JOURNAL OF CHEMICAL INFORMATION AND MODELING

Perspective

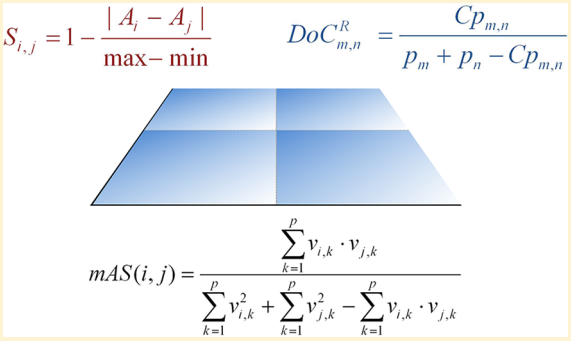
pubs.acs.org/jcim

Scanning Structure−Activity Relationships with Structure−Activity Similarity and Related Maps: From *Consensus Activity Cliffs* to *Selectivity Switches*

José L. Medina-Franco*

Torrey Pines Institute for Molecular Studies, 11350 SW Village Parkway, Port St. Lucie, Florida 34987, United States

ABSTRACT: Systematic description of structure−activity relationships (SARs) of data sets and structure−property relationships (SPRs) is of paramount importance in medicinal chemistry and other research fields. To this end, structure−activity similarity (SAS) maps are one of the first tools proposed to describe SARs using the concept of activity landscape modeling. One of the major goals of the SAS maps is to identify activity cliffs defined as chemical compounds with high similar structure but unexpectedly very different biological activity. Since the first publication of the SAS maps more than ten years ago, these tools have evolved and adapted over the years to analyze various types of compound collections, including structural diverse and combinatorial sets with activity for one or multiple biological end points. The development of SAS maps has led to general concepts that are applicable to other activity landscape methods such as "consensus activity cliffs" (activity cliffs common to a series of representations or descriptors) and "selectivity switches" (structural changes that completely invert the selectivity pattern of similar compounds against two biological end points). Herein, we review the development, practical applications, limitations, and perspectives of the SAS and related maps which are intuitive and powerful informatics tools to computationally analyze SPRs.



1. INTRODUCTION

Analysis of structure−activity relationships (SARs) and, more general structure−property relationships (SPRs), is of the main interest to medicinal chemists and groups working on different research areas. Over the years, qualitative and quantitative methods have been proposed to describe and predict the property of interest. Since the seminal paper of Maggiora[1] which clearly emphasizes the importance of activity cliffs (chemical compounds with high similar structure but unexpectedly very different biological activity) in the study of SARs, a plethora of methods has emerged to quantify the relationship between structure and property to detect, if any, activity cliffs. Understanding the SAR of data sets and the early detection of activity cliffs can be essential to a medicinal and computational chemist to guide lead-optimization efforts and to develop predictive in-silico models.[2,3]

Activity landscape modeling is becoming an established approach to systematically analyze the SAR of data sets, especially containing several compounds.[4] A number of methods have been proposed and rich reviews describing these methods are published by Bajorath and colleagues[4−7] and Guha.[8] Indeed, Bajorath and his group have made outstating contributions to the field, pioneering a number of methods and concepts.[6,7] Guha and Van Drie, Agrafiotis et al., and Rarey et al., to name few groups, have also made very important contributions to the detection and interpretation of activity cliffs.[2,3,9,10] Efforts to predict activity cliffs are also emerging.[11,12]

Structure−activity similarity (SAS) maps were published by Maggiora and Shanmugasundaram in 2001 and are one of the early methods to navigate through the SAR of data sets and rapidly detect activity cliffs.[13] Since then, the SAS maps have been applied in several projects and adapted to study data sets from different types. Table 1 summarizes the SAS maps and related methods indicating the main focus of each approach, applications, and remarks that are elaborated in this manuscript. This review is focused on the application and development of the SAS maps to analyze SAR using the concept of activity landscape highlighting the advantages and limitations. The review is organized in six major sections. After the Introduction, section 2 describes the basic principles and interpretation of the SAS maps. Section 3 describes the concept of consensus activity cliffs and consensus SAS maps. Section 4 presents the extension of the maps to model multitarget activity landscapes and describes the dual-activity difference (DAD) maps and structure−multiple activity similarity (SmAS) maps. Section 4 also discusses the concept of selectivity switches. Section 5 presents the application of the SAS and SAS-like maps to describe property landscapes other than biological activity. Finally, section 6 contains conclusions and additional potential applications of the SAS maps.

Received: August 1, 2012

   dx.doi.org/10.1021/ci300362x | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

Table 1. SAS and SAS-like Maps and Their Applications to Model Activity Landscapes Discussed in This Review

map	property end points[a]	main features[b]	applications and remarks[b]	ref
structure−activity similarity (SAS)	one	plot of structure similarity vs activity similarity or potency difference (Figure 1A); identify activity cliffs, scaffold hops, and continuous regions in the SAR	SAR analysis of 48 bicyclic guanidines with κ-opioid receptor binding affinity; activity cliffs identified consistently by several representations are termed *consensus activity cliffs*	13, 27
			prototype analysis of structure−flavor relationships and systematic identification of *flavor cliffs*	55
binned SAS maps	one	categorical representations of SAS map; represents the distribution of pairs of compounds in each region as pie charts.	SAR analysis of 48 bicyclic guanidines with κ-opioid receptor binding affinity	27
dual-activity difference (DAD)	two	plot of activity difference for three targets (Figure 2); the distribution of the data points is independent of the molecular similarity; identifies dual-target cliffs with inverse and direct SAR and *selectivity switches*.	SAR analysis of 55 benzimidazole derivatives with biological activity against two parasites	40
			SAR analysis of 168 molecules tested with three PPAR subtypes	47
			SAR analysis of compounds sharing the same scaffold	48
triple-activity difference (TAD)	three	plot of activity difference for three targets. These maps can be interpreted as the combination of three DAD maps	SAR analysis of benchmark data sets used in activity landscape modeling	41
structure−multiple activity similarity (SmAS)	multiple	plot of structure similarity vs activity similarity measured across two or more biological end points; suitable to analyze bioactivity profile data.	SAR analysis of benchmark data sets with biological activity across 3 and 4 targets	49
			Analysis of bioassay data from PubChem	52
consensus methods	one or more	address the dependence of chemical space with structure representation and choice of descriptors; identify consensus regions in the landscape including *consensus activity cliffs*	consensus SAS map for the SAR analysis of 32 benzimidazoles with activity against two protozoan; deep and shallow activity cliffs are distinguished	23
			consensus SAS map combines the 3D molecular similarity of multiple conformers; analysis of the SAR of 54 compounds with activities against cathepsins	29
			consensus DAD map used for SAR analysis of benchmark data sets	41

