrospective validation sets for the scaffold hopping examples. Additionally, the file includes the list of the top 100 ranked scaffolds (weighted SFP) and top 100 ranked monocycles for the case of the reference scaffold of GW803430, the MCH-R1 antagonist. Another Excel file is included that contains the top 100 ranked scaffolds by the weighted SFP-based search as smiles together with the SFP descriptor. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +34 948 194700. E-mail: julenoyarzabal@unav.es.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the Foundation for Applied Medical Research (FIMA) of the University of Navarra for financial support. Additionally, we thank Dr. Steven Muskal, Eidogen-Sertanty Inc., for providing access to the KKB database. O.R. was partially supported by MINECO and FSE (Inncorpora-Torres Quevedo grant), PTQ-1-04781. Finally, we would like to thank OpenEye Scientific Software for providing us with an academic license for use of its software.

■ REFERENCES

(1) Sun, H.; Tawa, G.; Wallqvist, A. Classification of Scaffold-Hopping Approaches. *Drug Discovery Today* **2012**, *17*, 310−324.

(2) Pitt, W. R.; Parry, D. M.; Perry, B. G.; Groom, C. R. Heteroaromatic Rings of the Future. *J. Med. Chem.* **2009**, *52*, 2952−2963.

(3) Lewell, X. Q.; Jones, A. C.; Bruce, C. L.; Harper, G.; Jones, M. M.; McLay, I. M.; Bradshaw, J. Drug Rings Database with Web Interface. A Tool for Identifying Alternative Chemical Rings in Lead Discovery Programs. *J. Med. Chem.* **2003**, *46*, 3257−3274.

(4) Ertl, P. Database of Bioactive Ring Systems with Calculated Properties and Its Use in Bioisosteric Design and Scaffold Hopping. *Bioorg. Med. Chem.* **2012**, *20*, 5436−5442.

(5) Tu, M.; Rai, B. K.; Mathiowetz, A. M.; Didiuk, M.; Pfefferkorn, J. A.; Guzman-Perez, A.; Benbow, J.; Guimarães, C. R.; Mente, S.; Hayward, M. M.; Liras, S. Exploring Aromatic Chemical Space with NEAT: Novel and Electronically Equivalent Aromatic Template. *J. Chem. Inf. Model.* **2012**, *52*, 1114−1123.

(6) Vainio, M. J.; Kogej, T.; Raubacher, F.; Sadowski, J. Scaffold Hopping by Fragment Replacement. *J. Chem. Inf. Model.* **2013**, *53*, 1825−1835.

(7) Oyarzabal, J.; Howe, T.; Alcazar, J.; Andrés, J. I.; Alvarez, R. M.; Dautzenberg, F.; Iturrino, L.; Martínez, S.; Van der Linden, I. Novel Approach for Chemotype Hopping Based on Annotated Databases of Chemically Feasible Fragments and a Prospective Case Study: New Melanin Concentrating Hormone Antagonists. *J. Med. Chem.* **2009**, *52*, 2076−2089.

(8) Saluste, G.; Albarran, M. A.; Alvarez, R. M.; Rabal, O.; Ortega, M. A.; Blanco, C.; Kurz, G.; Salgado, A.; Pevarello, P.; Bischoff, J. R.; Pastor, J.; Oyarzabal, J. Fragment-Hopping-Based Discovery of a Novel Chemical Series of Protooncogene PIM-1 Kinase Inhibitors. *PLoS One* **2012**, *7*, e45964.

(9) Ertl, P. Intuitive Ordering of Scaffolds and Scaffold Similarity Searching Using Scaffold Keys. *J. Chem. Inf. Model.* **2014**, *54*, 1617−1622.

(10) Oyarzabal, J.; Zarich, N.; Albarran, M. I.; Palacios, I.; Urbano-Cuadrado, M.; Mateos, G.; Reymundo, I.; Rabal, O.; Salgado, A.; Corrionero, A.; Fominaya, J.; Pastor, J.; Bischoff, J. R. Discovery of Mitogen-Activated Protein Kinase-Interacting Kinase 1 Inhibitors by a Comprehensive Fragment-Oriented Virtual Screening Approach. *J. Med. Chem.* **2010**, *53*, 6618−6628.

