JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

Multitarget Structure−Activity Relationships Characterized by Activity-Difference Maps and Consensus Similarity Measure

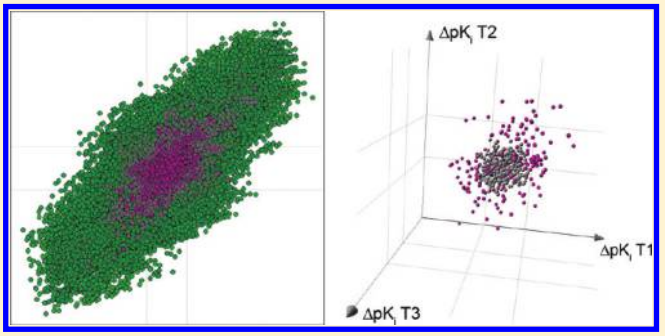
José L. Medina-Franco,*[,†] Austin B. Yongye,[†] Jaime Pérez-Villanueva,[‡] Richard A. Houghten,[†] and Karina Martínez-Mayorga[†]

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

[‡]Departamento de Sistemas Biológicos, División de Ciencias Biológicas y de la Salud, UAM-X, México DF 04960, México

Supporting Information

ABSTRACT: Dual and triple activity-difference (DAD/TAD) maps are tools for the systematic characterization of structure−activity relationships (SAR) of compound data sets screened against two or three targets. DAD and TAD maps are two- and three- dimensional representations of the pairwise activity differences of compound data sets, respectively. Adding pairwise structural similarity information into these maps readily reveals activity cliff regions in the SAR for one, two, or three targets. In addition, pairs of compounds in the smooth regions of the SAR and scaffold hops are also easily identified in these maps. Herein, DAD and TAD maps are employed for the systematic characterization of the SAR of a benchmark set of 299 compounds screened against dopamine, norepinephrine, and serotonin transporters. To reduce the well-known dependence of the activity landscape on the structural representation, five selected 2D and 3D structure representations were used to characterize the SAR. Systematic analysis of the DAD and TAD maps reveals regions in the landscape with similar SAR for two or the three targets as well as regions with *inverse SAR*, i.e., changes in structure that increase activity for one target, but decrease activity for the other target. Focusing the analysis on pairs of compounds with high structure similarity revealed the presence of single-, dual-, and triple-target activity cliffs, i.e., small changes in structure with high changes in potency for one, two, or the three targets, respectively. Triple-target scaffold hops are also discussed. Activity cliffs and scaffold hops were also quantified and represented using two recently proposed approaches namely, mean Structure Activity Landscape Index (mean SALI) and *Consensus* Structure−Activity Similarity (SAS) maps.



INTRODUCTION

The analysis of structure−activity relationships (SAR) of a set of compounds with measured biological activity is a central topic in drug design.[1] SAR studies of small-to-medium size data sets are frequently performed by medicinal chemists on a routine basis with no need for computational approaches. However, the systematic study of the SAR of large data sets requires the application of automated methods. Computational approaches applied to SAR analysis can be divided into predictive and descriptive methods.[2] Predictive methods are meant to anticipate the biological activity of molecules before synthesis and/or biological evaluation in order to focus the experimental efforts on the most promising compounds. Predictive methods include, but are not limited to, quantitative structure−activity relationships (QSAR), pharmacophore modeling, and machine learning approaches.[3−6] In contrast, descriptive approaches are designed to access, visualize, and to help understand the data obtained from biological screening or chemical optimization.[2] *Activity landscape* methods, as they currently stand, are descriptive approaches that are increasingly being used to systematically characterize the SAR of large data sets with activity against one or more targets.[1,2,7,8] An activity landscape can be conceptualized as a chemical space with the addition of biological activity as another dimension.[7] Several methodologies like traditional QSAR make assumptions that do not necessarily hold true and, thus, may present misleading results including nonpredictive models.[9] For example, one common assumption is that a lead series of compounds has a common binding mode or mechanism of action. For this reason, it has been largely recognized that understanding the activity landscape and the early detection of activity cliffs (chemical compounds with highly similar structures but significantly different biological activities)[10] can be crucial to the success of predictive computational models.[11,12] Indeed, the recent editorial by Maggiora[10] highlights the implications of the presence of activity cliffs in QSAR modeling. For example, *apparent* outliers in the data may reflect the presence of activity cliffs and may not be due to statistical fluctuations or to

Received: June 17, 2011
Published: August 15, 2011

measurement errors.[10] Multitarget activity landscape modeling is particularly attractive to characterize the SAR of compound data sets associated with selectivity[13] or promiscuity, e.g., 'scaffold-based promiscuity' of the targets.[14]

Different approaches are emerging to characterize systematically the SAR of compound data sets with one biological end point, e.g., single-target activity landscapes, and are reviewed in Bajorath et al.[7] These include the Structure−Activity Relationship Index (SARI),[15] Structure-Landscape Index (SALI),[11,12] Structure−Activity Similarity (SAS) maps,[16] and network-like similarity graphs (NSG).[17] Some of these approaches have been recently adapted to explore the SAR of data sets with biological activity across different biological end points, e.g., multitarget activity landscapes.[13,18−20] Of note, an important aspect in single- or multitarget activity landscape modeling is the strong influence of structure representation on chemical space.[21,22] Since different structure descriptors capture different aspects of molecular structures,[23,24] each set of descriptors can give a different landscape. Efforts to address this issue in single-target activity landscape modeling include using multiple structural representations to derive a consensus model for the activity landscape and identify consensus activity cliffs.[25,26] Consensus models are designed to prioritize the SAR analysis of activity cliffs and other consistent regions in the activity landscape that are captured by several structure representations. As pointed out previously, consensus data points are not meant to eliminate data, disregarding for example, "true" activity cliffs that are not identified by some structure representations.[26]

We recently introduced two-dimensional (2D) activity-difference (DAD) maps as a data analysis approach to characterize the SAR of a data set of 55 benzimidazole derivatives screened against two parasites.[27] In a 2D plot, DAD maps depict the pairwise activity differences for all compounds in the data set and can be considered as a pairwise landscape representation.[20] Since DAD maps represent pairwise relationships, it is easy to add structure similarity information and, consequently, identify dual-target scaffold hops (chemical compounds with highly dissimilar structures but significantly similar biological activities),[28] single-target activity cliffs and dual-target activity cliffs (pairs of compounds with highly similar structures but significantly different biological activities for only one or the two targets, respectively).

Expanding on our previous work, the pairwise SAR of a diverse set of 299 compounds tested for inhibition against three monoamine transporters is explored in this work.[20] The multitarget activity landscape was characterized using activity-difference maps annotated with an aggregated similarity value. To this end, 2D and 3D fingerprint similarities were combined in a single similarity measure by mean fusion.[25,26] Consensus single-, dual-, and triple-target activity cliffs as well as scaffold hops were identified. Activity cliffs and scaffold hops were also analyzed using two recently proposed approaches namely, mean Structure Activity Landscape Index (mean SALI)[11,12,25] and *Consensus* Structure−Activity Similarity (SAS) maps.[19,26]

The approaches described here are not restricted to the data set employed, the structure representations, similarity functions, or molecular targets. The emphasis in this work is on the activity difference maps for the comprehensive modeling of multitarget activity landscape of compounds tested across different targets using multiple structure representations.

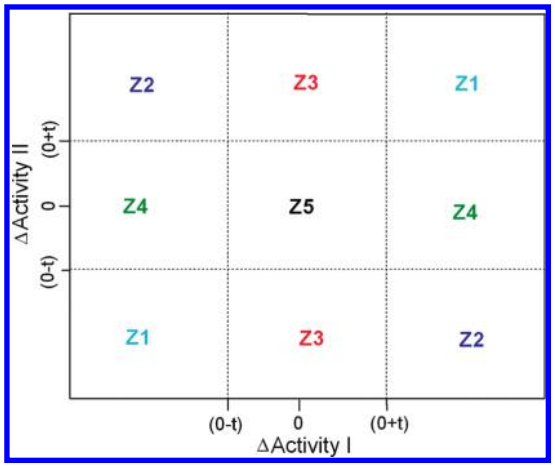


Figure 1. Prototype dual activity-difference (DAD) map for targets **I** and **II**. The dashed lines intersect the axes at potency difference values of $0 \pm t$ e.g., t = 1 (one log unit). The regions are as follows: Z1, structural modifications result in a significant decrease or increase of activity toward both targets; Z2, changes in structure increase activity for one target, while decreasing activity for the other target significantly; Z3 and Z4, structural changes result in significant changes in activity toward one target but not an appreciable change toward the other. See text for details.