[a]Several applications have been focused on molecular targets. However, the methods are applicable to any biological end point or property of interest. [b]See text for full discussion of each map.

dx.doi.org/10.1021/ci300362x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

2. SAS MAPS: PRINCIPLES AND INTERPRETATION

The SAS maps were proposed with the goal of finding a relationship between structure and activity and are highly related to the similarity principle: "similar structures have similar activity".[13,14] SAS maps are based on a systematic pairwise comparison of all the compounds in the data set and the structure similarity or potency difference for each pair of compounds is plotted against their structure similarity. The structure similarity can be computed with any structure representation, for example, two- or three-dimensional (2D/3D) fingerprints, descriptors obtained from various pharmacoporic tools recently reviewed,[15] descriptors obtained from quantum mechanical calculations, or the extensive set of descriptors compiled by Todeschini.[16] Also, any similarity metric can be used. Most of the studies performed so far have been conducted using the well-known Tanimoto similarity[17,18] but again, any other metric can be employed. The activity similarity can be computed with the equation:

$$S_{i,j} = 1 - \frac{|A_i - A_j|}{\max - \min} \quad (1)$$

where A_i and A_j are the activities of the ith and jth molecules, e.g., pIC_{50} or pK_i values, and max − min indicates the range of activities in the data set. A characteristic feature of the similarity values computed with eq 1 is that the values are in the same scale of the similarity values, namely zero to one. A disadvantage, however, is that the scaled values depend on the range of activities of the data set making difficult the direct comparison of SAS maps of data sets with different ranges of activity.

Alternatively, potency differences can be computed as the absolute difference between the activity values of each pair, for example, pIC_{50} values:

$$|\Delta pIC_{50}(T)_{i,j}| = pIC_{50}(T)_i - pIC_{50}(T)_j \quad (2)$$

where $pIC_{50}(T)_i$ and $pIC_{50}(T)_j$ are the activities of the ith and jth molecules ($j > i$) for target T.

Figure 1A presents a general form of a SAS map where the molecular similarity and potency differences are represented on the X- and Y-axes, respectively. The map can be roughly divided into four major regions. Pairs of compounds in region RI have low structural similarity and low potency difference (i.e., high activity similarity) and are indicative of scaffold or R-group hopping.[19,20] RII denotes pairs of compounds with high structure similarity and low potency difference and represents compounds with smooth or continuous SARs. Compounds in RIII have high structure similarity and high potency difference (i.e., low activity similarity) and, therefore, correspond to activity cliffs or discontinuous SARs. RIII identifies structural patterns that are critical for activity; therefore, it has been associated with high SAR information content.[21,22] In a subsequent development of the SAS maps, it was proposed that activity cliffs can be further classified as either shallow or deep.[23] Shallow activity cliffs are comprised of pairs of compounds displaying large (but not extreme) differences in activity as a result of slight changes in structure, and they occupy the borderline area between regions II and III. In contrast, a deep activity cliff is formed by pairs of compounds with very high structure similarities but remarkably low activity similarities. RIV is the least interesting to describe SARs because it contains pairs of molecules with low molecular similarity and high potency difference.

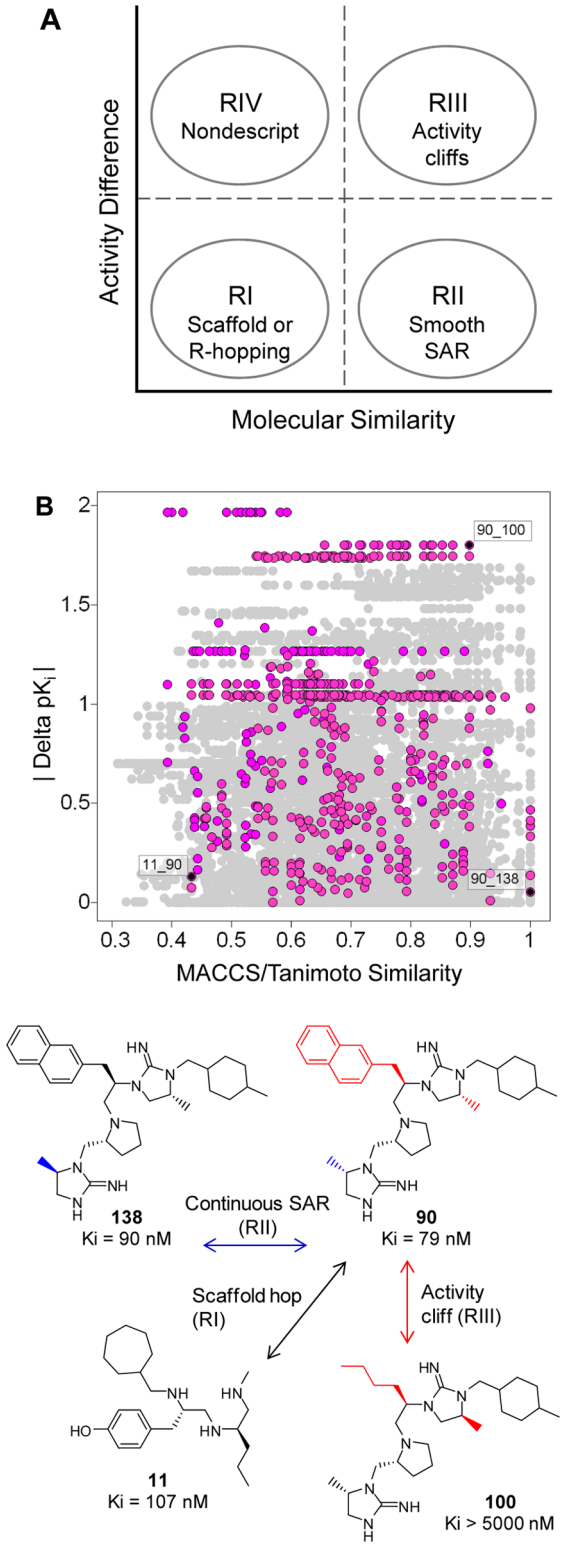


Figure 1. (A) Prototype map roughly divided in four major quadrants. Regions are as follows: RI and RII contain pairs of compounds that represent scaffold hopping and smooth SAR, respectively; RIII identifies activity cliffs. (B) Example of a map for a data set 160 compounds (12 720 data points total) tested in a μ-opioid receptor binding assay. The 785 points with at least one active compound ($K_i \leq$ 100 nM) in the pair are in magenta. The remaining pairs are displayed in light gray. Three selected pairs are marked in black; the chemical structures and activity data of the chosen compounds are also shown.

dx.doi.org/10.1021/ci300362x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Figure 1B shows an example of a SAS map for 160 compounds tested in a μ-opioid receptor binding assay.[24−26] The map has a total of 12 720 data points. The most relevant data points to interpret the SAR are the ones that contain at least one active compound in the pair defined as "active pairs".[27] Thus, 785 "active pairs" are colored in magenta. The remaining pairs are in light gray. In this example, a compound is considered "active" if $K_i \leq 100$ nM. Figure 1B also shows three selected pairs of compounds in different regions of the map illustrating an activity cliff, scaffold hop, and smooth region of the SAR. From the structures and activity values presented in the figure, it is clear that small structural changes in the activity cliff **90_100** (marked in red) are associated with a large change in potency difference (1.8 log units); for the pair of compounds **90_138**, that represent a smooth region in the SAR, the small structural change (marked in blue) does not dramatically change the biological activity; finally, the pair of compounds **11_90** have a very different chemical scaffold, although the potency is similar (0.13 log units) and exemplify a scaffold hop.

