

# Activity cliffs and activity cliff generators based on chemotype-related activity landscapes

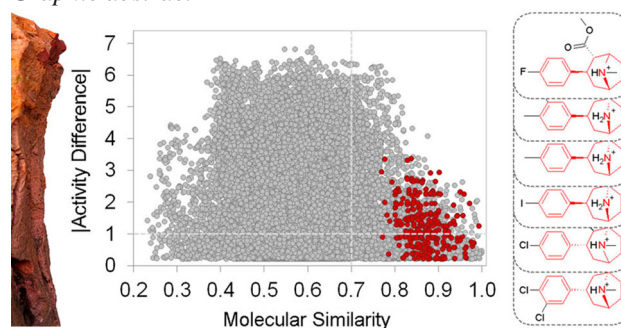
Jaime Pérez-Villanueva<sup>1</sup>  · Oscar Méndez-Lucio<sup>2,3</sup> · Olivia Soria-Arteche<sup>1</sup> · José L. Medina-Franco<sup>2</sup>

Received: 15 April 2015 / Accepted: 24 June 2015 / Published online: 7 July 2015  
© Springer International Publishing Switzerland 2015

**Abstract** Activity cliffs have large impact in drug discovery; therefore, their detection and quantification are of major importance. This work introduces the metric *activity cliff enrichment factor* and expands the previously reported *activity cliff generator* concept by adding chemotype information to representations of the activity landscape. To exemplify these concepts, three molecular databases with multiple biological activities were characterized. Compounds in each database were grouped into chemotype classes. Then, pairwise comparisons of structure similarities and activity differences were calculated for each compound and used to construct chemotype-based structure–activity similarity (SAS) maps. Different landscape distributions among four major regions of the SAS maps were observed for different subsets of molecules grouped in chemotypes. Based on this observation, the activity cliff enrichment factor was calculated to numerically detect chemotypes enriched in activity

cliffs. Several chemotype classes were detected having major proportion of activity cliffs than the entire database. In addition, some chemotype classes comprising compounds with smooth structure activity relationships (SAR) were detected. Finally, the activity cliff generator concept was applied to compounds grouped in chemotypes to extract valuable SAR information.

## Graphic abstract



**Electronic supplementary material** The online version of this article (doi:10.1007/s11030-015-9609-z) contains supplementary material, which is available to authorized users.

This paper is dedicated to the Memory of Dra. Maria Concepción Lozada García.

✉ Jaime Pérez-Villanueva  
jpvillanueva@correo.xoc.uam.mx

<sup>1</sup> División de Ciencias Biológicas y de la Salud, Departamento de Sistemas Biológicos, Universidad Autónoma Metropolitana Unidad Xochimilco (UAM-X), 04960 Mexico, DF, Mexico

<sup>2</sup> Departamento de Farmacia, Facultad de Química, Universidad Nacional Autónoma de México (UNAM), 04510 Mexico, DF, Mexico

<sup>3</sup> Present Address: Unilever Centre for Molecular Science Informatics Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

**Keywords** Activity cliff generators · Activity cliff enrichment factor · Chemotype · SAS maps · Structure–activity relationships

## Abbreviations

ACEF	Activity cliff enrichment factor
COX	Cyclooxygenase
DAT	Dopamine transporter
ECFP	Extended connectivity fingerprint
EstateIndices	Electrotopological state indices
MACCS	Molecular ACCess System
MATs	Monoamine transporters

MEQI	Molecular Equivalence Indices
MEQNUM	Molecular equivalence number
NAC/CF	Number of activity cliffs / chemotype frequency
NET	Norepinephrine transporter
NSGs	Network-like similarity graphs
PPAR	Peroxisome proliferator-activated receptor
QSAR	Quantitative structure–activity relationships
ROCS	Rapid overlay of chemical structures
SALI	Structure–activity landscape index
SARI	SAR index
SAS	Structure–activity similarity
SERT	Serotonin transporter
SAR	Structure–activity relationships
TopAtomPairs	Topological atom pairs
TopPh4AtomPairs	Topological pharmacophore atom pairs
TopAtomTorsions	Topological atom torsions
TopAtomTriplets	Topological atom triplets
TopPh4AtomTriplets	Topological pharmacophore atom triplets

## Introduction

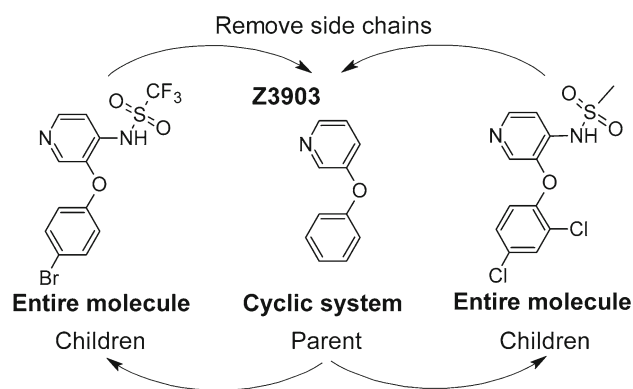
*Activity cliffs* can be defined as pairs of similar compounds with very high differences in activity [1]. Certainly, these kinds of compounds are of interest in medicinal chemistry because they are associated with high SAR information content [2,3]. It is worth mentioning that *activity cliffs* fall out of the similarity-property principle and are usually incorrectly predicted by quantitative structure–activity relationships (QSAR) models [1,4,5]; therefore, their visualization and quantification in screening datasets are of interest on SAR and QSAR studies [2]. Graphical methods designed to detect activity cliffs are inspired by the emerging concept of the activity landscape [2]. An activity landscape is defined as any representation that integrates similarity and potency relationships between compounds sharing the same biological activity [3,6]. These methods include structure–activity similarity (SAS) maps [7], network-like similarity graphs (NSGs) [8], structure–activity landscape index (SALI) graphs [9], etc. In addition to graphical methods, numerical SAR analysis functions have been introduced. These functions aid to characterize the SAR nature of molecular databases based on indices, i.e., Structure–Activity Landscape Index (SALI) and SAR Index (SARI) [9–11]. Furthermore, some methods to characterize local SAR information and activity cliffs have been reported [12,13]. In a recent work, Kayastha et al. proposed local SAS maps as representations formed