METHODS

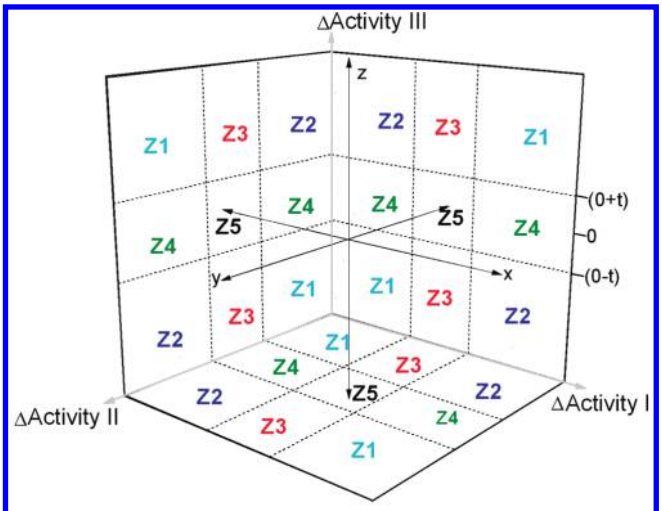
Data Set. We employed the set of 299 compounds with activity against three monoamine transporters recently used in the work of Dimova et al.[20,29] Each molecule in the data set has reported K_i values against dopamine, norepinephrine, and serotonin transporters, DA, NE, and 5HT, respectively. The distribution of activities for each target is summarized in Figure S1 (Supporting Information). This data set was used previously to model the multitarget activity landscape using NSG,[20] and it was used as a benchmark set in this work. In the previous study the activity profile was encoded classifying the compounds in three categories.[20] In contrast, in this work we used the potency values with no previous encoding. We also employed a consensus fingerprint representation as opposed to a single representation used in the previous analysis.[20]

Dual and Triple Activity-Difference Maps. The SAR of data sets tested with two biological end points, e.g., molecular targets **I** and **II**, can be characterized using DAD maps that are based on pairwise comparisons.[27] For a data set of N compounds tested with targets **I** and **II**, the DAD map depicts $N(N\text{-}1)/2$ 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

$$\Delta \mathrm{p}K_i(\mathrm{T})_{i,j} = \mathrm{p}K_i(\mathrm{T})_i - \mathrm{p}K_i(\mathrm{T})_j \quad (1)$$

where $\mathrm{p}K_i(\mathrm{T})_i$ and $\mathrm{p}K_i(\mathrm{T})_j$ are the activities of the ith and jth molecules ($j > i$). In this work, T = DA, NE, 5HT. Note that $\Delta \mathrm{p}K_i(\mathrm{T})_{i,j}$ can have positive or negative values. As described below, using the sign of the potency difference values in the activity-difference maps (as opposed to using the absolute potency difference) provides additional information concerning the *direction* of the SAR (*vide infra*).[30]

A prototype DAD map is shown in Figure 1. Vertical and horizontal lines at $\Delta \mathrm{p}K_i \pm t$ determine boundaries for low/high potency difference for targets **I** and **II**, respectively. In this work



Zone	Structure similarity category	
	High	Low
Z1Z1Z1	Triple target activity cliffs	Uninformative
Z1Z4Z4	Single target activity cliffs	Dual target scaffold hops
Z1Z2Z2	Triple target activity cliffs	Uninformative
Z3Z3Z1	Single target activity cliffs	Dual target scaffold hops
Z3Z5Z4	Continuous SAR	Triple target scaffold hops
Z3Z3Z2	Single target activity cliffs	Dual target scaffold hops
Z2Z4Z4	Single target activity cliffs	Dual target scaffold hops
Z4Z1Z3	Single target activity cliffs	Dual target scaffold hops
Z4Z4Z5	Continuous SAR	Triple target scaffold hops
Z4Z2Z3	Single target activity cliffs	Dual target scaffold hops
Z5Z3Z3	Continuous SAR	Triple target scaffold hops
Z5Z5Z5	Continuous SAR	Triple target scaffold hops

Figure 2. Prototype triple activity-difference (TAD) map for targets **I**-**III**. In essence, a TAD map is a combination of three DAD maps (Figure 1) to generate 29 symmetry-related regions that can be represented by 12 unique regions described in the table.

t = 1 (one log unit) so that data points were considered with low potency difference if $-1 \leq \Delta pK_i \leq 1$ for each target. Of course, other potency difference values ($t \neq 1$) can be used. The boundaries give rise to five general zones labeled as Z1-Z5 (Figure 1). Structural changes for molecule pairs that fall into zone 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 fall into zone 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 in 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 in zone Z5 denote a pair of compounds with similar activity (or identical if ΔActivity = 0 for both targets) against **I** and **II**. In other words, structural changes in the pairs of compounds in Z5 have little or no impact on the activity against the two targets. Of note, the classification of data points in an activity-difference map is independent of the structure similarity.

DAD maps can be extended to triple activity-difference (TAD) maps by adding a third dimension that represents the activity difference for a third target **III** (Figure 2). TAD maps can be interpreted as the grouping of three DAD maps for each combination of two targets, i.e., **I**-**II**, **I**-**III**, and **II**-**III**.

Since dual activity-difference maps are based on pairwise comparisons, it is straightforward to incorporate pairwise structure similarity into the DAD and TAD maps (for example, using a discrete or continuous color coding). Adding structure similarity information easily reveals the presence of single-, dual-, or triple-target activity cliffs as well as regions with continuous SAR for one or several targets. For example, in the DAD map in Figure 1 the pairs of compounds with high structure similarity 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 denote *dual-target activity cliffs* with *similar SAR*, i.e., small changes in structure are associated with a high change in activity for both targets, either increase or decrease. In contrast, pairs of similar compounds in Z2 indicate *dual-target activity cliffs* with *inverse SAR*, 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. Similar conclusions can be obtained from the TAD maps and are summarized in Figure 2. Therefore, mapping structure similarity in activity-difference maps distinguishes the *directionality* of the dual-target *activity cliffs*, either a similar effect or an opposite effect.

2D and 3D Fingerprint Representations and Structure Similarity. Ten 2D and four 3D fingerprints were computed. The 2D fingerprints included MACCS keys (166 bits), pharmacophore graph triangle (i.e., graph-based three point pharmacophores) (GpiDAPH3), typed graph distance (TGD), and typed graph triangle (TGT) implemented in Molecular Operating Environment (MOE);[31] Canvas[32] dendritic, topological, atom triplets, MOLPRINT 2D, atom pairs, and radial. Radial fingerprints are equivalent to the extended connectivity fingerprints (ECFPs).[33] The 3D fingerprints included three- and four-point pharmacophores from Canvas (3p and 4p) and MOE (piDAPH3 and piDAPH4). Despite the inherent issues of employing a single conformer to represent 3D structures, the 3D fingerprints are valuable to characterizing activity landscapes.[19,25,26] To simplify the analysis, a single low-energy conformation of each molecule was used.[25,34] Several conformations can also be considered as we recently reported.[26] Structure similarities were computed with the Tanimoto coefficient,[35] which has been successfully applied in a number of cases of activity landscape modeling. For example refer to the several cases reviewed in refs 1 and 2 and recent applications by our and other groups.[19,20,25−27] However, other measures such as Euclidean distance can be used.

Activity Cliffs and Scaffold Hops in Consensus SAS Maps. For reference, activity cliffs and scaffold hops were represented in consensus SAS maps for each target. SAS maps were designed to characterize pairwise SAR of individual targets.[16] Briefly, in a SAS map, activity similarity or the absolute potency difference for each molecule pair is represented on the *Y*-axis and the corresponding structure similarity is plotted on the *X*-axis. SAS maps can be roughly divided in four major quadrants. Activity cliffs will be identified in the quadrant that corresponds to high-structure similarity and high-potency difference; the scaffold hops will be