The qualitative interpretation and use of the SAS maps are straightforward as illustrated in the example above and the numerous examples found in the literature (Table 1). The quantitative characterization of the SAS maps, however, is challenging because of the need of imposing activity similarity (or potency difference) and molecular similarity thresholds along the *Y*- and *X*-axis, respectively (Figure 1A), and then counting the number of data pairs in each region. For example, the so-called "binned SAS maps" represent the distribution of the data points in each region of the map as pie charts.[27] Nevertheless, pairs of compounds near the borderlines of the quadrants can have different classifications depending on the criteria used to define the thresholds. There are no unique criteria to define the borderlines. The threshold for activity similarity or potency difference depends on the data set(s) and the specific goals of the study. Examples of thresholds used are an activity difference of one log unit;[27] two thresholds of one and two log units,[23] or the mean activity similarity of the data set.[29]

Defining thresholds for structure similarity has the issue of defining "high similarity". Statistical approaches have been proposed to address this point.[28] It has been argued that a pair of compounds is similar as long as it is intuitive or interpretable for a medicinal chemist. Although this is largely true, it is not easy to use thresholds to define high similarity for different structure representations (for example, pair of compounds that "look" similar to a medicinal chemist have, for example, Tanimoto similarity values of 0.4 and 0.9 for two different structure representations). In practical applications, various criteria have been used to set up a threshold along the *x*-axis of the SAS maps, for example; the median similarity of the most active compounds in the data set (using different definition of "active compound" depending on the goals of the project),[23,27] and the mean structure similarity of the data set depending on the structural fingerprint.[29] Of note, the challenge to quantitatively define "high structure similarity" applies to any other activity landscape method.

Since SAS maps depict all pairwise comparisons in the data set, the number of data points represented in a SAS or SAS-like map increases dramatically with the size of the database making difficult their interpretation. This is a general limitation of several activity landscape methods based on pairwise comparisons. However, one needs to keep in mind that these tools are designed to analyze and interpret only the most important pairs of compounds such as the activity cliffs and scaffold hops. Adding information for each molecular pair, for example, the activity of the most active compound in the pair, it is possible to rapidly filter out "inactive pairs" e.g., Figure 1B. A second, perhaps more general, approach for large data sets is to focus the pairwise comparison on selected subsets of molecules as recently reported by Yongye et al. to analyze a subset of compounds selected from a database of more than 15 000 compounds screened across 100 sequence-unrelated proteins.[30]

As discussed in the next sections, SAS maps and their variants have been applied to a number of data sets providing valuable information on different ongoing projects (Table 1). To name few examples, SAS or SAS-like maps have been used to describe the SAR of 48 bicyclic guanidines with κ-opioid receptor binding affinity;[27] 32 benzimidazoles with activity against two protozoan[23] and 54 compounds with activities against human cathepsin B, human cathepsin L, and *Trypanosoma brucei* cathepsin B.[29] Additional applications are described below.

3. MULTIPLE STRUCTURE REPRESENTATIONS: CONSENSUS MODELS

The dependence of chemical space on structure representations is well-recognized.[22,31,32] Therefore, the distribution of the pairs of compounds in an SAS map will depend on the specific fingerprint or any other set of descriptors used to represent the chemical structures. For example, the large effect of the type of 2D and 3D fingerprints used to compute structure similarity on the distribution of the pairs of compounds in an SAS map has been shown for a number of data sets.[23,27,29] In light of this dependence and to *reduce* the effect of the activity landscape with structure representation, it was proposed the analysis of the SAR of data sets using a number of SAS maps, each generated using a different 2D and 3D structure fingerprints, and then derive general conclusions common to several representations.[27,33] This approach is based on the fact that there are not perfect representations or descriptors that capture all possible relevant aspects of the molecule structures.[34] Thus, the goal is not to compare and select "the best" representation but to combine the information that provides each different type of representation. It can be argued that one representation is enough to provide interpretable results.[7] However, depending on the data set one representation can miss important information; for example, the largely used MACCS keys miss stereoisomers that are crucial for many biological systems. In contrast, computing similarities with multiple structure fingerprints is very fast and little additional computing time can provide significant robustness in the results. In fact, it has been shown that the use of 3D fingerprint representations is particularly convenient to study data sets that include stereoisomers. Of course, working with 3D conformations is not an easy task because of the limitations to use biologically relevant conformers.[35] To address this issue, one can compute 3D similarities calculated for multiple conformers.[29]

As part of the further development of the SAS maps to analyze SARs, it was proposed using not only fingerprint-based representations but also molecular properties, e.g., drug-like properties, to compute property similarity using scaled Euclidean distance.[23,29] The Euclidean distance (d_{ij}) between a pair of compounds *i* and *j* based on *k* properties are computed with the equation:[36]

dx.doi.org/10.1021/ci300362x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

$$d_{ij} = [\sum_{k=1}^{K} (p_{ki} - p_{kj})^2]^{1/2} \quad (3)$$

where p_{ki} and p_{kj} represent the value of property k. Then, the Euclidean distances can be scaled to values between 0 and 1 using the equation:[23]

$$sd_{ij} = \frac{d_{ij} - \min d_{ij}}{\max d_{ij} - \min d_{ij}} \quad (4)$$

where sd_{ij} is the scaled distance, while min d_{ij} and max d_{ij} specify the range of the computed distances. Finally, the pairwise property similarities (PS_{ij}) are determined from the scaled distances using the equation:[23]

$$PS_{ij} = 1 - sd_{ij} \quad (5)$$

Once several SAS maps have been generated for a data set using multiple structure representations, the next step is to extract general conclusions from the different representations. A general approach to obtain such general conclusions is based on the principles of data fusion[37−39] and is discussed in the next subsection.