by plotting all compound pairs formed by a specific compounds [12]. This concept is related to our previous work where an activity cliff generator was defined as a molecule with high probability to form activity cliffs with structurally similar molecules tested in the same biological assay [13]. Additionally, Hu and Bajorath proposed the structural categorization of activity cliffs as *R-group cliffs* (compound pairs with same scaffold and different R-groups), *chirality and topology cliffs* (compound pairs with same scaffold and R-groups), and *scaffold cliffs* (compound pairs with different scaffold and same R-group) [14]. As part of our-continued effort to further advance the concept of local activity landscapes and their characterization, i.e., activity cliff detection, herein we use a chemotype-based classification to define local activity landscapes. Then, chemotype-based activity landscapes were analyzed employing the concepts of activity cliff enrichment factor (ACEF) and activity cliff generators. To illustrate our results, three screening datasets with different sizes were analyzed including compounds screened against three monoamine transporters (MATs): dopamine (DAT), norepinephrine (NET), and serotonin (SERT) transporters, two cyclooxygenases (COX-1 and -2), and three peroxisome proliferator-activated receptors (PPAR  $\alpha$ ,  $\delta$  and  $\gamma$ ). MATs have received considerable attention as targets for psychiatric and neurological disorders as depression, attention deficit, hyperactivity disorder, Parkinson's disease, and schizophrenia [15,16]. Cyclooxygenases are important since they are targeted by non-steroidal anti-inflammatory drugs and have been proposed as target for cancer therapy in combination with other chemotherapeutic agents [17,18]. Finally, the activation of PPARs offers a promising strategy for the treatment of diabetes mellitus, obesity, and related cardiovascular complications [19–21].

## Materials and methods

### Dataset

Three molecular databases previously used for activity landscape modeling studies were used in this work as benchmark datasets [21–24]. The first database includes 299 compounds tested against three monoamine transporters recently used by Dimova et al. [22]. Each molecule in the dataset has reported  $K_i$  values against DAT, NET, and SERT. A second set has 658 cyclooxygenase inhibitors obtained from the Binding Database [25–27]. These compounds have reported activity (IC<sub>50</sub>) against COX-1 and COX-2. The third set of compounds comprises 168 PPAR ligands, with IC<sub>50</sub> values against  $\alpha$ ,  $\delta$ , and  $\gamma$  receptors, obtained from the Binding Database. A general overview of activities, for each target, is summarized in the parameters presented in the Supporting Information (Table S1).



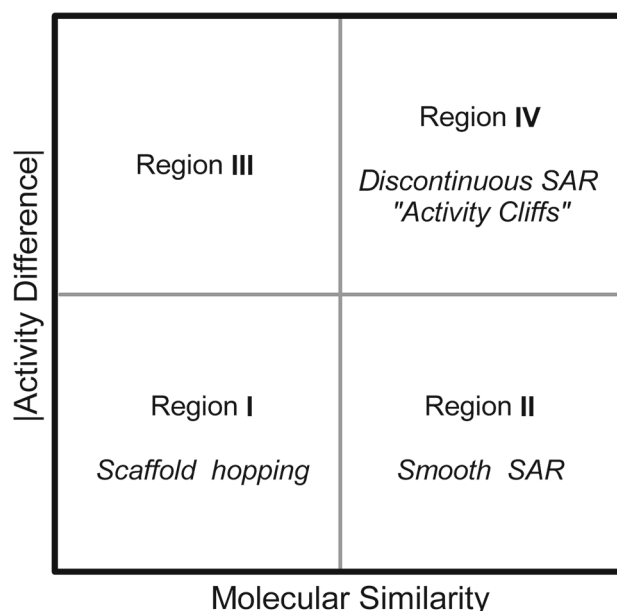
**Fig. 1** Example of the generation of chemotypes using cyclic systems resolution level. The resulting chemotypes were used to group compounds with the same molecular scaffold

### Molecular classification into chemotype classes

The chemotype of each molecule in the datasets was calculated using molecular equivalence indices (MEQI) at cyclic system resolution [28–32]. Those compounds that presented the same chemotype were grouped together. The generation of chemotypes at cyclic system resolution consists of removing all side chains from the molecule, while bond orders and atom types are conserved as shown in Fig. 1. Exocyclic bonds of carbonyls, imines, sulfones, and sulfoxides were considered as part of the cyclic system if they are directly bonded to a ring, as well as all functional groups forming linkers between two rings. Although cyclic system resolution was considered in this work, different resolutions or any other definition of molecular scaffold can be used as long as they are generated in a consistent manner.

### Molecular similarity

Molecular similarities were calculated using eleven 2D fingerprints implemented in MayaChemTools [33] and two 3D shape methods available in Rapid Overlay of Chemical Structures (ROCS) [34]. For 2D and 3D representations, the Tanimoto coefficient was used to calculate  $N(N - 1) / 2$  pairwise structural similarities ( $SS_{ij}$ ) for each pair of molecules  $i$  and  $j$  [35,36]. The 2D fingerprints namely atom neighborhoods [37], atom types, electrotopological state indices (EstateIndices) [38], extended connectivity (ECFP4) [39], MACCS (322 bits) [40], path length, topological atom pairs (TopAtomPairs) [41], topological atom torsions (TopAtomTorsions) [42], topological atom triplets (TopAtomTriplets), topological pharmacophore atom pairs (TopPh4AtomPairs) [43], and topological pharmacophore atom triplets (TopPh4AtomTriplets) [44] implemented in a set of Perl scripts available in MayaChemTools were used. To apply 3D similarity methods, the structures were opti-