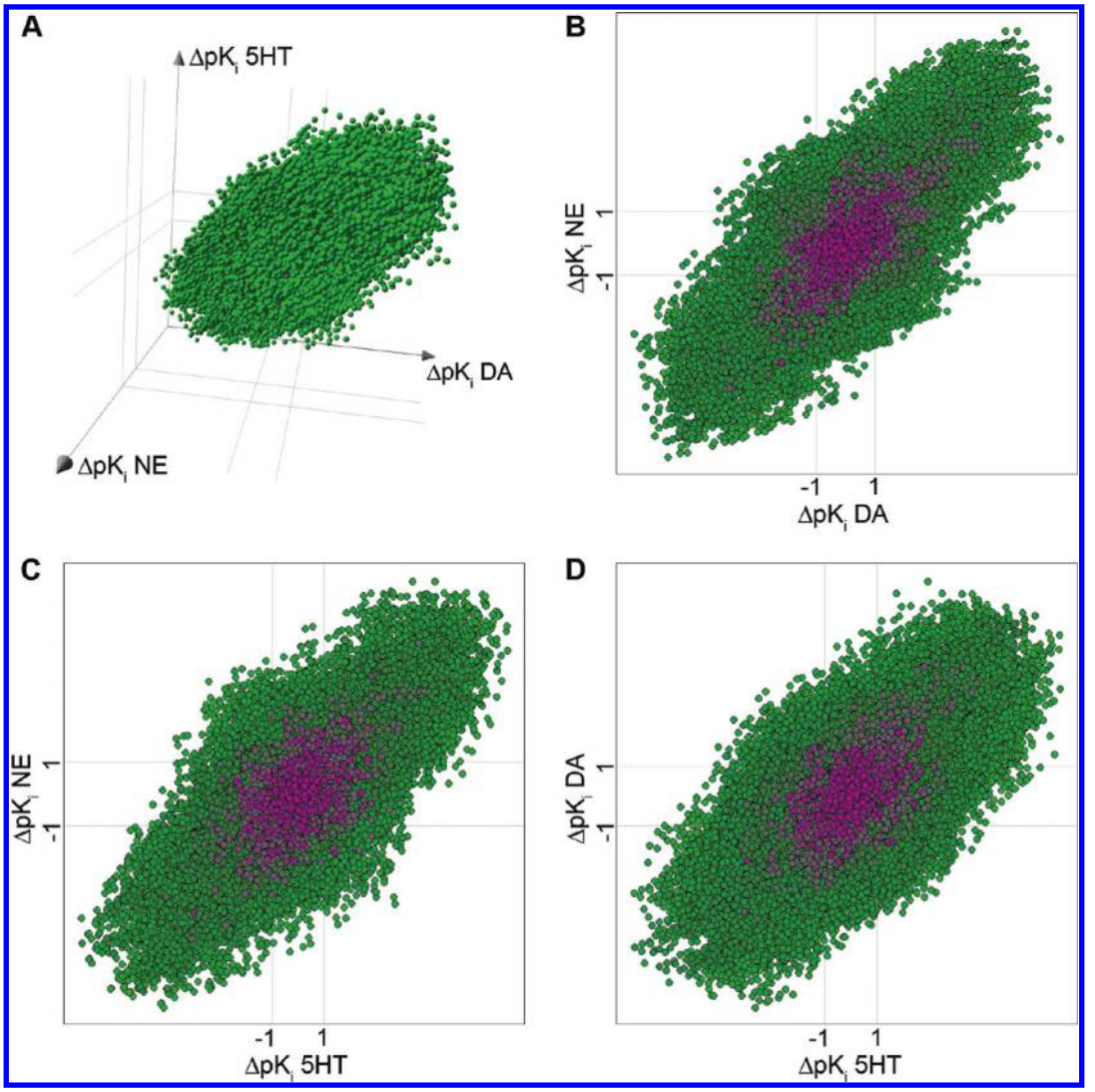


Figure 3. Dual-activity difference maps for the 299 compounds. Each data point represents a pairwise comparison of 299 compounds (i.e., 44551 data points total). Data points are color-coded by the mean structure similarity using a continuous scale from green (less similar) to magenta (more similar). (A) TAD map for DA-NE-5HT, and DAD maps for (B) DA-NE; (C) 5HT-NE; and (D) 5HT-DA.

identified in the quadrant with low structure similarity and low potency difference. Further details are presented elsewhere.[16,26] As we have shown before, consensus SAS maps can be obtained when the structure similarity is a combination of values computed with structure representations that capture different aspects of the chemical structures.[19,26] Using this approach, consensus SAS maps for each of the three targets were obtained in this work by plotting the corresponding absolute potency differences against the mean structure similarity of five selected 2D and 3D representations (*vide infra*).

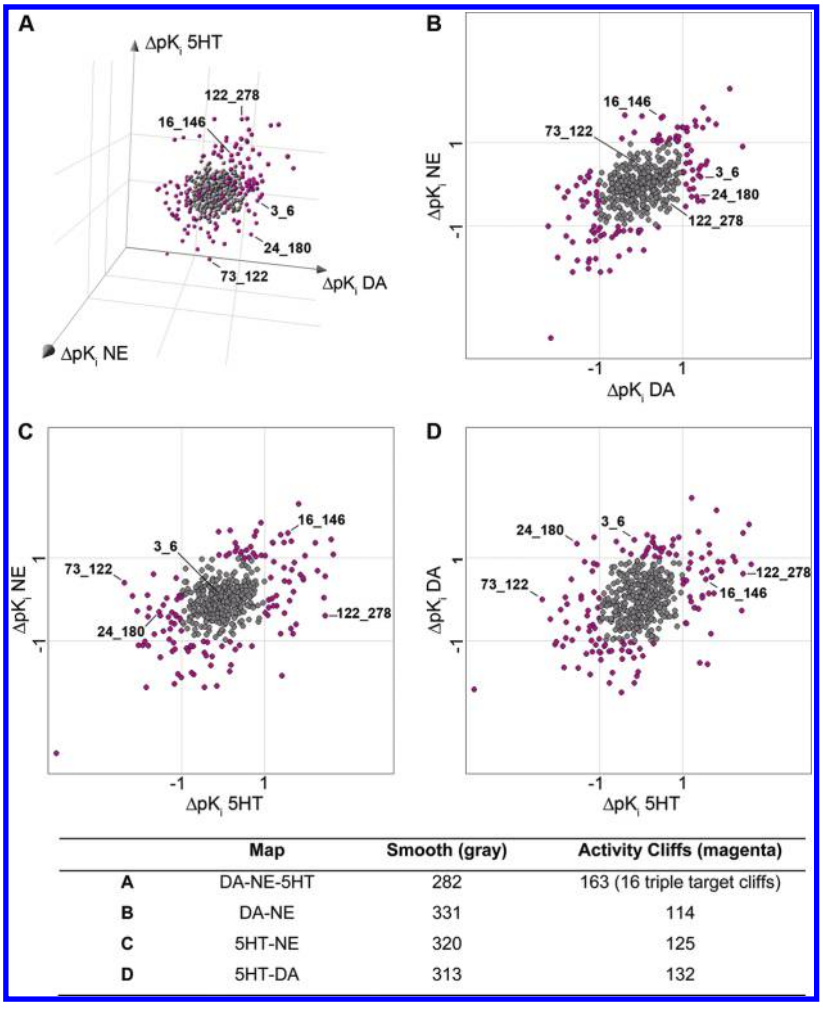
Activity Cliffs with SALI. The presence of activity cliffs for each target was also evaluated quantitatively by computing the SALI values using the expression[11,12]

$$\mathrm{SALI}_{i,j} = \frac{|A_i - A_j|}{1 - \mathrm{sim}(i,j)} \qquad (2)$$

where A_i and A_j are the activities of the ith and jth molecules, and $\mathrm{sim}(i,j)$ is the similarity coefficient between the two molecules. Any similarity method can be used to compute SALI. Mean SALI values were computed in this work using the mean structure similarity of five selected 2D and 3D molecular representations (*vide infra*). This extension of the SALI approach has been used before to detect consensus activity cliffs.[25] As will be shown later, the mean SALI values were consistent in interpreting the SAR of the data set highlighting the feasibility of using mean fusion similarity values in characterizing activity landscapes.

RESULTS AND DISCUSSION

Activity-Difference Maps. Figure 3 shows TAD and DAD maps with the 44551 pairwise potency differences of the 299 compounds in the data set. Of note, the distribution of the data points is independent of the structure similarity. Overall, the DAD map for DA-NE (Figure 3B) showed the highest correlation (Pearson's correlation coefficient between the 44551 potency differences of 0.76), while the map for DA-5HT (Figure 3D) showed the lowest (correlation coefficient of 0.58). These values mirror the correlations for the 299 pK_i values for DA, NA, and 5HT summarized in Figure S1 (Supporting Information) suggesting an overall closer agreement between the



	Map	Smooth (gray)	Activity Cliffs (magenta)
A	DA-NE-5HT	282	163 (16 triple target cliffs)
B	DA-NE	331	114
C	5HT-NE	320	125
D	5HT-DA	313	132

Figure 4. Dual-activity difference maps showing the top 1% (445) molecule pairs with the highest mean structure similarity of the data set. Pairs in gray have potency difference equal or less than one log unit for each target in the corresponding plot (zones Z5 in Figure 1 and Z5Z5Z5 in Figure 2). Pairs in magenta have potency difference greater than one log unit and represent activity cliffs. (A) TAD map for DA-NE-5HT, and DAD maps for (B) DA-NE; (C) 5HT-NE; and (D) 5HT-DA. The table summarizes the number of activity cliffs. Selected pairs are labeled with the compound numbers.

SAR of molecules targeting DA and NE. These results are in agreement with the chemical structures of the natural substrates of these transporters; the neurotransmitters dopamine and norepinephrine are structural analogues synthesized from phenylalanine, whereas 5-hydroxytryptamine (serotonin) is synthesized from tryptophan.[36] We want to emphasize that that correlations of the 299 pK_i values for DA, NA, and 5HT do not provide information concerning the structural modifications of the inhibitors associated with changes in potency against the three targets.