(11) Rabal, O.; Oyarzabal, J. Using Novel Descriptor Accounting for Ligand-Receptor Interactions To Define and Visually Explore Biologically Relevant Chemical Space. *J. Chem. Inf. Model.* **2012**, *52*, 1086−1102.

(12) Rabal, O.; Oyarzabal, J. Biologically Relevant Chemical Space Navigator: from Patent and Structure-Activity Relationship Analysis to Library Acquisition and Design. *J. Chem. Inf. Model.* **2012**, *52*, 3123−3137.

(13) Crisman, T. J.; Jenkins, J. L.; Parker, C. N.; Hill, W. A.; Bender, A.; Deng, Z.; Nettles, J. H.; Davies, J. W.; Glick, M. "Plate Cherry Picking": A Novel Semi-Sequential Screening Paradigm for Cheaper, Faster, Information-Rich Compound Selection. *J. Biomol. Screen.* **2007**, *12*, 320−327.

(14) Langdon, S. R.; Brown, N.; Blagg, J. Scaffold Diversity of Exemplified Medicinal Chemistry Space. *J. Chem. Inf. Model.* **2011**, *51*, 2174−2185.

(15) Broughton, H. B.; Watson, I. A. Selection of Heterocycles for Drug Design. *J. Mol. Graphics Modell.* **2004**, *23*, 51−58.

(16) Ertl, P.; Jelfs, S.; Mühlbacher, J.; Schuffenhauer, A.; Selzer, P. Quest for the Rings. In Silico Exploration of Ring Universe To Identify Novel Bioactive Heteroaromatic Scaffolds. *J. Med. Chem.* **2006**, *49*, 4568−4573.

(17) Nicolaou, C. A.; Tamura, S. Y.; Kelley, B. P.; Bassett, S. I.; Nutt, R. F. Analysis of Large Screening Data Sets via Adaptively Grown Phylogenetic-Like Trees. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1069−1079.

(18) Shelat, A. A.; Guy, R. K. Scaffold Composition and Biological Relevance of Screening Libraries. *Nat. Chem. Biol.* **2007**, *3*, 442−446.

(19) Over, B.; Wetzel, S.; Grütter, C.; Nakai, Y.; Renner, S.; Rauh, D.; Waldmann, H. Natural-Product-Derived Fragments for Fragment-Based Ligand Discovery. *Nat. Chem.* **2013**, *5*, 21−28.

(20) 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.

(21) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.

(22) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC-A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757−1768.

(23) ChEMBL. European Bioinformatics Institute (EBI): Cambridge, UK, 2010. ftp://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/releases/ (accessed June 18, 2014).

(24) Kinase SARfari database. European Bioinformatics Institute (EBI): Cambridge, UK, 2010. ftp://ftp.ebi.ac.uk/pub/databases/chembl/KinaseSARfari/releases/5.01/ (accessed June 18, 2014).