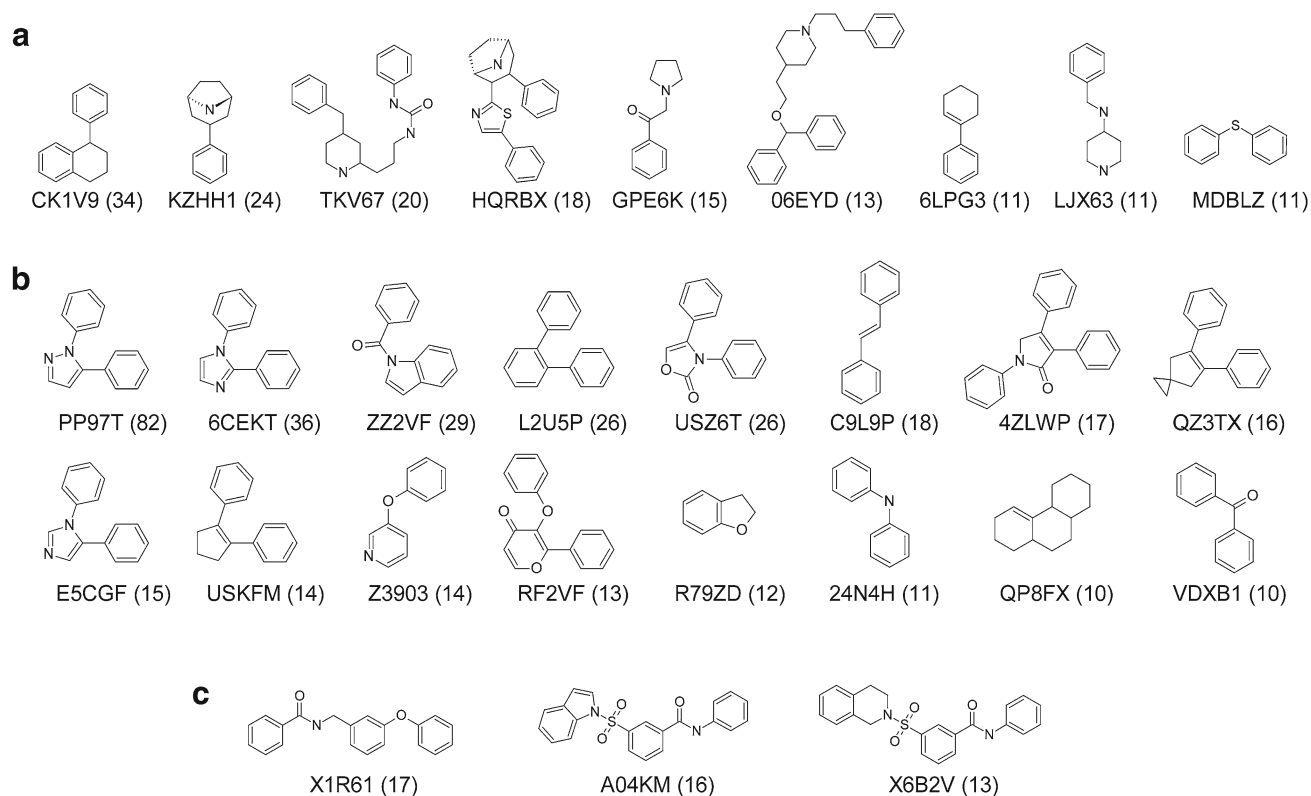
**Fig. 2** General form of the structure–activity similarity (SAS) maps showing four major regions. Regions I and II are associated with scaffold hopping and smooth SAR, respectively. Region IV indicates discontinuous SAR and activity cliffs

mized to obtain a single low-energy conformation using a modified Dreiding force field implemented in Vconf software [45,46]. Although the use of multiple conformations is desirable for activity landscape studies [47], we use a single low-energy conformation to calculate 3D similarity representations as an approximation to simplify this work by introducing the less variables as possible. 3D representations included Tanimoto Shape Index (ShapeTanimoto) and ComboScore/2 (mean value of ShapeTanimoto and Color Score) [48,49]. Despite the inherent conformational issues, the use of 3D structural representations is valuable in activity landscape studies [50].

Since structural similarities have a strong dependence with the molecular representation used [50], three uncorrelated structure similarity representations namely MACCS keys, TopPh4AtomPairs, and ComboScore/2 were combined in a mean similarity measure, taking into account the principles of data fusion [36,51]. Similar criteria to select fingerprint representations to obtain consensus models of activity landscape have been used in several studies and extensively discussed [52].

### Activity differences

For each dataset of  $N$  compounds tested against target  $T$ ,  $N(N - 1) / 2$  absolute value of pairwise activity differences corresponding to each possible pair of compounds in the dataset was calculated from the following equation [21,53]:



**Fig. 3** Most frequent cyclic systems found in the analyzed databases (frequency  $\geq 10$ ). Chemotype identifier and frequency are displayed. **a** Compounds screened against monoamine transporters. **b** Cyclooxygenase inhibitors. **c** PPAR ligands

$$|\Delta A(T)_{i,j}| = |A(T)_i - A(T)_j| \quad (1)$$

where  $A(T)_i$  and  $A(T)_j$  are the activities of the  $i$ th and  $j$ th molecules ( $j > i$ ), in  $\text{pIC}_{50}$  or  $\text{pK}_i$  values, tested against the target  $T$ . In this work,  $T = \text{COX-1}$ ,  $\text{COX-2}$ ,  $\text{NET}$ ,  $\text{SERT}$ ,  $\text{DAT}$ ,  $\text{PPAR}\alpha$ ,  $\text{PPAR}\delta$ , and  $\text{PPAR}\gamma$ .