Activity-Difference Maps Annotated with Structure Similarity. Activity-difference maps provide systematic information concerning changes in structure associated with changes in potency for each target. To analyze the distribution of structural relationships between pairs of compounds, structural similarity was encoded in the activity difference maps. Figure 3 shows the TAD and DAD maps with the data points colored by structure similarity using a continuous scale from less similar (green) to more similar (magenta) structures. For *illustration* purposes, data points are colored by the mean similarity of five selected fingerprints (see next section). Of note, similarity values obtained with any structure representation and similarity measure can be used. Mapping similarity values into the activity-difference maps clearly revealed activity differences associated with large and small differences in chemical structures. Therefore, as described below, it was straightforward to systematically identify single-, dual-, and triple-target activity cliffs and scaffold hops.

Table 1. Counts of Activity Cliffs

type of activity cliff		total (very high potency difference)[a]
single-target	5HT	49 (5)
	NE	31 (2)
	DA	38 (1)
dual-target	DA-NE	11[b]
	5HT-NE	10[b]
	5HT-DA	8[c]
triple-target	5HT-NE-DA	16[d]

[a] Number in parentheses indicate *deep* activity cliffs with $|\Delta pK_i| > 2$. [b] All direct SAR. [c] Five direct SAR; three with inverse SAR. [d] One cliff (**168_225**) with inverse SAR for 5HT-NE and 5HT-DA but with direct SAR for DA-NE.

dx.doi.org/10.1021/ci200281v |*J. Chem. Inf. Model.* 2011, 51, 2427–2439

Table 2. Representative Consensus Activity Cliffs and Scaffold Hops Identified in the TAD and DAD Maps[a]

	pair	δpK_i DA	δpK_i NE	δpK_i 5HT	mean sim	radial	atom pairs	MACCS	TGD	piDAPH3
				Type of Activity Cliff						
single 5HT	**73_122**	0.00 (0.00)	0.395 (2.12)	−2.380 (12.76)	0.813	0.115	1.000	1.000	1.000	0.952
single 5HT	**122_278**	0.620 (3.527)	−0.395 (2.24)	2.462 (13.97)	0.824	0.210	1.000	1.000	1.000	0.909
single DA	**3_6**	1.435 (7.59)	0.177 (0.94)	−0.161 (0.85)	0.811	0.477	0.831	0.914	0.956	0.876
single NE	**118_275**	−0.755 (3.03)	−2.093 (8.40)	−0.155 (0.62)	0.751	0.248	0.716	0.826	1.000	0.965
dual 5HT-NE *direct*	**16_146**	0.483 (1.756)	1.597 (5.81)	1.559 (5.67)	0.725	0.447	0.769	0.848	0.966	0.595
dual 5HT-DA *inverse*	**24_180**	1.342 (5.23)	−0.301 (1.17)	−1.548 (6.04)	0.743	0.313	0.669	0.960	0.987	0.788
dual DA-NE *direct*	**114_195**	−1.260 (4.97)	−1.378 (5.43)	−0.264 (1.04)	0.746	0.469	0.705	0.864	0.832	0.862
triple	**91_137**	−2.174 (8.68)	−3.708 (14.82)	−4.027 (16.08)	0.750	0.367	0.789	0.902	0.822	0.868
triple	**137_196**	2.140 (13.04)	2.293 (13.96)	1.821 (11.08)	0.836	0.179	1.000	1.000	1.000	1.000
triple	**2_77**	1.805 (7.25)	1.444 (5.80)	2.609 (10.48)	0.751	0.132	0.781	0.891	1.000	0.952
triple	**15_225**	−1.651 (8.02)	−1.398 (6.79)	−1.519 (7.38)	0.794	0.422	0.873	0.931	1.000	0.744
triple	**168_225**	−1.534 (7.99)	−1.845 (9.61)	1.420 (7.40)	0.808	0.576	0.872	1.000	0.903	0.689
				Triple-Target Scaffold Hops						
	44_224	−0.471 (0.66)	0.009 (0.01)	−0.512 (0.71)	0.280	0.039	0.173	0.469	0.721	0.000
	44_148	0.089 (0.12)	0.425 (0.59)	−0.318 (0.44)	0.282	0.074	0.184	0.469	0.681	0.000
	148_225	−0.077 (0.14)	0.186 (0.34)	0.426 (0.78)	0.453	0.074	0.330	0.515	0.784	0.560
	156_197	−0.401 (0.53)	−0.458 (0.60)	−0.223 (0.29)	0.241	0.060	0.144	0.368	0.632	0.000
	49_152	0.358 (0.54)	0.170 (0.26)	0.459 (0.69)	0.339	0.042	0.172	0.281	0.778	0.422
	13_233	−0.150 (0.23)	0.407 (0.63)	−0.326 (0.50)	0.354	0.024	0.272	0.540	0.590	0.342

[a] Mean SALI values are in parentheses.

Consensus or Aggregated Similarity Measure. The number of activity cliffs identified in the activity-difference maps depends, among other factors, on the structure representation and the criteria to define 'high' structure similarity. We want to emphasize that any similarity method can be used in the TAD/DAD maps to explore the activity cliffs in the data set. To *exemplify* the distribution of the different types of activity cliffs in these maps, several 2D and 3D structure representations were considered. This strategy has been used to *reduce* the well-known dependence of the activity landscape with structure representation.[19,25−27] One approach to reduce such dependence is identifying activity cliffs common to a series of structure representations, i.e., consensus activity cliffs.[25] An alternative strategy, which was used in this work, is combining similarity measures obtained by different methods using the principles of data fusion,[37−39] for example by computing the mean similarity ("mean fusion") of selected representations.[26,27]

Herein, with the main goal of characterizing activity cliffs in the TAD and DAD maps (as opposed to comparing different fingerprints and select the "best" representation), we computed the *mean structure similarity* of radial, atom pairs, MACCS keys (166-bits), TGD (2D), and piDAPH3 (3D) fingerprints. Following a similar approach recently reported,[26] from the initial pool of 14 2D and 3D similarities (cf. the Methods section) we selected fingerprints that capture different aspects of the chemical structures[23] and showed relatively low linear correlations for the 44551 pairwise structure similarities (correlations are shown in Table S2 in the Supporting Information). The maximum correlation between any of the five selected representations was 0.72 (Table S2). Selected fingerprints have conceptually different designs. For example, MACCS keys used in this work are a predefined set of 166 structural keys; TGD are graph distances employing the following atom typing: donor, acceptor polar, anion, cation, and hydrophobe; radial fingerprints entail growing a set of fragments radially from each heavy atom over a series of iterations;[33,40] piDAPH3 are spatial three-points pharmacophores employing the following atom types: in pi system, is donor, is acceptor.

For a combined similarity measure defining quantitatively 'high' structure similarity is not straightforward. In addition to statistical significance, the threshold used to describe 'highly similar' pairs of compounds should provide meaningful and interpretable results. In this work, to *illustrate* the systematic identification of consensus activity cliffs and scaffold hops in the activity-difference maps (which is the main focus of this work), we selected the top 1% (445) pairs of compounds with the highest mean structure similarity. Table S1 (Supporting Information) summarizes the distribution of the mean-similarity for all the 44551 pairwise comparisons. Table S3 (Supporting Information) summarizes the distribution of the mean similarity for the 445 molecule pairs selected. Of note, the lowest mean similarity value of the 445 pairs (0.723) has a Z-score of 3.6 (Table S3), and it is much higher than the upper confidence interval of the mean at 95% (U95) for the mean similarity for all the 44551 pairs (0.33; Table S1). For reference, Table S3 also summarizes the distribution of the similarity values computed with radial, atom pairs, MACCS keys (166-bits), TGD, and piDAPH3 fingerprints for the 445 selected pairs. As it is shown below, the strategy employed here to define compounds with high consensus structure similarity rendered the identification of scaffold hops and activity cliffs that have an interpretable SAR meaningful for a chemoinformatitian or a medicinal chemist. Other approaches to define 'high' similarity can be explored.[25,41,42]

Characterization of Activity Cliffs. Figure 4 shows TAD and DAD maps displaying only the 445 molecule pairs with the highest mean structure similarity. Data points in gray are pairs of compounds with $-1 \leq \Delta pK_i \leq 1$ for the corresponding target

representing *smooth* regions in the activity landscape for all three targets (Figure 4A) and for each pair of targets (Figure 4B-D), i.e., pairs of compounds with high structural similarity and low potency difference. In contrast, data points in magenta correspond to the total number of activity cliffs (including single-, dual-, and triple-target cliffs). As discussed above, TAD and DAD maps keep the sign of the potency difference (in contrast to representing the absolute potency difference) to provide information about the *direction* of the SAR (*vide infra*). Figure 4 also summarizes the total number of activity cliffs and molecule pairs with smooth SAR in each map. In the TAD map, out of the 163 activity cliffs, just 16 are triple-target cliffs (the remaining are dual-target cliffs). The DAD map for 5HT-DA showed the highest number of cliffs of the three DAD maps suggesting less agreement between the SAR for these two targets. Thus, depending on the direction of the SAR, TAD and DAD maps provide a quick access to SAR information that can be used to optimize compounds to enhance activity toward both targets or to increase selectivity for one target.