3.1. Consensus Activity Cliffs and Consensus SAS Maps. As discussed above, one of the major practical applications of the SAS maps is the rapid identification of activity cliffs. As such, the concept of consensus activity cliffs was introduced to describe activity cliffs that are common to a series of representations or descriptors.[27] An apparent limitation of this approach is that it is possible that "true" activity cliffs are only captured by just few representations and missed by others.[7] However, it has been largely emphasized in the literature that consensus activity cliffs and, more generally, consensus pairs of compounds in other regions of the SAS map are not intended to eliminate data and disregard information that is captured by few molecular representations. Consensus activity cliffs and consensus models of the landscape are conceptualized to prioritize the SAR analysis of activity cliffs and other consistent regions in the activity landscape that are captured by several molecular representations.[29,33]

Two general strategies have been used to identify consensus activity cliffs. One approach is to identify pairs of compounds that are consistently are found in RIII of the prototype map in Figure 1A for a number of representations.[23,27] This method can be extended to identify consensus pairs of compounds in other regions of the SAS map, such as consensus scaffold hops. In order to compare the SAS map obtained from different structural similarity methods the number of data points consistently put into the same region for a given combination of representations can be computed with the degree of consensus (DoC). The DoC measure for each region R is a Tanimoto-like equation:[27]

$$\mathrm{DoC}^{\mathrm{R}}_{m,n} = \frac{\mathrm{Cp}_{m,n}}{p_m + p_n - \mathrm{Cp}_{m,n}} \quad (6)$$

where $\mathrm{Cp}_{m,n}$ is the number of *consensus pairs* in region R (R = I−IV in Figure 1A) between methods m and n; p_m is the number of pairs of molecules assigned by method m in region R; and p_n is the number of pairs of molecules assigned by method n in the same region. DoC represents the "similarity" that two methods exhibit for assigning molecules to a given region in a SAS map. Of note, DoC depends on the thresholds used to define regions I−IV. The DoC measure has been employed in several studies.[23,27,29]

The second general strategy to identify consensus activity cliffs or other consensus pairs is combining similarity measures obtained from different methods using data fusion. One example is computing the mean similarity i.e., "mean fusion" of various representations.[29,40] Other measures or fusion rules to combine pairwise similarities can be used. The aggregation of SAS maps obtained with different structure representations generates a "consensus SAS map".[23] Quantitative analysis of a consensus SAS maps have the challenge to establish a meaningful threshold for structure similarity in Figure 1, in other words, to define high/low similarity based on a combined similarity measure that provides interpretable and significant results of the SAR.[41]

The combination of structure representations has led to question of what set of descriptors should be selected to generate consensus SAS maps and identify consensus activity cliffs. A major criterion is to use fingerprints of different design that captures distinctive aspects of the molecular structures.[42] For example, dictionary-based fingerprints, pharmacophoric fingerprints, fingerprints based on the concept of atom pairs, or fingerprints that entail growing a set of fragments radially from each heavy atom over a series of iterations such as the extended connectivity fingerprints (ECFPs).[43] Of course, fingerprints that are available to the research group performing the analysis play an important role (it is worth mentioning that there are several structure fingerprints in the public domain such as the MayaChemTools[44]). A second, debatable criterion that has been used is to select fingerprints with "low" linear correlation between all pairwise similarity values for the particular data set under study. Here the drawback is to provide a statistical measure that defines low correlation. Currently, heuristic thresholds to define low correlation have been employed.[23,27,29] Approaches to select uncorrelated fingerprints are actively pursued.

4. FROM SINGLE- TO MULTITARGET LANDSCAPES

4.1. DAD Maps. The need to analyze the SAR for compounds tested across two biological end points gave rise to dual-activity difference (DAD) maps.[40] For a set of N molecules with biologic activity for targets **I** and **II**, the DAD map depicts $N(N-1)/2$ pairwise potency differences for each possible pair in the data set against both targets. The potency differences for target T for each molecule pair are calculated with the expression:

$$\Delta\mathrm{pIC}_{50}(T)_{i,j} = \mathrm{pIC}_{50}(T)_i - \mathrm{pIC}_{50}(T)_j \quad (7)$$

where $\mathrm{pIC}_{50}(T)_i$ and $\mathrm{pIC}_{50}(T)_j$ are the activities of the ith and jth molecules ($j > i$). Other activity values such as pK_i values can be used. A remarkable feature of eq 7 is that $\Delta\mathrm{pIC}_{50}(T)_{i,j}$ can have positive or negative values in contrast to the absolute potency difference in eq 2. Therefore, the DAD maps preserve the information of the *directionality* of the SAR and are able to differentiate pair of molecules where the structural change increases the activity for one target but decreases the activity for the other target (see below). Note that the eq 7 is different from eq 8 that measures selectivity of molecule m:[45]

$$S_m = \mathrm{pIC}_{50}(T1)_m - \mathrm{pIC}_{50}(T2)_m \quad (8)$$

dx.doi.org/10.1021/ci300362x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

where S_m is the selectivity of molecule m for targets $T1$ and $T2$, and $pIC_{50}(T1)_m$, $pIC_{50}(T2)_m$ are the potency values for molecule m against the two targets $T1$ and $T2$, respectively.

A general form of a DAD map is shown in Figure 2. Vertical and horizontal lines at $\Delta pIC_{50} \pm t$ determine boundaries for

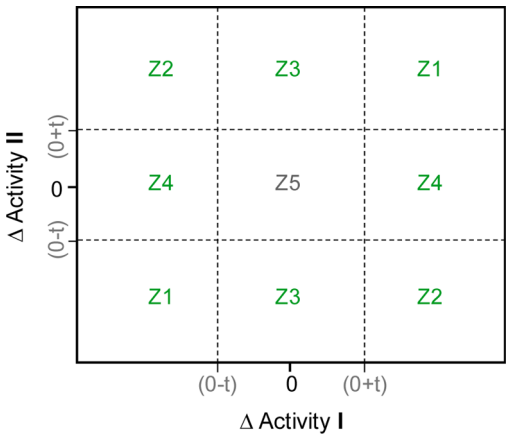


Figure 2. Prototype DAD map for two biological end points, **I** and **II**. The dashed lines intersect the axes at activity difference values of $0 \pm t$ generating five major regions: Z1, structural modifications result in significant a decrease or increase of activity toward both biological end points; Z2, changes in structure increase activity for one end point, while significantly decreasing activity for the other end point; Z3 and Z4, structural changes result in significant changes in activity toward one end point, but not a large change toward the other.