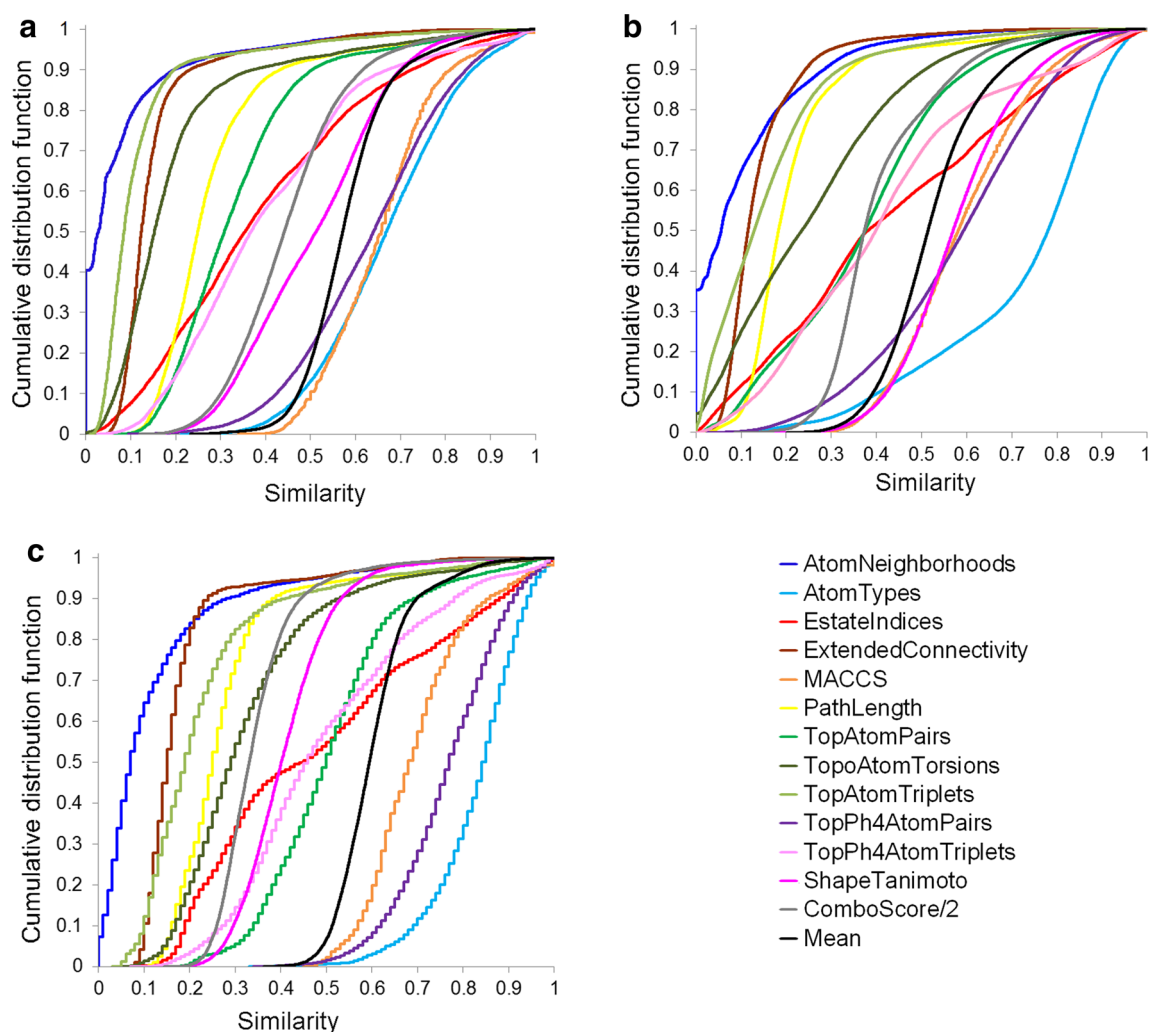
### Activity landscape with SAS maps

SAS maps were generated by plotting the structural similarity in the X-axis against the absolute value of the activity difference in the Y-axis for each pair of compounds; thus, for a set of  $N$  compounds tested against target  $T$ , each SAS map contains  $N(N-1)/2$  data points [13]. As is shown in Fig. 2, SAS maps can be roughly divided in four zones (I–IV) by imposing activity difference and molecular similarity threshold values to aid their interpretation [13]. The X-axis in SAS maps was divided using, as a heuristic threshold, the lowest similarity value observed in the top 10 % compounds with higher similarity. The activity difference axis (Y-axis) threshold was set to a value of 1 log unit to define boundaries between regions I/III and II/IV. Other criteria to define structure similarity and activity difference thresholds can be employed as reviewed by Medina-Franco [52]. Points

that fall in each region can be associated with different SAR, briefly, data points that fall in region I are associated with scaffold hops having low structure similarity and low activity differences [52]. Points in region II are associated with compounds with smooth SAR having high structure similarity and low activity differences [52]. Region III corresponds to low structure similarity and high activity differences. Region IV contains molecular pairs that correspond to activity cliffs having high structure similarity and high activity difference [1,3,52].

### SAS maps defined with chemotype information

All compound pairs in each database were distinguished depending on whether both molecules in the pair share the same chemotype or not. Data points where both compounds contain the same chemotype were further differentiated in the SAS maps using different colors. In this work, only chemotypes containing at least ten molecules were analyzed. Each chemotype  $\lambda$  comprises  $N_\lambda(N_\lambda - 1)/2$  pairwise comparisons of structure similarity and activity differences, where  $N_\lambda(N_\lambda \geq 10)$  is the number of compounds classified in chemotype  $\lambda$ .



**Fig. 4** Cumulative distribution functions of structural similarities using different 2D and 3D fingerprint representations. **a** 299 compounds screened against monoamine transporters (44,551 structural

similarities), **b** 658 compounds with activity against COX-1 and COX-2 (216,153 structural similarities), **c** 168 compounds screened against PPAR  $\alpha$ ,  $\delta$ , and  $\gamma$  (14,028 structural similarities)

### Activity cliff enrichment factor

For each chemotype  $\lambda$ , activity cliff enrichment factor (ACEF) was calculated using the following equation:

$$\text{ACEF}(\lambda) = \frac{\text{AC}(x_\lambda)}{\text{AC}(x_\chi)} \quad (2)$$

where  $\text{ACEF}(\lambda)$  is the activity cliff enrichment factor for the  $\lambda$ th chemotype, which relates the fraction of activity cliffs formed with molecules that fall in the  $\lambda$ th chemotype  $\text{AC}(x_\lambda)$  with respect to the fraction of activity cliffs in the entire database  $\text{AC}(x_\chi)$ .

$\text{AC}(x_\lambda)$  was calculated as follows:

$$\text{AC}(x_\lambda) = \frac{|C_\lambda^{IV}|}{|C_\lambda|} \quad (3)$$

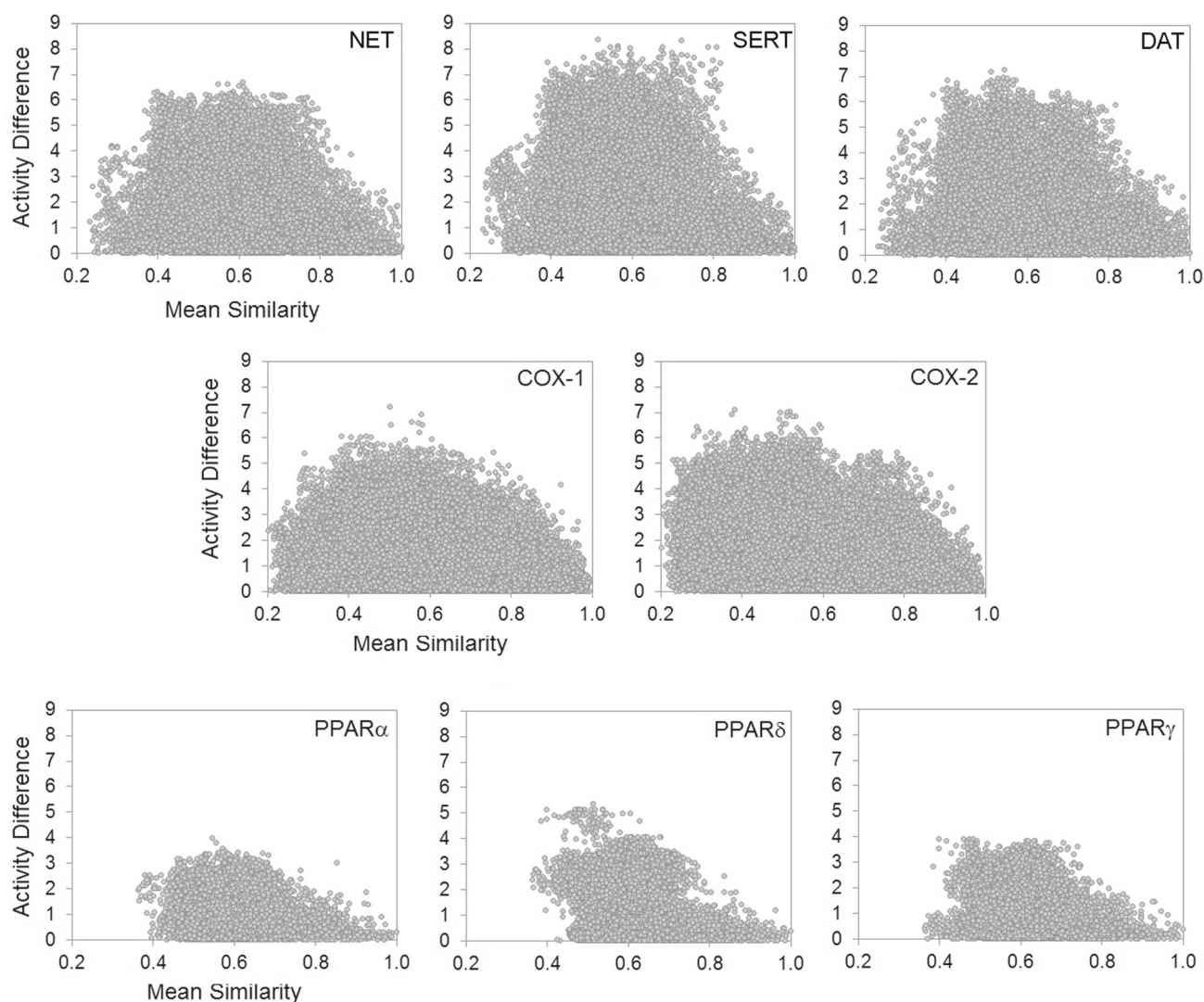
where  $|C_\lambda^{IV}|$  is the number of activity cliffs (pairs of compounds that fall in region IV in the SAS maps) in the chemotype class  $\lambda$ , and  $|C_\lambda|$  is the total number of pairs of compounds in the same chemotype class ( $\lambda$ ).

$\text{AC}(x_\chi)$  was calculated as follows:

$$\text{AC}(x_\chi) = \frac{|C_\chi^{IV}|}{|C_\chi|} \quad (4)$$

where  $|C_\chi^{IV}|$  is the total number of activity cliffs, and  $|C_\chi|$  is the total number of pairs in the database. Noteworthy,  $\text{AC}(x_\chi)$  is fraction of the total number of activity cliffs (as defined in the SAS maps) that does not depend on the chemotype classification. Similar indices were previously reported to identify chemotypes enriched with active and selective compounds [23,54].





**Fig. 5** SAS maps for databases analyzed in this study. Each data point indicates a pairwise comparison of 299 compounds (44,551 datapoints) screened against monoamine transporters (NET, SERT, and DAT), 658 compounds (216,153 datapoints) with activity against cyclooxygenases

(COX-1 and COX-2) and 168 compounds (14,028 datapoints) screened against PPARs ( $\alpha$ ,  $\delta$ , and  $\gamma$ ). Mean similarity was calculated using three selected similarity representations in all cases (MACCS keys, TopPh4AtomPairs and ComboScore/2)

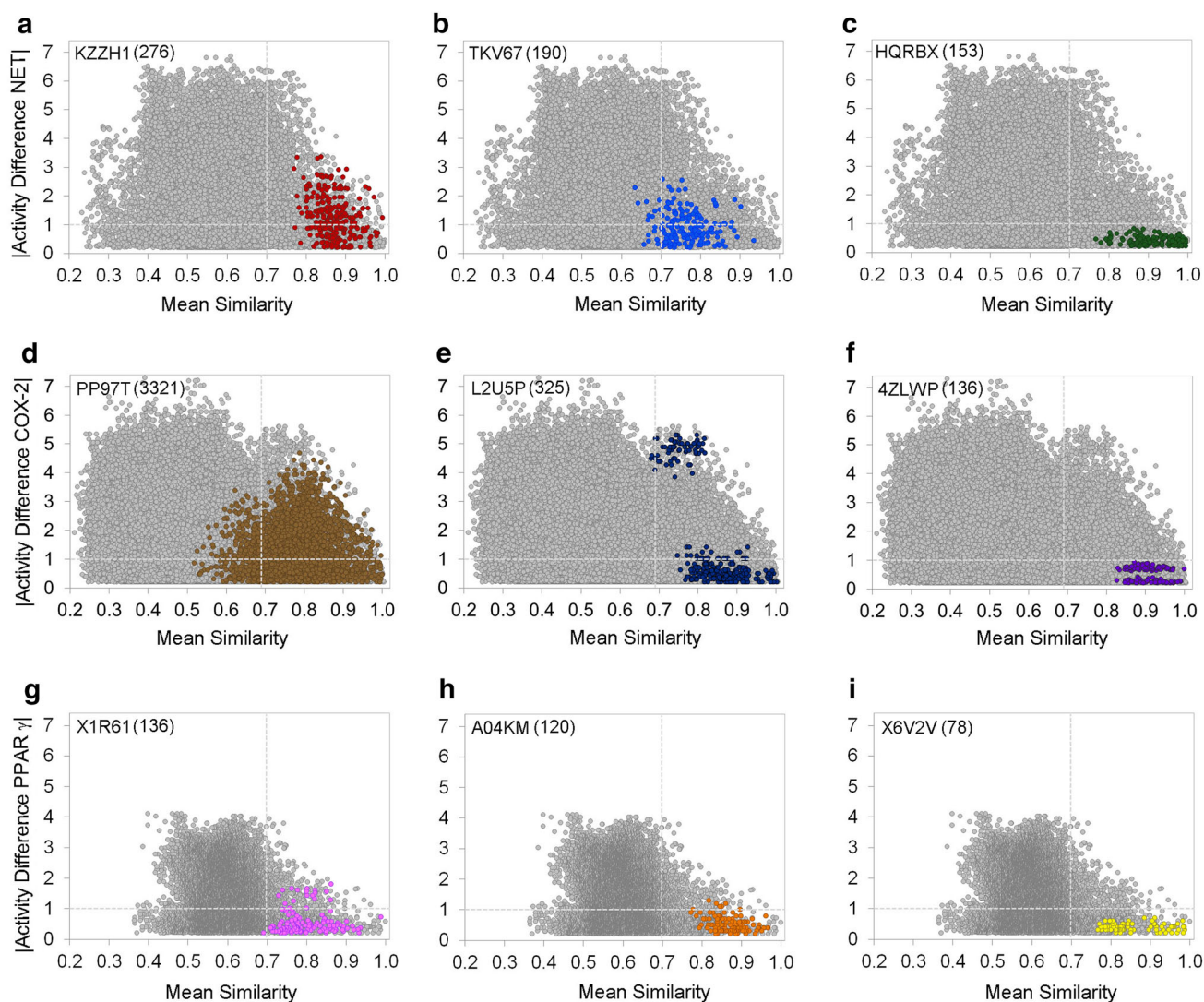
### Activity cliff generators based on chemotypes