The number of single-, dual-, and triple-target activity cliffs can be easily deduced from the TAD and DAD maps shown in Figure 4. Table 1 summarizes the number of single-, dual-, and triple-target activity cliffs indicating that few cliffs (less than 50, 0.1%) were identified. This is not unexpected, considering that activity cliffs are not anticipated to be the most prevalent features of an activity landscape, although they certainly are among the most interesting from a SAR perspective.[43] Significantly, activity cliffs indicate regions in the activity landscapes that contain maximum information on SARs. This is because changes in molecular similarity make it easier to identify what features may be responsible for the dramatic shifts observed in activity.[43] For all single-target cliffs pairs of compounds with *very* high potency difference (more than two log units) were identified, and the number is indicated in parentheses (Table 1). These activity cliffs are distinct as they can be considered *deep* activity cliffs.[19,44] 5HT showed the highest number of single-target activity cliffs (49) and deep cliffs (5). Hence, 5HT was the most sensitive monoamine transporter to the compounds in the data set, suggesting a relative ease in optimizing the compounds toward 5HT. In contrast, no deep activity cliffs were identified in any of the three DAD maps. Of note, out of the eight cliffs in the DAD map of 5HT-DA, five showed *direct SAR*, i.e., the small change in the structure of the molecules in the pair is associated with an increase (or decrease) in the potency for 5HT and DA, while three cliffs showed an *inverse SAR*, i.e., the small change in structure of the molecules in the pair is associated with an increase in the potency for one target but a decrease in potency in the second target. Concerning triple-target cliffs, one (out of 16) showed *inverse SAR* for 5HT-NE and 5HT-DA but direct SAR for DA-NE.

Table 2 shows several examples of single-, dual- (with direct and inverse SAR), and triple-target activity cliffs. Table 2 also lists the potency difference, the corresponding mean similarity as well as the similarity values obtained with the five selected structure representations. A side-by-side comparison of the chemical structures for selected pairs along with their position in the corresponding DAD and TAD maps is illustrated in Figures 3-5.

Single- and Dual-Target Activity Cliffs. The pairs of compounds **73_122** and **122_278**, in zone Z4 of the DAD maps 5HT-NE (Figure 4C) and 5HT-DA (Figure 4D), exemplify single-target cliffs for 5HT. Both pairs are deep cliffs with potency differences larger than two log units. The same two

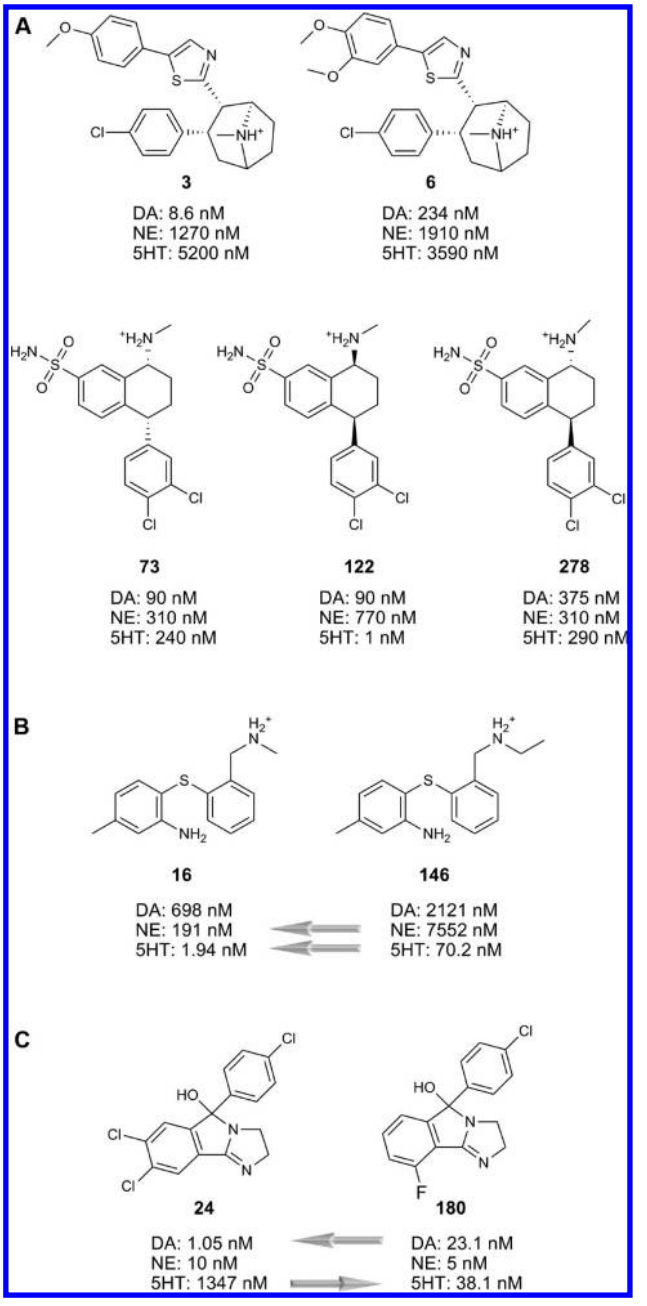


Figure 5. Chemical structures and K_i values for examples of activity cliffs: (A) single-target; (B) dual-target with *direct* SAR; and (C) dual-target with *inverse* SAR. In panels B and C the *direction* of the activity (increase in potency) is indicated with arrows. The position of the molecule pairs in the DAD/TAD maps is indicated in Figure 4. The values of potency difference and structure similarity are summarized in Table 2. See text for details.

pairs, however, are in zone Z5 of the DAD map DA-NE (Figure 4B) with potency differences of less than one log unit for DA and NE (Table 2). The only difference in the chemical structures of **73_122** and **122_278** is the stereochemistry at two chiral centers (Figure 5A). Of note, several fingerprints cannot distinguish the two pairs; the similarity for several representations is 1.0, except for the radial fingerprint (Table 2). This observation further emphasizes the convenience of considering multiple structure representations to characterize activity landscapes as suggested previously.[19,25,26]