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18

(25) GPCR SARfari database. European Bioinformatics Institute (EBI): Cambridge, UK, 2010. ftp://ftp.ebi.ac.uk/pub/databases/chembl/GPCRSARfari/releases/3.00/ (accessed June 18, 2014).
(26) *Kinase KnowledgeBASE (KKB)*; Eidogen-Sertanty, Inc.: San Diego, CA, 2011.
(27) *Oncology KnowledgeBASE (OKB)*; Eidogen-Sertanty, Inc.: San Diego, CA, 2011.
(28) Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A. C.; Wishart, D. S. DrugBank 3.0: A Comprehensive Resource for 'Omics' Research on Drugs. *Nucleic Acids Res.* **2011**, *39*, D1035−D1041.
(29) Chen, C. Y. TCM Database@Taiwan: The World's Largest Traditional Chinese Medicine Database for Drug Screening In Silico. *PLoS One* **2011**, *6*, e15939; http://tcm.cmu.edu.tw/ (accessed June 18, 2014).
(30) *Pipeline Pilot*, version 8.5; Accelrys, Inc.: San Diego, CA, 2011.
(31) Chen, H.; Yang, Y.; Engkvist, O. Molecular Topology Analysis of the Differences Between Drugs, Clinical Candidate Compounds, and Bioactive Molecules. *J. Chem. Inf. Model.* **2010**, *50*, 2141−2150.
(32) Wester, M. J.; Pollock, S. N.; Coutsias, E. A.; Allu, T. K.; Muresan, S.; Oprea, T. I. Scaffold Topologies. 2. Analysis of Chemical Databases. *J. Chem. Inf. Model.* **2008**, *48*, 1311−1324.
(33) Pollock, S. N.; Coutsias, E. A.; Wester, M. J.; Oprea, T. I. Scaffold Topologies. 1. Exhaustive Enumeration up to Eight Rings. *J. Chem. Inf. Model.* **2008**, *48*, 1304−1310.
(34) Lipkus, A. H. Exploring Chemical Rings in a Simple Topological-Descriptor Space. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 430−438.
(35) Nilakantan, R.; Bauman, N.; Haraki, K.; Venkataraghavan, R. A Ring-Based Chemical Structural Query System: Use of a Novel Ring-Complexity Heuristic. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 65−68.
(36) Xu, Y. J.; Johnson, M. Algorithm for Naming Molecular Equivalence Classes Represented by Labeled Pseudographs. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 181−185.
(37) 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.
(38) Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem.* **2009**, *52*, 6752−6756.
(39) Taylor, R. D.; Maccoss, M.; Lawson, A. D. Rings in Drugs. *J. Med. Chem.* **2014**, *57*, 5845−5859.
(40) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. Scaffold-Hopping by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 2894−2896.
(41) Wagener, M.; Lommerse, J. P. The Quest for Bioisosteric Replacements. *J. Chem. Inf. Model.* **2006**, *46*, 677−685.
(42) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(43) *OMEGA*, version 2.5.1.4; OpenEye Scientific Software.: Santa Fe, NM, 2014; http://www.eyesopen.com (accessed June 18, 2014).
(44) Hawkins, P. C. D.; Skillman, G. L.; Warren, B. A.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572−584.
(45) Hawkins, P. C. D.; Nicholls, A. N. Conformer Generation with OMEGA: Learning from the Dataset and Analysis of Failures. *J. Chem. Inf. Model.* **2012**, *52*, 2919−2936.
(46) *ROCS*, version 3.1.1; OpenEye Scientific Software.: Santa Fe, NM, 2014. http://www.eyesopen.com (accessed June 18, 2014). (b) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. J. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50*, 74−82.
(47) *EON*, version 2.2.0.5; OpenEye Scientific Software.: Santa Fe, NM, 2014; http://www.eyesopen.com (accessed June 18, 2014).
(48) Bollag, G.; Hirth, P.; Tsai, J.; Zhang, J.; Ibrahim, P. N.; Cho, H.; Spevak, W.; Zhang, C.; Zhang, Y.; Habets, G.; Burton, E. A.; Wong, B.; Tsang, G.; West, B. L.; Powell, B.; Shellooe, R.; Marimuthu, A.; Nguyen, H.; Zhang, K. Y.; Artis, D. R.; Schlessinger, J.; Su, F.; Higgins, B.; Iyer, R.; D'Andrea, K.; Koehler, A.; Stumm, M.; Lin, P. S.; Lee, R. J.; Grippo, J.; Puzanov, I.; Kim, K. B.; Ribas, A.; McArthur, G. A.; Sosman, J. A.; Chapman, P. B.; Flaherty, K. T.; Xu, X.; Nathanson, K. L.; Nolop, K. Clinical Efficacy of a RAF Inhibitor Needs Broad Target Blockade in BRAF-Mutant Melanoma. *Nature* **2010**, *467*, 596−599.