We computed the number of activity cliffs with chemotype  $\lambda$  ( $NAC_{x\lambda}$ ) as the frequency of each compound ( $x$ ) in points that fall in region IV of SAS maps if the corresponding chemotype frequency  $\lambda(CF_\lambda)$  is equal or greater than 10.  $NAC_{x\lambda}$  was scaled by the chemotype frequency using the expression  $(NAC_{x\lambda}/CF_\lambda)$ . Compounds with  $NAC_{x\lambda}/CF_\lambda \geq 2$  standard deviations relative to the corresponding mean of the distribution of the total  $NAC_{x\lambda}/CF_\lambda$  values were considered as activity cliff generators.

### Results and discussion

#### Chemotype classification

Molecules in each database were classified in cyclic systems. Figure 3 shows the most common cyclic systems in each database. In this work, only cyclic systems found in at least ten molecules (frequency  $\geq 10$ ) were considered as relevant, since they correspond to the most common molecular scaffolds. For compounds screened against monoamine transporters, 91 cyclic systems were identified of which nine



**Fig. 6** SAS maps including chemotype information for databases analyzed in this study. Each data point indicates a pairwise comparison of 299 compounds (44,551 datapoints) screened against norepinephrine transporter (a–c), 658 compounds (216,153 datapoints) with activity against cyclooxygenase-2 (d–Medina-Franco), and 168 compounds (14,028 datapoints) screened against PPAR $\gamma$  (g–i). Representative

chemotypes were highlighted in color code represented by  $N_\lambda(N_\lambda - 1)/2$  pairwise comparisons of structure similarity and potency differences where  $N_\lambda$  is the number of compounds classified in the chemotype  $\lambda$ . Also, chemotype code and the number of data points where both compounds fall in a particular chemotype are indicated at the top of each map

of them have high frequency ( $\geq 10$ ). Also, 191 cyclic systems were found for cyclooxygenase inhibitors, and 16 of them have high frequency. Furthermore, PPAR ligands database contains 66 cyclic systems of which only three have high frequency.

### Molecular similarity

For each database, pairwise structural similarities were calculated using 13 structural representations. This resulted in 44,551 pairwise structural similarities for 299 compounds tested against monoamine transporters, 216,153 for 658

cyclooxygenase inhibitors, and 14,028 for 168 PPAR ligands. Cumulative distribution functions for each representation and for the mean similarity are shown in Fig. 4. Results show that most of the 2D and 3D representations have different ranges of similarity values. This is consistent with previous studies where different ranges of similarity values are identified for molecular fingerprints obtained from different design [13,20,45]. The statistics of structural similarities are summarized in Table S2 in the Supporting Information indicating the maximum, third and first quartile, median, mean, and standard deviation. Results indicate that Atom neighborhoods, ExtendedConnectivity, and TopAtomTriplets have

the lowest similarity values for all databases analyzed. Also, median values are close to mean values in all the cases. Orthogonal representations were selected taking into account the correlation of pairwise similarities between different descriptors and their ability to capture molecular features by different methodologies. The correlation matrix for different 2D and 3D representations for each database can be found in the Supporting Information (Tables S3–S5). Three low correlated representations namely MACCS keys, TopPh4AtomPairs, and ComboScore/2 were selected to compute mean similarity. Noteworthy, all selected representations show low correlation between them in all databases studied (correlation < 0.6). MACCS and TopPh4AtomPairs have correlations of 0.25, 0.31, and 0.53 for MATs, COXs, and PPARs databases, respectively. Also, MACCS and ComboScore/2 have correlations of 0.28, 0.50, and 0.38, for the same databases, respectively. Finally, TopPh4AtomPairs and ComboScore/2 have correlations of 0.33, 0.32, and 0.30, respectively. It is worth mentioning that the selected structure representations are able to capture different molecular information since they have different design. The representation MACCS is based on substructures, TopPh4AtomPairs is based on atom pairs, and ComboScore/2 is based on 3D molecular shape and color features (related to 3D pharmacophoric elements) [40,43,48,49].

### Activity landscape with SAS maps

Figure 5 shows a general overview of the activity landscape with SAS maps for all databases analyzed. It is worth noting the different point distributions in the databases under study. SAS maps for monoamine transporters and cyclooxygenases are characterized by compound pairs that cover a wide range of mean molecular similarity and potency differences. In other words, these databases are characterized by structurally diverse compounds, which cover a wide potency range (see Tables S6–S7 in the Supporting Information). In contrast, SAS maps for PPARs are characterized by the lower activity differences and the lowest structure similarity coverage. Therefore, this database is characterized by structurally similar compounds and low potency variations as compared with compounds tested against cyclooxygenases and monoamine transporters. This is an indication that a smaller part of the activity landscape is being captured by this database, and it has the most continuous SAR among the databases included in this study.