low/high potency difference for targets (or any other biological end point) **I** and **II**, respectively. Typically, $t = 1$ (one log unit) so that data points are considered with low potency difference if $-1 \leq \Delta pIC_{50} \leq 1$ for each target. However, potency difference values different than one, $t \neq 1$ can be used. In fact, it is possible to employ more than one threshold if one want to identify data pairs with multiple levels of activity difference such, for example, high and extremely high potency difference e.g., shallow and deep activity cliffs discussed above. When one threshold value is used, which has been the case for the applications developed so far, the boundaries defined with $\Delta pIC_{50} \pm t$ generate five zones, Z1−Z5 (Figure 2). Structural changes for molecule pairs that are in Z1 (either a small or a large structural change) have a similar impact on the activity against the two targets (either an increase or decrease in activity). Therefore, zone Z1 is associated with similar SAR of the pair of compounds for both targets. Data points that are in Z2 indicate that the change in activity for the compounds in the pair is opposite for **I** and **II**. Thus, the structural changes in the pair of compounds in Z2 are associated with an inverse SAR, i.e., increases the activity for one target but decreases the activity for the other target. Data points that falls into Z3 and Z4 correspond to pairs of molecules with the same or similar activity for one target (**I** or **II**, respectively), but different activity for the other target (**II** or **I**, respectively). Data points that fall into Z5 indicate pair of compounds with identical or similar activity against **I** and **II** and, therefore, have little or no impact on the activity against the two targets. Compounds in Z5 might be of significant interest if the goal if the analysis is to identify promiscuous hits. Of note, the classification of data points in an activity-difference map is independent of the structure similarity.

The prototype DAD map in Figure 2 does not contain information of how similar or dissimilar are the molecule pairs. However, it is straightforward to add such information. At least, two general methods can be proposed that mostly depend on the type of data set to be analyzed:

- *Annotate the data points with pairwise structure or property similarity*. This approach is most appropriate for data sets with diverse molecules although it has been applied to analogues series. Here, the structure similarity for each molecule pair can be computed using any molecular representation and similarity method. For several data sets, 2D, 3D fingerprints, and drug-like properties have been used, and the similarity has been measured by the Tanimoto coefficient. Moreover, similar to the consensus SAS maps discussed above, multiple-structure representations can be used and merged to generate a consensus representation.[41]
- *Distinguish the data points with the number of substitutions*. This method is intended to analyze combinatorial data sets or analogues series. Data points in the DAD map are further differentiated by the number of different substitutions around a common core scaffold. Perhaps the most interpretable SAR occurs for data points with one or two substitutions.

Adding structure similarity information for diverse sets or the number of substitutions for analogues series easily identifies in the DAD maps single- or dual-target activity cliffs and R-cliffs.[10] R-cliffs are pair of compounds where a single or double substitution around the core scaffold dramatically modifies the biological activity. Similarly, structure similarity or number of substitutions information represented in the DAD maps easily identifies scaffold-hoops or R-hops.[20,46] For example, the pairs of compounds in a DAD map (Figure 2) with high structure similarity (or one or two substitutions for analogues series) in Z3 and Z4 correspond to single-target activity cliffs for target **I** or **II**, respectively. Pairs of compounds with high structure similarity in Z1 indicate dual-target activity cliffs with similar SAR (small changes in structure are associated with a high change in activity for both targets, either increase or decrease). In sharp contrast, pairs of similar compounds in Z2 indicate dual-target activity cliffs with inverse SAR or "selectivity switches", i.e., small changes in structure are associated with a high increase in the activity for one target but high decrease in activity for the second target.[41] In other words, selectivity switches point to structural changes (such as one or two substitutions in the case of combinatorial data sets) that completely invert the selectivity pattern.

As discussed in section 2, since DAD maps depict all pairwise potency differences for each pair in the data set, the number of data points represented in the map increases dramatically with the size of the database making difficult the interpretation for large data sets. However, similar to the SAS maps, adding information for each molecular pair with the structure similarity or number of substitutions, it is possible to easily filter out data points and focus the analysis and interpretation on the most interesting cases.

Adding a third dimension to DAD maps that represents the activity difference for a target **III** gives rise to triple activity-difference (TAD) maps.[41] TAD maps can be interpreted as a group of three DAD maps for each combination of two targets, i.e., **I-II**, **I-III**, and **II-III**.[41]

dx.doi.org/10.1021/ci300362x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

DAD maps have been used to describe the SAR of 55 benzimidazole derivatives with biological activity against two parasites, *T. vaginalis* and *G. intestinalis*,[40] 299 compounds screened against dopamine, norepinephrine, and serotonin transporters,[41] and 168 molecules tested with three proliferator-activated receptor (PPAR) subtypes.[47] DAD maps have also been adapted to scan the SAR of compounds with active and selective chemotypes with activity against cyclooxygenase-1 and -2[48] (Table 1).

4.2. SmAS Maps. As discussed above, DAD and TAD maps offer an intuitive manner to explore the SAR of data sets with activity against two or three targets, respectively. For data sets with activity across more than three biological end points, the SAS maps were adapted by computing multiple-activity similarity (mAS).[49] The pairwise activity similarities across p targets, i.e., mAS, can be calculated using the Tanimoto coefficient for real-valued vectors:[22]

$$\mathrm{mAS}(i,j) = \frac{\sum_{k=1}^{p} v_{i,k} \cdot v_{j,k}}{\sum_{k=1}^{p} v_{i,k}^{\,2} + \sum_{k=1}^{p} v_{j,k}^{\,2} - \sum_{k=1}^{p} v_{i,k} \cdot v_{j,k}} \quad (9)$$

where mAS(i,j) is the multiple activity similarity of the ith and jth molecules, p is the number of targets, and $v_{i,k}$ and $v_{j,k}$ denote the value of the potency of the kth target for the ith and jth molecules, respectively. Thus, mAS capture with a single measure the similarity of the bioactivity profile of each pair of molecules across all p targets, and it is is similar to the "binding profile similarity" measure reported by Steffen et al.[50] Of note, the mAS measure does not provide information for the change in activity for each target. Plotting mAS against molecular similarity gives rise to the SmAS maps that are an extension of the SAS maps for multiple targets,[49] and their interpretation is essentially the same. For example, multitarget activity cliffs can be rapidly located in RIII of a SmAS map (Figure 1). By analogy with the discussion of the SAS maps, any structure representation and similarity measure can be used. Similarity, "consensus SmAS maps" can be produced by combining multiple representations.[49] Clearly, the mAS measure in eq 9 can be applied to other approaches that model activity landscapes such as the structure multiple-activity landcape index (SmALI)[49] that is a natural extension to the SALI metric developed by Van Drie and Guha for single targets.[2,3]