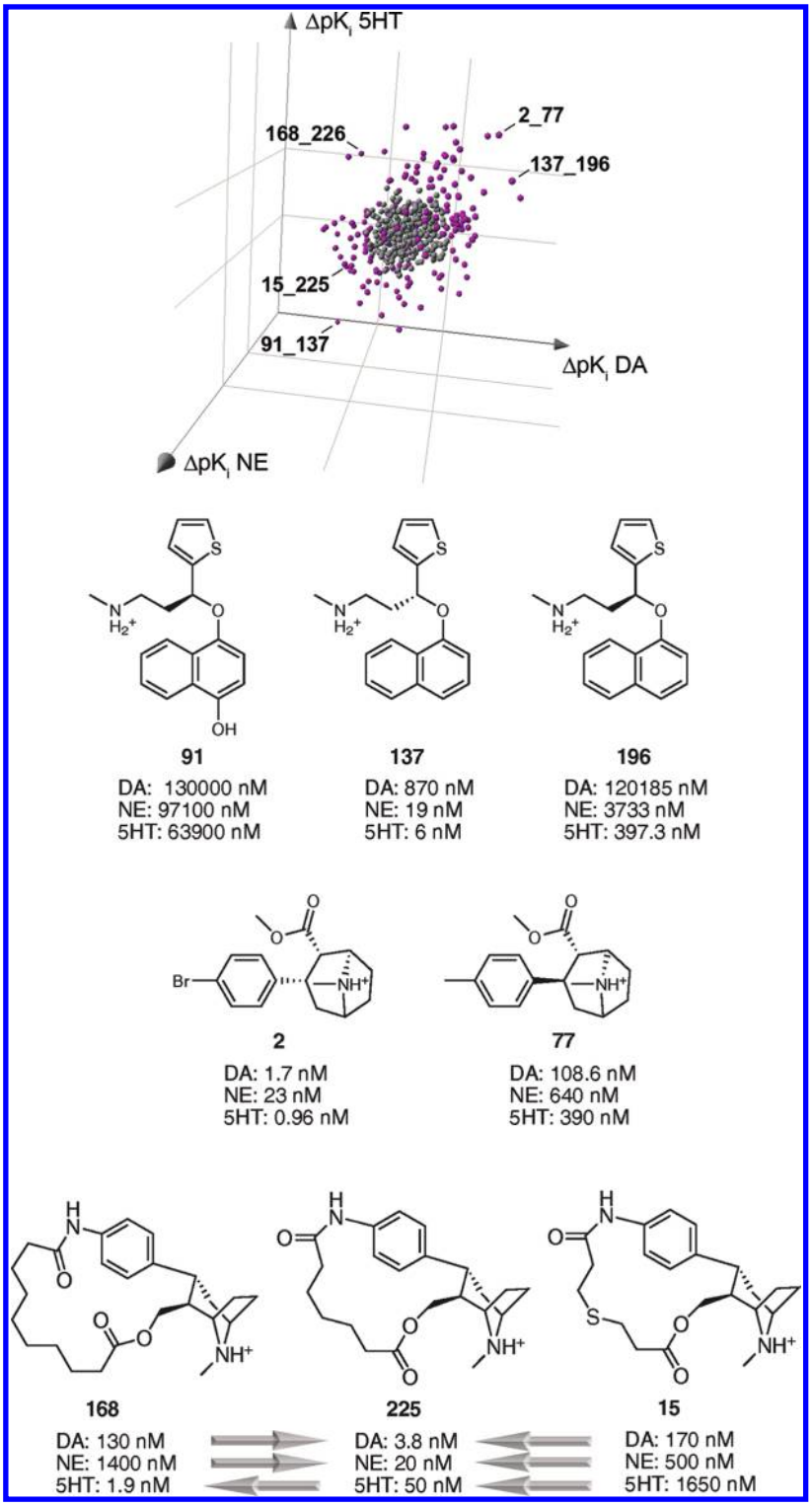


Figure 6. Chemical structures and K_i values for five representative triple-target activity cliffs. The position of the molecule pairs in the TAD map is indicated. The values of potency difference and structure similarity are summarized in Table 2. For the molecule pairs **168_225** and **15_225** the *direction* of the activity (increase in potency) is indicated with arrows. See text for details.

Pairs **3_6** and **118_275** (Table 2) are examples of single-target cliffs for DA and NE, respectively. The position of **3_6** in the corresponding DAD maps is indicated in Figure 4 that shows the large potency difference for DA but low potency difference for 5HT and NE; Figure 5A clearly shows that while compound **3** is selective for DA over 5HT and NE, addition of a methoxy group in **6** gives rise to a significant decrease in selectivity.

Representative examples of dual-target activity cliffs for 5HT-NE and 5HT-DA are the pairs **16_146** and **24_180**, respectively. The pair **16_146** has a *direct* SAR (located in region Z1 of the

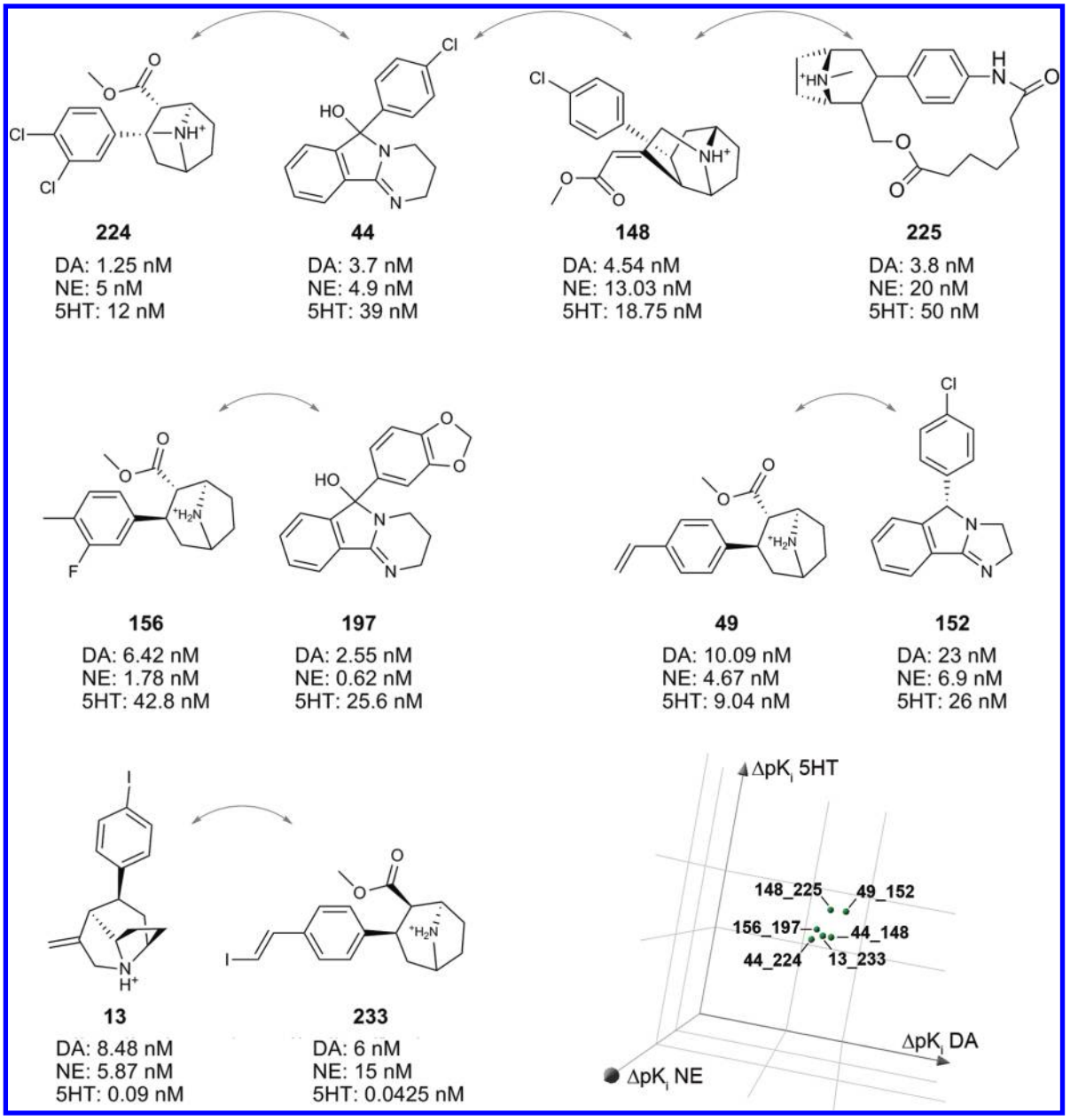


Figure 7. Chemical structures and K_i values for six representative consensus scaffold hops for all three targets. The arrows schematically illustrate the scaffold "hopping" between each pair. The position of the molecule pairs in the TAD map is indicated (zone Z5Z5Z5 in Figure 2). The values of potency difference and structure similarity are summarized in Table 2.

DAD map for 5HT-NE, Figure 4C); the change of an *N*-methyl group in **16** to an *N*-ethyl group in **146** leads to a significant decrease in the activity for both 5HT and NE (Figure 5B). In sharp contrast, the pair **24_180** has an *inverse* SAR (located in region Z2 of the DAD map for 5HT-DA, Figure 4D); the change of two chlorine aromatic substitutions in **24** to a single fluorine substitution in **180** leads to a large increase in activity for 5HT but to a large decrease in activity for DA (Figure 5C).

Triple-Target Activity Cliffs. Figure 6 shows the chemical structures of five representative activity cliffs for all three targets and their positions in the TAD map (corresponding to zones Z1Z1Z1 or Z1Z2Z2 in the prototype TAD map in Figure 2). Concerning the cliff **137_196**, change in the stereochemistry from *R* in **137** to *S* in **196** dramatically reduces the activity for DA, NE, and 5HT by more than 1.5 log units (Table 2). The pair **91_137** is another example of a deep activity cliff; the change in the stereochemistry from *R* in **137** to *S*, and the addition of a hydroxyl group to the naphthalene moiety in **91** leads to a large decrease in the activity for all three targets in more than two log units (more than three log units for 5HT and NE). These results are in agreement with previous studies reporting the importance in the activity of the chiral center in compound **137** and the effect of introducing hydroxyl groups in the naphthalene moiety.[20] Another example of a triple-target activity cliff is the pair **2_77**; change of a bromine atom in **2** to a methyl group in **77** decreases significantly the activity for all targets, in particular for 5HT (more than 2.5 log units). Additional examples of triple-target cliffs are shown in Figure S2 of the Supporting Information.