### Mapping chemotypes in SAS maps

Although SAS maps give a general overview of the activity landscape, extract specific SAR information from these representations is not straightforward; therefore, additional information can be mapped in SAS maps to simplify their

**Table 1** Quantitative characterization of molecular pairs in regions I–IV for monoamine transporters database

T-10 % <sup>a</sup> (0.702)	Total number of pairs	I			II			III			IV Activity cliffs		
		NET	SERT	DAT	NET	SERT	DAT	NET	SERT	DAT	NET	SERT	DAT
	44551	15248	12026	14862	2520	2429	2314	24811	28033	25197	1972	2063	2178
CK1V9	561	18	13	16	441	372	352	6	11	8	96	165	185
KZHH1	276	0	0	0	138	111	188	0	0	0	138	165	88
TKV67	190	9	12	10	129	110	130	8	5	7	44	63	43
HQRBX	153	0	0	0	153	150	121	0	0	0	0	3	32
GPE6K	105	1	2	2	84	88	89	2	1	1	18	14	13
06EYD	78	0	0	0	78	74	67	0	0	0	0	4	11
6LPG3	55	5	1	4	36	39	26	1	5	2	13	10	23
LJX63	55	0	0	0	30	51	55	0	0	0	25	4	0
MDBLZ	55	0	0	0	27	25	51	0	0	0	28	30	4

<sup>a</sup> Threshold based on molecular pairs with higher molecular similarity (top 10 %)



**Table 2** ACEF for all databases studied using a threshold of 1 in activity differences and 0.702, 0.681, and 0.698 in structure similarity for MATs, COXs, and PPARs databases, respectively

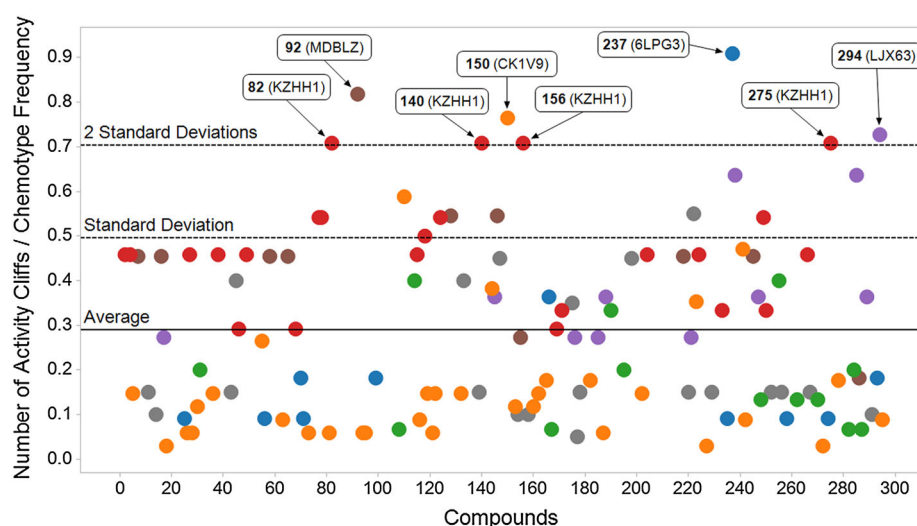
Chemotype	Frequency (pairs)	ACEF NET	ACEF SERT	ACEF DAT
<i>Monoamine transporters database</i>				
Database	299 (44551)	1.00	1.00	1.00
CK1V9	34 (561)	3.87	6.35	6.75
KZHH1	24 (276)	11.30	12.91	6.52
TKV67	20 (190)	5.23	7.16	4.63
HQRBX	18 (153)	0.00	0.42	4.28
GPE6K	15 (105)	3.87	2.88	2.53
06EYD	13 (78)	0.00	1.11	2.88
6LPG3	11 (55)	5.34	3.93	8.55
LJX63	11 (55)	10.27	1.57	0.00
MDBLZ	11 (55)	11.50	11.78	1.49
Chemotype	Frequency (pairs)	ACEF COX-1	ACEF COX-2	
<i>Cyclooxygenases database</i>				
Database	658 (216153)	1.00	1.00	
PP97T	82 (3321)	10.66	10.73	
6CEKT	36 (630)	7.50	7.92	
ZZ2VF	29 (406)	8.59	6.42	
L2U5P	26 (325)	2.98	5.09	
USZ6T	325 (26)	2.01	1.26	
C9L9P	18 (153)	6.33	5.78	
4ZLWP	17 (136)	0.00	0.00	
QZ3TX	16 (120)	7.87	4.53	
E5CGF	15 (105)	9.69	0.65	
USKFM	14 (91)	5.33	5.73	
Z3903	14 (91)	2.93	8.47	
RF2VF	13 (78)	1.55	0.00	
R79ZD	12 (66)	8.81	10.65	
24N4H	11 (55)	10.13	7.01	
QP8FX	10 (45)	1.62	5.04	
VDXB1	10 (45)	4.31	3.02	
Chemotype	Frequency (pairs)	ACEF PPAR $\alpha$	ACEF PPAR $\delta$	ACEF PPAR $\gamma$
<i>PPAR ligands database</i>				
Database	168 (14028)	1.00	1.00	1.00
X1R61	17 (136)	9.24	6.85	4.65
A04KM	16 (120)	0.42	4.44	1.05
X6B2V	13 (78)	0.00	0.00	0.00

interpretation such as chemotype information. SAS maps including chemotype information visually depicts the local activity landscape of subsets of compounds sharing the same chemotype. Using this approach, it is possible to analyze the distribution of the pairs of compounds for each frequent cyclic system and then identify the cyclic systems with the highest fraction of activity cliffs. Also, is of interest the detection of the most common cyclic systems present in compounds that frequently form activity cliffs; this is activity

cliff generators. While detection of chemotypes enriched in activity cliffs gives a general overview of local activity landscape based on chemotypes, i.e., continuous or discontinuous chemotype landscape, the detection of activity cliff generators gives important advantage to highlight small structural changes on substituents that a specific scaffold should have to exhibit high activity differences.