4.3. Bioassay Landscape Modeling. Several drug and probe discovery efforts involve the profiling of sets of compound across multiple assays. A prominent example is found in PubChem,[51] a public database which contains activity data for compounds tested across a number of bioassays. SmAS maps have been adapted to navigate through the SAR of sets of compounds with activity across different bioassays. Since the bioassays reported in PubChem may have distinctive nature and the assays can be performed by different screening centers, each bioassay has its own quantitative definition of "active", "inactive", and "inconclusive". To address this feature, a categorical classification of the activity data was used[52] and the bioassay activity profile was represented as a multiset fingerprint encoding of the activity data as follows: active is set to "2"; inactive as "1"; inconclusive or not tested as "0". Then, the pairwise bioassay activity-profile similarity (bAPS) across the n bioassays was calculated using the Tanimoto coefficient:

$$\mathrm{bAPS}(i,j) = \frac{\sum_{k-1}^{n} \min[m_k(i), m_k(j)]}{\sum_{k=1}^{n} \max[m_k(i), m_k(j)]} \quad (10)$$

where bAPS(i,j) is the bioassay activity-profile similarity of the ith and jth molecules, $m_k(i)$ and $m_k(j)$ are the activity encodings of the ith and jth molecules, respectively, and n is the total number of assays that the molecules were screened across. bAPS is plotted against a measure of structure similarity that leads to the rapid identification of "bioassay activity profile cliffs". Similar to the SAS and SmAS maps, any similarity measure, structure representation or combination of structure representations can be used.[52] The bAPS measure has been employed to describe the bioassay activity landscape of 618 compounds obtained from an in-house collection and deposited in PubChem.[52] The approach can be extended to analyze any other subset of compounds included in this or other large public database annotated with biological activity.[53,54]

5. BEYOND BIOLOGICAL ACTVIVTY: MODELING OTHER PROPERTY LANDSCAPES

In principle, scanning the SAR of data sets using SAS and SAS-like maps can be extended to analyze any other property landscape by replacing the activity similarity (or potency difference) measures discussed above e.g., eqs 1, 2, 7, 9, and 10, with the property of interest. As part of a study to explore the relationship between chemical structure and flavor of a large data set, the description of the flavor for each compound was coded into a binary flavor fingerprint. On the basis of the fingerprint-based representation of the "flavor space", the flavor similarity was computed using the Tanimoto coefficient. Plotting flavor similarity against structure similarity generated a structure−flavor similarity map.[55] Using these maps it is possible to systematically and easily identify "flavor cliffs" or "odor cliffs".[56] The latter applies if odorants (fragrance chemicals) are studied. Indeed, there is an increased interest to integrate the methods commonly used in chemical informatics with food chemistry as exemplified by "FoodInformatics" that integrates chemical information with food chemistry.[57]

Similar to the analysis of structure−flavor associations using SAS-like maps, the analysis of other properties can be envisioned. For example, the analysis of protein−ligand interactions is in current development using protein−ligand interaction fingerprints (PLIF).[58−60] The application of activity landscape methods to analyze PLIF has been proposed[9] using the SALI metric (see above). Therefore, it is quite feasible to analyze the relationships between PLIF and structure similarity or other associations, such as docking score similarities, using SAS-like maps described above. Likewise, properties related to solve toxicological and ecotoxicological problems[61−63] can be analyzed using SAS and related maps.

6. CONCLUSIONS

SAS maps are powerful chemoinformatics tools that have been used to systematically analyze structure−activity and structure−property relationships of numerous data sets. SAS and SAS-like maps allow the SAR analysis of structural diverse and combinatorial data sets or analogues series; sets of molecules tested with a single target or tested across multiple biological end points. SAS and related maps provide not only a visual and qualitative analysis of the data but also a quantitative assessment of the property landscape. In particular, these maps represent an intuitive way to identify the most pronounced examples of activity cliffs that point to specific

dx.doi.org/10.1021/ci300362x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

structural changes crucial for biological activity. Similarly, the identification of the most representative scaffold hops and selectivity switches is straightforward. Most of the current applications of the SAS and SAS-like maps have been focused on modeling biological activity. Applications are emerging to model other properties, such as flavor landscapes. SAS and associated maps are very versatile; any similarity metric, structure representation (fingerprint- or property-based), or combination of representations can be used. The combination of structure representations has given rise to the general concept of consensus activity cliffs. One of the challenges to quantitatively characterize the SAS and related maps, which is a limitation of many other activity landscape methods, is the definition of "high similarity". For practical applications, different criterion can be used that largely depend on the particular data set, the structure representation used, and the goals of the project. Other limitations of the SAS and SAS-like maps is the analysis of large data sets. Since all possible pairwise comparisons are represented, the SAS maps eventually get crowded so that the interpretation is difficult. This limitation is not only for SAS maps but for several other methods that visually represent the SAR of data sets. A workaround of this limitation is to focus the analysis of activity cliffs on subsets of compounds selected from very large data sets. Also, one must consider that not all compound pairs are interpreted in a SAS maps, but those compound pairs that have the highest information content such as the activity cliffs. It is expected that the SAS and associated maps reviewed here will continue to evolve and assist in navigating through the structure−property relationships.

AUTHOR INFORMATION

Corresponding Author

*Tel.: +1-772-345-4685. Fax: +1-772-345-3649. E-mail: jmedina@tpims.org.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The most sincere thanks to Prof. Gerald M. Maggiora for motivating and inspiring the work reviewed here and invaluable discussions. We also acknowledge the extensive work and helpful insights with all the colleagues and students whose names are listed in the references cited in this paper related to the SAS and SAS-like maps. We thank Jacob Waddell for writing the scripts to generate the examples presented in this review. This work was supported by the State of Florida, Executive Office of the Governor's Office of Tourism, Trade, and Economic Development and the Multiple Sclerosis National Research Institute.

ABBREVIATIONS

2D, two-dimensional; 3D, three-dimensional; bAPS, bioactivity activity-profile similarity; DAD, dual-activity difference; PLIF, protein−ligand interaction fingerprints; SAR, structure−activity relationships; SAS, structure−activity similarity; SmAS, structure−multiple activity similarity; SPA, structure−property activity; SPR, structure−property relationships; TAD, triple-activity difference

REFERENCES

(1) Maggiora, G. M. On Outliers and Activity Cliffs-Why QSAR Often Disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535−1535.

(2) Guha, R.; Van Drie, J. H. Assessing How Well a Modeling Protocol Captures a Structure-Activity Landscape. *J. Chem. Inf. Model.* **2008**, *48*, 1716−1728.

(3) Guha, R.; VanDrie, J. H. Structure-Activity Landscape Index: Identifying and Quantifying Activity Cliffs. *J. Chem. Inf. Model.* **2008**, *48*, 646−658.

(4) Wassermann, A. M.; Wawer, M.; Bajorath, J. Activity Landscape Representations for Structure-Activity Relationship Analysis. *J. Med. Chem.* **2010**, *53*, 8209−8223.

(5) Bajorath, J.; Peltason, L.; Wawer, M.; Guha, R.; Lajiness, M. S.; Van Drie, J. H. Navigating Structure-Activity Landscapes. *Drug Discovery Today* **2009**, *14*, 698−705.

(6) Bajorath, J. Modeling of Activity Landscapes for Drug Discovery. *Expert Opin. Drug Discovery* **2012**, *7*, 463−473.