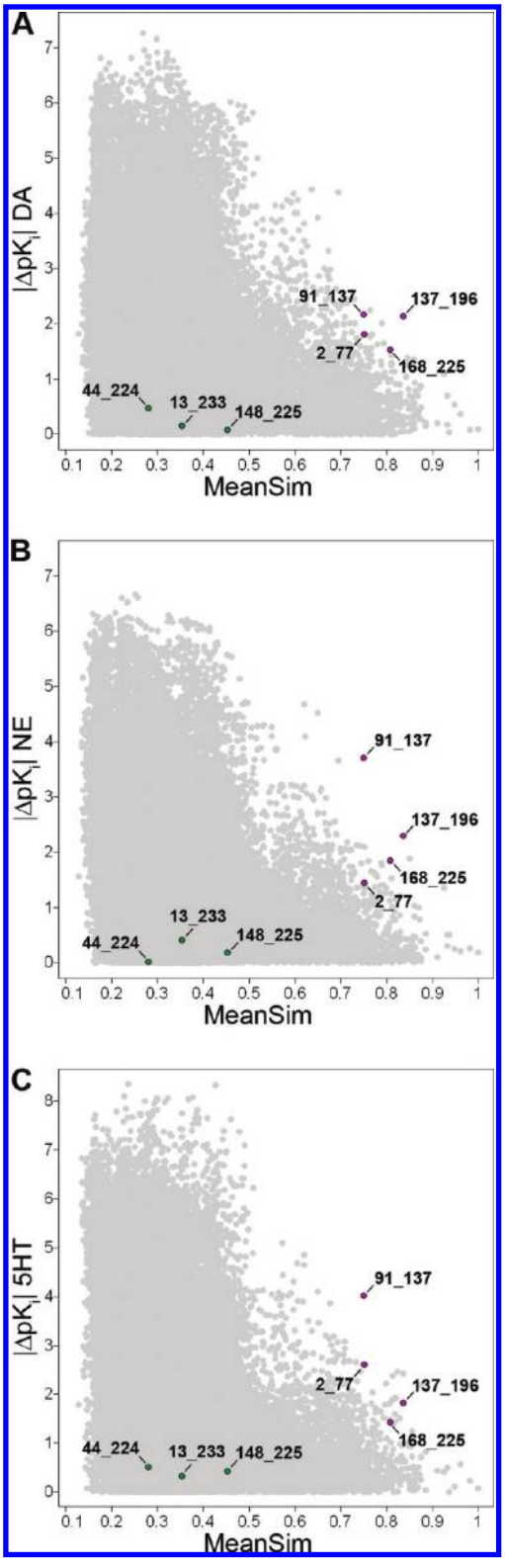


Figure 8. Consensus SAS maps for the three monoamine transporters. (A) DA; (B) NE; and (C) 5HT. Each data point indicates a pairwise comparison of 299 compounds (i.e., a total of 44551 data points). Selected consensus triple-target activity cliffs (magenta) and scaffold hops (green) are labeled with the compound numbers. The remaining pairs are displayed in light gray. The values of potency difference and structure similarity are summarized in Table 2.

Table 3. Distribution of Mean SALI Values for Each Monoamine Transporter

target	max	Q3[a]	median	Q1[b]	min	U95[c]	mean	L95[c]	STD
DA	14.60	3.61	2.07	0.97	0.00	2.51	2.49	2.48	1.92
NE	18.14	3.47	2.02	0.93	0.00	2.39	2.37	2.35	1.79
5HT	17.50	4.27	2.53	1.19	0.00	2.98	2.96	2.94	2.21

[a] Third quartile. [b] First quartile. [c] 95% confidence of the mean upper (U95) and lower (L95) limits.

All triple-target activity cliffs discussed above lead to the same effect (increase or decrease) in activity for all three targets i.e., direct SAR (all are located in zone Z1Z1Z1 of the prototype TAD map, Figure 2). In contrast, the pair **168_225** is a triple-target cliff where a small change in the structure (size of the macrocyclic structure) gives rise to a large *increase* in activity for DA and NE but to a large *decrease* in activity for 5HT, i.e., direct SAR for DA-NE but inverse SAR for 5HT-NE and 5HT-DA. This pair is in zone Z1Z2Z2. However, in the molecule pair **15_225** (zone Z1Z1Z1), the change of a carbon atom in the cyclic structure of **225** to a sulfur atom in **15** decreases the activity for all three targets by more than one log unit (Table 2 and Figure 6). Compound **225** forms additional triple-target cliffs with **214** and **273** with direct SAR (chemical structures and activity data is shown in Figure S2 of the Supporting Information).

Scaffold Hops. Scaffold hopping is a useful technique in drug design to "jump" in different areas of chemical space.[28,45] Scaffold hops are characterized by having structurally different templates but equivalent biological activity. Scaffold hopping is used by medicinal chemists for example to avoid some undesirable ADME-tox properties; move from complex natural products to more easily synthesizable small molecules or for Intellectual Property (IP) reasons, to name a few.[45,46] Mapping structure similarity into the activity-difference maps clearly reveals such scaffold hops, i.e., pairs of compounds with very low potency difference but significantly different chemical structures, i.e., low structure similarity. Scaffold hops will be identified in the center of the activity-difference maps. To *illustrate* the presence and position of scaffold hops in the data set, Table 2 lists examples of triple-target scaffold hops along with the potency difference and similarity measures. A side-by-side comparison of the chemical structures for selected pairs is shown in Figure 7, including their position in the corresponding TAD map. It is clear from Table 2 and Figure 7 that pairs of compounds such as **44_224**, **44_148**, and **148_225** have very different chemical structures as captured by all representations. Indeed, the chemical scaffolds are different. However, all ten compounds in Figure 7 are active against the three targets and are among the most active compounds (pK_i <100 nM) in the data set for DA, NE, and 5HT. These results are significant in light of recent trends pointing to a desire to seek compounds that target these three monoamine transporters.[47] Single- and dual-target scaffolds hops can also be explored in the activity-difference maps (data not shown).

Activity Cliffs and Scaffold Hops in Consensus SAS Maps. SAS maps and consensus SAS maps are well-established methods to describe the relationship between structure similarity and activity similarity for all pairs of compounds in a data set. As a *reference*, Figure 8 shows consensus SAS maps for DA, NE, and 5HT showing the position of selected triple-target activity cliffs and triple-target scaffold hops discussed above (Table 2). The mean structure similarity (cf. the Methods section) is plotted in

the *X*-axis. Figure 8 clearly shows that the pairs of compounds **2_77**, **91_137**, **137_196**, **168_225** are located in the activity-cliff region (high molecular similarity, high potency difference) for all three targets. In turn, molecule pairs **44_224**, **13_223**, **148_225** are in the scaffold-hop region (low molecular similarity, low potency difference) for all targets. Similar conclusions are obtained from other molecule pairs. It is worth pointing out that these classifications are also consistent with those of the DAD and TAD maps. The SAS maps can be further characterized comprehensively using the approaches we have described previously for several data sets.[19,25,26] However, such comprehensive characterization is beyond the scope of this work that is mainly focused on the activity-difference maps. Despite the fact that SAS maps were initially designed to explore the SAR of single targets, it is possible to generate 2D and 3D SAS maps by plotting absolute potency differences for two or three targets, respectively; structure similarity can be then mapped into the 2D/3D SAS map (data not shown). Single-, dual-, or triple-target activity cliffs are also straightforward to uncover in these plots. Note, however, that the *directionally* of the SAR is lost.

Activity Cliffs (And Scaffold Hops) with SALI. The SALI parameter in eq 2 is a metric to easily detect activity cliffs. This index has been used to represent SAR as a graph (SALI networks).[12] An extension of this approach is the *mean SALI* value that combines the information from different structure representations to detect consensus activity cliffs.[25] We computed the mean SALI for the data set as described in the Methods, and a summary of the distribution of the values for each target is presented in Table 3. DA and NE showed similar distributions. Overall, 5HT had slightly higher values of mean SALI as indicated by the mean and median values. This result is consistent with the higher number of activity cliffs identified in the activity-difference maps (*vide supra*). The corresponding mean SALI values for selected molecule pairs are indicated in Table 2. Single-target cliffs such as **73_122** and **122_278** have high SALI values only for one target. For example, the SALI value for 5HT of **73_122** (12.76) is higher than the U95 value of the SALI distribution for 5HT (2.51, Table 3). Triple-target cliffs such as **91_137** had high SALI values for DA, NE, and 5HT namely, 8.68, 14.81, and 16.08, respectively. Of note, these values are higher than the corresponding U95 values of SALI for each target: 2.51, 2.39, and 2.98, respectively (Table 3). Similar conclusions can be obtained for all single-, dual-, and triple-target cliffs in Table 2. It is worth noting that in a previous study using the same data set triple-activity cliffs were not identified.[20] This difference is due to the approach used to encode the activity data (*vide supra*).

Table 2 also summarizes the mean SALI values of representative triple-target scaffold hops. Note that the SALI values for the scaffold hops are lower than the corresponding L95 (lower limit of the mean at 95% of confidence) value of the corresponding distribution of SALI values (Table 3). The low mean SALI values for these molecular pairs, e.g., **44_224**, **44_148**, **13_233**, are in agreement with the "scaffold hop" nature of these pairs.