(49) *Scifinder*, web version; Chemical Abstracts Service: Columbus, OH, 2012.
(50) Wenglowsky, S.; Ahrendt, K. A.; Buckmelter, A. J.; Feng, B.; Gloor, S. L.; Gradl, S.; Grina, J.; Hansen, J. D.; Laird, E. R.; Lunghofer, P.; Mathieu, S.; Moreno, D.; Newhouse, B.; Ren, L.; Risom, T.; Rudolph, J.; Seo, J.; Sturgis, H. L.; Voegtli, W. C.; Wen, Z. Pyrazolopyridine Inhibitors of B-RafV600E. Part 2: Structure-Activity Relationships. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 5533−5537.
(51) Cui, J. J.; Deal, J. G.; Gu, D.; Guo, C.; Johnson, M. C.; Kania, R. S.; Kephart, S. E.; Linton, M. A.; McApline, I. J.; Pairish, M. A.; Palmer, C. L. Pyrazole Compounds. PCT WO2009016460 (A2), July 28, 2008.
(52) Johansson, A. Recent Progress in the Discovery of Melanin-Concentrating Hormone 1-Receptor Antagonists. *Expert Opin. Ther. Pat.* **2011**, *21*, 905−925.
(53) MacNeil, D. J. The Role of Melanin-Concentrating Hormone and Its Receptors in Energy Homeostasis. *Front Endocrinol (Lausanne).* **2013**, *4* (49), 1−14.
(54) Hertzog, D. L.; Al-Barazanji, K. A.; Bigham, E. C.; Bishop, M. J.; Britt, C. S.; Carlton, D. L.; Cooper, J. P.; Daniels, A. J.; Garrido, D. M.; Goetz, A. S.; Grizzle, M. K.; Guo, Y. C.; Handlon, A. L.; Ignar, D. M.; Morgan, R. O.; Peat, A. J.; Tavares, F. X.; Zhou, H. The Discovery and Optimization of Pyrimidinone-Containing MCHR1 Antagonists. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4723−4727.
(55) Haning, H.; Niewoehner, U.; Schenke, T.; Lampe, T.; Hillisch, A.; Bischoff, E. Comparison of Different Heterocyclic Scaffolds as Substrate Analog PDE5 Inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3900−3907.
(56) Plummer, M. S.; Cornicelli, J.; Roark, H.; Skalitzky, D. J.; Stankovic, C. J.; Bove, S.; Pandit, J.; Goodman, A.; Hicks, J.; Shahripour, A.; Beidler, D.; Lu, X. K.; Sanchez, B.; Whitehead, C.; Sarver, R.; Braden, T.; Gowan, R.; Shen, X. Q.; Welch, K.; Ogden, A.; Sadagopan, N.; Baum, H.; Miller, H.; Banotai, C.; Spessard, C.; Lightle, S. Discovery of Potent, Selective, Bioavailable Phosphodiesterase 2 (PDE2) Inhibitors Active in an Osteoarthritis Pain Model, Part I: Transformation of Selective Pyrazolodiazepinone Phosphodiesterase 4 (PDE4) Inhibitors into Selective PDE2 Inhibitors. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 3438−3442.
(57) Tworowski, D.; Matsievitch, R. Heterocyclic Compounds and Uses Thereof in the Treatment of Sexual Disorders. PCT WO2007110868 (A2), March 28, 2007.
(58) Medina-Franco, J. L.; Martinez-Mayorga, K.; Bender, A.; Scior, T. Scaffold Diversity Analysis of Compounds Data Sets Using an Entropy-Based Measure. *QSAR Comb. Sci.* **2009**, *11*−*12*, 1551−1560.
(59) McFayden, I.; Walker, G.; Alvarez, J. Enhancing Hit Quality and Diversity within Assay Throughput Constraints. In *Chemoinformatics in Drug Discovery*; Oprea, T., Ed.; Wiley-VCH Verlag GmbH & Co.: Weinheim, Germany, 2005; pp 143−173.
(60) Life Chemicals: Niagara-on-the-Lake, Canada, 2014; http://www.lifechemicals.com (accessed June 18, 2014).
(61) Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. Through the "Gatekeeper Door": Exploiting the Active Kinase Conformation. *J. Med. Chem.* **2010**, *53*, 2681−2694.
(62) Aronov, A.; Bandarage, U. K.; Cottrell, K.; Davies, R.; Krueger, E.; Ledeboer, M.; Ledford, B.; Le Tiran, A.; Liao, Y.; Messersmith, D.; Wang, T.; Xu, J. Tetrahydrothiazolopyridine Inhibitors of Phosphatidylinositol 3-Kinase. PCT WO2010096389 (A1), February 16, 2010.

DOI: 10.1021/ci500542e
J. Chem. Inf. Model. 2015, 55, 1−18