(7) Stumpfe, D.; Bajorath, J. Exploring Activity Cliffs in Medicinal Chemistry. *J. Med. Chem.* **2012**, *55*, 2932−2942.

(8) Guha, R. Exploring Structure−Activity Data Using the Landscape Paradigm. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, in press. doi: 10.1002/wcms.1087.

(9) Seebeck, B.; Wagener, M.; Rarey, M. From Activity Cliffs to Target-Specific Scoring Models and Pharmacophore Hypotheses. *ChemMedChem* **2011**, *6*, 1630−1639.

(10) Agrafiotis, D. K.; Wiener, J. J. M.; Skalkin, A.; Kolpak, J. Single R-Group Polymorphisms (SRPs) and R-Cliffs: An Intuitive Framework for Analyzing and Visualizing Activity Cliffs in a Single Analog Series. *J. Chem. Inf. Model.* **2011**, *51*, 1122−1131.

(11) Guha, R. Exploring Uncharted Territories: Predicting Activity Cliffs in Structure−Activity Landscapes. *J. Chem. Inf. Model.* **2012**, *52*, 2181−2191.

(12) Heikamp, K.; Hu, X.; Yan, A.; Bajorath, J. Prediction of Activity Cliffs Using Support Vector Machines. *J. Chem. Inf. Model.* **2012**, *52*, 2354−2365.

(13) Shanmugasundaram, V.; Maggiora, G. M. Characterizing Property and Activity Landscapes Using an Information-Theoretic Approach. CINF-032. In *222nd ACS National Meeting*, Chicago, IL, United States; August 26−30, 2001; American Chemical Society: Washington, D. C., 2001.

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

(15) Sanders, M. P. A.; Barbosa, A. J. M.; Zarzycka, B.; Nicolaes, G. A. F.; Klomp, J. P. G.; de Vlieg, J.; Del Rio, A. Comparative Analysis of Pharmacophore Screening Tools. *J. Chem. Inf. Model.* **2012**, *52*, 1607−1620.

(16) Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*; Wiley-VCH: Weinheim, 2009.

(17) Jaccard, P. Etude Comparative De La Distribution Florale Dans Une Portion Des Alpes Et Des Jura. *Bull. Soc. Vaudoise Sci. Nat.* **1901**, *37*, 547−579.

(18) Bender, A.; Glen, R. C. Molecular Similarity: A Key Technique in Molecular Informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204−3218.

(19) Brown, N.; Jacoby, E. On Scaffolds and Hopping in Medicinal Chemistry. *Mini-Rev. Med. Chem.* **2006**, *6*, 1217−1229.

(20) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. Scaffold-Hopping by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894−2896.

(21) Iyer, P.; Wawer, M.; Bajorath, J. Comparison of Two- and Three-Dimensional Activity Landscape Representations for Different Compound Data Sets. *Med. Chem. Comm.* **2011**, *2*, 113−118.

(22) Maggiora, G. M.; Shanmugasundaram, V. Molecular Similarity Measures. In *Chemoinformatics and Computational Chemical Biology, Methods in Molecular Biology*; Bajorath, J., Ed.; Springer: New York, 2011; Vol. *672*, pp 39−100.

(23) Pérez-Villanueva, J.; Santos, R.; Hernández-Campos, A.; Giulianotti, M. A.; Castillo, R.; Medina-Franco, J. L. Towards a Systematic Characterization of the Antiprotozoal Activity Landscape of Benzimidazole Derivatives. *Biorg. Med. Chem.* **2010**, *18*, 7380−7391.