CONCLUSIONS

We report two- and three-dimensional activity-difference (DAD/TAD) maps for the systematic characterization of the multitarget SAR of 299 diverse compounds screened against three monoamine transporters. DAD/TAD maps are based on pairwise comparisons of potency differences making straightforward to represent molecular similarity and systematically exploring single-, dual-, and triple-target activity cliffs as well as scaffold hops. To illustrate the application of the DAD/TAD maps, a single consensus or aggregated similarity measure was derived by combining the similarity values of five different similarity methods; namely radial, atom pairs, MACCS keys (166-bits), TGD (2D), and piDAPH3 (3D) fingerprints. The use of multiple fingerprints served to reduce the dependence of the activity landscape with molecular representation. However, DAD/TAD maps are not limited to a particular representation (or set of selected representations) or protocol to aggregate similarity measures.

For each DAD or TAD map the number of activity cliffs was quantified and several single-, dual-, and triple-target activity cliffs were identified. Deep activity cliffs were further differentiated from regular cliffs as those pairs of compounds with particularly large potency difference (in this work with more than two log units in potency difference in contrast to one log unit for regular cliffs). Dual-target cliffs were also analyzed considering the *direction* of the SAR as *direct* SAR if the change in the structure increased (or decreased) the activity of the pair for both targets, and *inverse SAR* if the structural modification increased the activity for one target but decreased the activity for the second target. Mapping molecular similarity into the DAD/TAD maps was used to identify representative triple-target scaffold hops. Selected activity cliffs were also represented in the recently developed consensus SAS maps. In addition, consensus activity cliffs were quantitatively assessed using a proposed extension of the SALI metric that considers multiple structure representations. Taken together, the results presented in this work show that DAD/TAD maps are complementary tools to systematically explore the multitarget activity landscape of compound data sets.

We want to emphasize that the present study is a systematic description of the activity landscape of a data set using DAD/TAD maps. A major perspective of this work is to explore the predictive capabilities of the activity landscape models to anticipate the SAR of new molecules in a prospective manner. This is an area of intense research in our and other research groups.

ASSOCIATED CONTENT

Supporting Information. Summary of the biological activity data (Figure S1); summary of the distribution of the 44551 pairwise structure similarities (Table S1); correlation matrix for the 44551 pairwise structure similarities (Table S2); distribution of similarity values of the top 1% (445) molecule pairs with the highest mean similarity (Table S3); additional examples of triple-target activity cliffs. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

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

ACKNOWLEDGMENT

Helpful discussions and comments of Dr. Gerald M. Maggiora are gratefully acknowledged. We also thank Prof. Dr. Jürgen

Bajorath for making available the data set used in this work. Authors are also grateful to Jacob Waddell for help with preparing figures for the manuscript. This work was supported by the State of Florida, Executive Office of the Governor's Office of Tourism, Trade, and Economic Development.

REFERENCES

(1) Wawer, M.; Lounkine, E.; Wassermann, A. M.; Bajorath, J. Data Structures and Computational Tools for the Extraction of SAR Information from Large Compound Sets. *Drug Discovery Today* **2010**, *15*, 630–639.

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

(3) Ooms, F. Molecular Modeling and Computer Aided Drug Design. Examples of Their Applications in Medicinal Chemistry. *Curr. Med. Chem.* **2000**, *7*, 141–158.

(4) Ma, X. H.; Jia, J.; Zhu, F.; Xue, Y.; Li, Z. R.; Chen, Y. Z. Comparative Analysis of Machine Learning Methods in Ligand-Based Virtual Screening of Large Compound Libraries. *Comb. Chem. High Throughput Screening* **2009**, *12*, 344–357.

(5) Yang, S. Y. Pharmacophore Modeling and Applications in Drug Discovery: Challenges and Recent Advances. *Drug Discovery Today* **2010**, *15*, 444–450.

(6) Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inf.* **2010**, *29*, 476–488.

(7) 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.

(8) Peltason, L.; Bajorath, J. Molecular similarity analysis in virtual screening. In *Chemoinformatics Approaches to Virtual Screening*; Varnek, A., Tropsha, A., Eds.; Royal Society of Chemistry: Cambridge, UK, 2008; pp 120−149.

(9) Scior, T.; Medina-Franco, J. L.; Do, Q. T.; Martínez-Mayorga, K.; Yunes Rojas, J. A.; Bernard, P. How to Recognize and Workaround Pitfalls in QSAR Studies: A Critical Review. *Curr. Med. Chem.* **2009**, *16*, 4297–4313.

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

(11) 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.

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

(13) Wassermann, A. M.; Peltason, L.; Bajorath, J. Computational Analysis of Multi-Target Structure-Activity Relationships to Derive Preference Orders for Chemical Modifications toward Target Selectivity. *ChemMedChem* **2010**, *5*, 847–858.

(14) Medina-Franco, J. L.; Martínez-Mayorga, K.; Bender, A.; Scior, T. Scaffold Diversity Analysis of Compound Data Sets Using an Entropy-Based Measure. *QSAR Comb. Sci.* **2009**, *28*, 1551–1560.

(15) Peltason, L.; Bajorath, J. SAR Index: Quantifying the Nature of Structure-Activity Relationships. *J. Med. Chem.* **2007**, *50*, 5571–5578.

(16) 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*, American Chemical Society: Washington, DC, Chicago, IL, United States, 2001.

(17) Wawer, M.; Peltason, L.; Weskamp, N.; Teckentrup, A.; Bajorath, J. Structure-Activity Relationship Anatomy by Network-Like Similarity Graphs and Local Structure-Activity Relationship Indices. *J. Med. Chem.* **2008**, *51*, 6075–6084.

(18) 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.

(19) 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. *Bioorg. Med. Chem.* **2010**, *18*, 7380–7391.

(20) Dimova, D.; Wawer, M.; Wassermann, A. M.; Bajorath, J. Design of Multitarget Activity Landscapes That Capture Hierarchical Activity Cliff Distributions. *J. Chem. Inf. Model.* **2011**, *51*, 258–266.

(21) Peltason, L.; Bajorath, J. Molecular Similarity Analysis Uncovers Heterogeneous Structure-Activity Relationships and Variable Activity Landscapes. *Chem. Biol.* **2007**, *14*, 489–497.

(22) 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.

(23) 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.

(24) 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.

(25) 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.

(26) 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.

(27) 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.

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

(29) Data set available at http://www.limes.uni-bonn.de/forschung/abteilungen/Bajorath/labwebsite (accessed May 14, 2011).

(30) The expression in eq 1 is different from a measure of selectivity of molecule m (S_m) for two targets, T1 and T2: $S_m = pK_i(T1)_m - pK_i(T2)_m$.

(31) Molecular Operating Environment (MOE), *version 2008*; Chemical Computing Group Inc.: Montreal, Quebec, Canada. Available at http://www.chemcomp.com (accessed June 16, 2011).

(32) Canvas, *version 1.3*; Schrödinger, LLC, New York, NY, 2010.

(33) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.

(34) The single low-energy conformation was obtained after 'washing' the compounds with MOE and energy-minimize the structures with the Merck Molecular Force Field 94x.

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

(36) Kalat, J. W. Chemical events in the synapse. In *Biological Psychology*, 9th ed.; Kalat, J. W., Ed.; Thompson Wadsworth: Belmont, 2007; pp 58−69.

(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) Sastry, M.; Lowrie, J. F.; Dixon, S. L.; Sherman, W. Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments. *J. Chem. Inf. Model.* **2010**, *50*, 771–784.

(41) 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.

(42) For example, we also investigated the mean similarity of the active compounds as criteria to define high similarity. However, the most active compounds in this data set e.g., 35 molecules with $K_i \leq 100$ nM for all targets, are as diverse as the entire set giving rise to not interpretable SAR.

(43) 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.

(44) Peltason, L.; Iyer, P.; Bajorath, J. Rationalizing Three-Dimensional Activity Landscapes and the Influence of Molecular Representations on Landscape Topology and the Formation of Activity Cliffs. *J. Chem. Inf. Model.* **2010**, *50*, 1021–1033.

(45) Ciapetti, P.; Giethlen, B. Molecular variations based on isosteric replacements. In *The Practice of Medicinal Chemistry*, 3rd ed.; Wermuth, C. G., Ed.; Elsevier: 2008; pp 290−342.

(46) López-Vallejo, F.; Castillo, R.; Yépez-Mulia, L.; Medina-Franco, J. L. Benzotriazoles and Indazoles Are Scaffolds with Biological Activity Against *Entamoeba Histolytica*. *J. Biomol. Screening* (**2011**), in press. DOI: 10.1177/1087057111414902.

(47) Chen, Z.; Skolnick, P. Triple Uptake Inhibitors: Therapeutic Potential in Depression and Beyond. *Expert Opin. Invest. Drugs* **2007**, *16*, 1365–1377.