(24) Houghten, R. A.; Pinilla, C.; Appel, J. R.; Blondelle, S. E.; Dooley, C. T.; Eichler, J.; Nefzi, A.; Ostresh, J. M. Mixture-Based Synthetic Combinatorial Libraries. *J. Med. Chem.* **1999**, *42*, 3743−3778.
(25) Houghten, R. A.; Pinilla, C.; Giulianotti, M. A.; Appel, J. R.; Dooley, C. T.; Nefzi, A.; Ostresh, J. M.; Yu, Y. P.; Maggiora, G. M.; Medina-Franco, J. L.; Brunner, D.; Schneider, J. Strategies for the Use of Mixture-Based Synthetic Combinatorial Libraries: Scaffold Ranking, Direct Testing, in Vivo, and Enhanced Deconvolution by Computational Methods. *J. Comb. Chem.* **2008**, *10*, 3−19.
(26) Yongye, A. B.; Appel, J. R.; Giulianotti, M. A.; Dooley, C. T.; Medina-Franco, J. L.; Nefzi, A.; Houghten, R. A.; Martínez-Mayorga, K. Identification, Structure−Activity Relationships and Molecular Modeling of Potent Triamine and Piperazine Opioid Ligands. *Bioorg. Med. Chem.* **2009**, *17*, 5583−5597.
(27) Medina-Franco, J. L.; Martínez-Mayorga, K.; Bender, A.; Marín, R. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. Characterization of Activity Landscapes Using 2D and 3D Similarity Methods: Consensus Activity Cliffs. *J. Chem. Inf. Model.* **2009**, *49*, 477−491.
(28) Baldi, P.; Nasr, R. When Is Chemical Similarity Significant? The Statistical Distribution of Chemical Similarity Scores and Its Extreme Values. *J. Chem. Inf. Model.* **2010**, *50*, 1205−1222.
(29) Yongye, A.; Byler, K.; Santos, R.; Martínez-Mayorga, K.; Maggiora, G. M.; Medina-Franco, J. L. Consensus Models of Activity Landscapes with Multiple Chemical, Conformer and Property Representations. *J. Chem. Inf. Model.* **2011**, *51*, 1259−1270.
(30) Yongye, A. B.; Medina-Franco, J. L. Data Mining of Protein-Binding Profiling Data Identifies Structural Modifications That Distinguish Selective and Promiscuous Compounds. *J. Chem Inf. Model.* **2012**, *52*, 2454−2461.
(31) Sheridan, R. P.; Kearsley, S. K. Why Do We Need So Many Chemical Similarity Search Methods? *Drug Discovery Today* **2002**, *7*, 903−911.
(32) Medina-Franco, J. L.; Martínez-Mayorga, K.; Giulianotti, M. A.; Houghten, R. A.; Pinilla, C. Visualization of the Chemical Space in Drug Discovery. *Curr. Comput.-Aided Drug Des.* **2008**, *4*, 322−333.
(33) Medina-Franco, J. L.; Yongye, A. B.; López-Vallejo, F. Consensus Models of Activity Landscapes. In *Statistical Modeling of Molecular Descriptors in QSAR/QSPR*; Matthias, D., Kurt, V., Danail, B., Eds.; Wiley-VCH: Weinheim, 2012; pp 307−326.
(34) Bender, A. How Similar Are Those Molecules after All? Use Two Descriptors and You Will Have Three Different Answers. *Expert Opin. Drug Discovery* **2010**, *5*, 1141−1151.
(35) Yongye, A. B.; Bender, A.; Martinez-Mayorga, K. Dynamic Clustering Threshold Reduces Conformer Ensemble Size While Maintaining a Biologically Relevant Ensemble. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 675−686.
(36) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.
(37) Willett, P. Similarity-Based Virtual Screening Using 2D Fingerprints. *Drug Discovery Today* **2006**, *11*, 1046−1053.
(38) Chen, B.; Mueller, C.; Willett, P. Combination Rules for Group Fusion in Similarity-Based Virtual Screening. *Mol. Inf.* **2010**, *29*, 533−541.
(39) Medina-Franco, J. L.; Maggiora, G. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. A Similarity-Based Data-Fusion Approach to the Visual Characterization and Comparison of Compound Databases. *Chem. Biol. Drug Des.* **2007**, *70*, 393−412.
(40) Pérez-Villanueva, J.; Santos, R.; Hernández-Campos, A.; Giulianotti, M. A.; Castillo, R.; Medina-Franco, J. L. Structure-Activity Relationships of Benzimidazole Derivatives as Antiparasitic Agents: Dual Activity-Difference (DAD) Maps. *Med. Chem. Commun.* **2011**, *2*, 44−49.
(41) Medina-Franco, J. L.; Yongye, A. B.; Pérez-Villanueva, J.; Houghten, R. A.; Martínez-Mayorga, K. Multitarget Structure-Activity Relationships Characterized by Activity-Difference Maps and Consensus Similarity Measure. *J. Chem. Inf. Model.* **2011**, *51*, 2427−2439.
(42) 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.
(43) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(44) Sud, M. MayaChemTools. http://www.Mayachemtools.org (accessed August, 2012).
(45) Peltason, L.; Hu, Y.; Bajorath, J. From Structure-Activity to Structure-Selectivity Relationships: Quantitative Assessment, Selectivity Cliffs, and Key Compounds. *ChemMedChem* **2009**, *4*, 1864−1873.
(46) Sun, H. M.; Tawa, G.; Wallqvist, A. Classification of Scaffold-Hopping Approaches. *Drug Discovery Today* **2012**, *17*, 310−324.
(47) Méndez-Lucio, O.; Pérez-Villanueva, J.; Castillo, R.; Medina-Franco, J. L. Activity Landscape Modeling of PPAR Ligands with Dual-Activity Difference Maps. *Bioorg. Med. Chem.* **2012**, *20*, 3523−3532.
(48) Pérez-Villanueva, J.; Medina-Franco, J. L.; Méndez-Lucio, O.; Yoo, J.; Soria-Arteche, O.; Izquierdo, T.; Lozada, M. C.; Castillo, R. CASE Plots for the Chemotype Based Activity and Selectivity Analysis: A CASE Study of Cyclooxygenase Inhibitors. *Chem. Biol. Drug Des.* **2012**, in press. DOI: 10.1111/cbdd.12019.
(49) Waddell, J.; Medina-Franco, J. L. Bioactivity Landscape Modeling: Chemoinformatic Characterization of Structure-Activity Relationships of Compounds Tested across Multiple Targets. *Bioorg. Med. Chem.* **2012**, *20*, 5443−5452.
(50) Steffen, A.; Kogej, T.; Tyrchan, C.; Engkvist, O. Comparison of Molecular Fingerprint Methods on the Basis of Biological Profile Data. *J. Chem. Inf. Model.* **2009**, *49*, 338−347.
(51) Wang, Y. L.; Xiao, J. W.; Suzek, T. O.; Zhang, J.; Wang, J. Y.; Zhou, Z. G.; Han, L. Y.; Karapetyan, K.; Dracheva, S.; Shoemaker, B. A.; Bolton, E.; Gindulyte, A.; Bryant, S. H. Pubchem's BioAssay Database. *Nucleic Acids Res.* **2012**, *40*, D400−D412.
(52) Medina-Franco, J. L.; Waddell, J. Towards the Bioassay Activity Landscape Modeling in Compound Databases. *J. Mex. Chem. Soc.* **2012**, *56*, 163−168.
(53) Bender, A. Databases Compound Bioactivities Go Public. *Nat. Chem. Biol.* **2010**, *6*, 309−309.
(54) Barbosa, A. J. M.; Rio, A. D. Freely Accessible Databases of Commercial Compounds for High- Throughput Virtual Screenings. *Curr. Top. Med. Chem.* **2012**, *12*, 866−877.
(55) Martínez-Mayorga, K.; Peppard, T. L.; Yongye, A. B.; Santos, R.; Giulianotti, M.; Medina-Franco, J. L. Characterization of a Comprehensive Flavor Database. *J. Chemom.* **2011**, *25*, 550−560.
(56) Martínez-Mayorga, K.; Medina-Franco, J. L. Chemoinformatics—Applications in Food Chemistry. In *Advances in Food and Nutrition Research*; Taylor, S., Ed.; Academic Press: Burlington, 2009; Vol. *58*, pp 33−56.
(57) Martínez-Mayorga, K.; Medina-Franco, J. L. Organizers. FoodInformatics: Applications of Chemical Information to Food Chemistry. Division of Chemical Information. In *245th ACS National Meeting*, New Orleans, LA, United States; April 7−11, 2013; American Chemical Society: Washington, D. C., April 7−11, 2013.
(58) Deng, Z.; Chuaqui, C.; Singh, J. Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein-Ligand Binding Interactions. *J. Med. Chem.* **2004**, *47*, 337−344.
(59) Pérez-Nueno, V. I.; Rabal, O.; Borrell, J. I.; Teixido, J. APIF: A New Interaction Fingerprint Based on Atom Pairs and Its Application to Virtual Screening. *J. Chem. Inf. Model.* **2009**, *49*, 1245−1260.
(60) Schreyer, A.; Blundell, T. CREDO: A Protein-Ligand Interaction Database for Drug Discovery. *Chem. Biol. Drug Des.* **2009**, *73*, 157−167.
(61) Cronin, M. T. D.; Worth, A. P. (Q)SARs for Predicting Effects Relating to Reproductive Toxicity. *QSAR Comb. Sci.* **2008**, *27*, 91−100.
(62) Michielan, L.; Pireddu, L.; Floris, M.; Moro, S. Support Vector Machine (SVM) as Alternative Tool to Assign Acute Aquatic Toxicity Warning Labels to Chemicals. *Mol. Inf.* **2010**, *29*, 51−64.
(63) Fayet, G.; Del Rio, A.; Rotureau, P.; Joubert, L.; Adamo, C. Predicting the Thermal Stability of Nitroaromatic Compounds Using Chemoinformatic Tools. *Mol. Inf.* **2011**, *30*, 623−634.