Figure 6 shows some examples of SAS maps highlighting with a color code those points where both molecules in the

**Fig. 7** Number of activity cliffs formed by each compound divided by the chemotype frequency for NET database; only chemotypes with frequency  $\geq 10$  were considered in this analysis. Compounds with the fraction [number of activity cliffs]/[chemotype frequency] higher than two standard deviations of the average are considered activity cliff generators. Different chemotypes are labeled by the corresponding MEQNUM and are represented with different colors



pair share the same chemotype. As described in the Methods, SAS maps in Fig. 6 were divided in four major regions (I–IV) using a threshold value of 1 for activity difference and the top 10 % pairs with the highest structure similarity for each database. Only the SAS maps for NE transporter, COX-2, and PPAR $\gamma$  are shown in Fig. 6; however, other biological activities and databases can be analyzed similarly.

Different distributions in regions I–IV were observed in SAS maps for each chemotype class showed in Fig. 6. Note that, in general, most of the data points are in region II. This is not surprising since molecules sharing the same scaffold will have generally high mean structure similarity (see Table S7 in the Supporting Information). However, some chemotypes have different distributions in regions I–IV as illustrated in Fig. 6. Figure 6a–c show local chemotype activity landscapes for compounds that fall in three chemotypes for the NET dataset. In this case, pairs that fall in chemotype KZZH1 are distributed between regions II and IV (Table 1). Noteworthy, some pairs in region IV for KZZH1 present high activity differences, in some cases more than two logarithmic units; hence, this chemotype is characterized by a heterogeneous SAR and the presence of activity cliffs for the current database (Table S6). Chemotype TKV67 shows pairs distributed in regions I–IV, although this observation is highly dependent of the structure similarity threshold used as reference, it is clear the overall lower structure similarity values for this chemotype (mean = 0.765 and median = 0.759) as compared with chemotype KZZH1 (mean = 0.870 and median = 0.865), see Table S7 in the Supporting Information. This last observation can be rationalized in function of the structural diversity of side chains of compounds with cyclic system TKV67 as compared with KZZH1. In contrast, chemotype HQRBX, where all molecular pairs fall in region II, is characterized by a flat SAR having molecules with high structure similarity (mean = 0.914 and median

= 0.918) and low activity differences (mean = 0.204 and median = 0.173). Interesting results were also found for COX-2 inhibitors; Fig. 6d shows that chemotype PP97T covers a broad area of the ‘landscape space.’ Also, chemotype L2U5P (Fig. 6e) shows an interesting pattern where molecular pairs were mainly found in two zones including pairs having low and very high activity differences. For the same database, 4ZLWP (Fig. 6f) is an example of chemotype characterized by a flat SAR. On the other hand, only chemotypes X1R61, A04KM, and X6V2V have high frequency for PPAR $\gamma$  ligands. The SAS maps including these chemotypes are shown in Fig. 6g–i. Data points for chemotypes X1R61 and A04KM are mainly located at region II and IV, whereas pairs with chemotype X6V2V fall only in region II. It is worth mentioning that pairs in region IV for chemotypes X1R61 and A04KM have activity difference values lower than 2 units; therefore, changes in structure have only low or moderate impact in potency for this database.

The number of molecule pairs that fall in each region of SAS maps was calculated for all databases. The results for monoamine transporters database are shown in Table 1; whereas the same data for COXs and PPARs databases can be found in Tables S8–S9 of the Supporting Information. As can be deduced from Table 1, quantitative comparison of each chemotype in SAS maps is meaningful only when comparing subsets with the same chemotype across regions I–IV, i.e., 50 % of the pairs with chemotype KZZH1 fall in region IV for NET database; therefore, activity cliffs having this chemotype are frequent. Also, the quantitative analysis in Table 1 enables the cross-comparison of molecules sharing the same scaffold and tested against multiple targets employing the same thresholds in activity and molecular similarity. For example, molecule pairs with chemotype HQRBX have a total flat SAR in NET landscape, whereas some activity cliffs (32 pairs with activity differences  $\geq 1$ ) can be found

**Table 3** Activity cliff generators for MATs, COXs, and PPARs databases

Chemotype	Compound	Act. cliffs	Factor ( $NAC_{x\lambda}/CF_{\lambda}$ )
NET (threshold = mean [ $NAC_{x\lambda}/CF_{\lambda}$ ] + 2STD = 0.7024)			
KZHH1	<b>82</b>	17	0.7083
KZHH1	<b>140</b>	17	0.7083
KZHH1	<b>156</b>	17	0.7083
KZHH1	<b>275</b>	17	0.7083
6LPG3	<b>237</b>	10	0.9091
MDBLZ	<b>92</b>	9	0.8182
CK1V9	<b>150</b>	26	0.7647
LJX63	<b>294</b>	8	0.7273
SERT (threshold = 0.7447)			
KZHH1	<b>233</b>	20	0.8333
KZHH1	<b>250</b>	19	0.7917
KZHH1	<b>68</b>	19	0.7917
KZHH1	<b>204</b>	18	0.7500
TKV67	<b>198</b>	16	0.8000
TKV67	<b>267</b>	15	0.7500
CK1V9	<b>162</b>	27	0.7941
DAT (threshold = 0.6168)			
CK1V9	<b>110</b>	26	0.7647
CK1V9	<b>223</b>	23	0.6765
CK1V9	<b>116</b>	22	0.6471
KZHH1	<b>115</b>	21	0.8750
KZHH1	<b>82</b>	17	0.7083
6LPG3	<b>237</b>	10	0.9091
COX-1 (threshold = 0.7055)			
PP97T	<b>141</b>	71	0.8659
PP97T	<b>144</b>	70	0.8537
PP97T	<b>123</b>	63	0.7683
PP97T	<b>142</b>	63	0.7683
PP97T	<b>483</b>	61	0.7439
PP97T	<b>504</b>	61	0.7439
PP97T	<b>485</b>	60	0.7317
PP97T	<b>488</b>	60	0.7317
PP97T	<b>128</b>	59	0.7195
PP97T	<b>112</b>	58	0.7073
PP97T	<b>116</b>	58	0.7073
PP97T	<b>151</b>	58	0.7073
E5CGF	<b>460</b>	12	0.8000
E5CGF	<b>450</b>	11	0.7333
QZ3TX	<b>56</b>	15	0.9375
6CEKT	<b>184</b>	28	0.7778
R79ZD	<b>195</b>	9	0.7500
ZZ2VF	<b>436</b>	21	0.7241
COX-2 (threshold = 0.7146)			
PP97T	<b>489</b>	69	0.8415
PP97T	<b>128</b>	63	0.7683

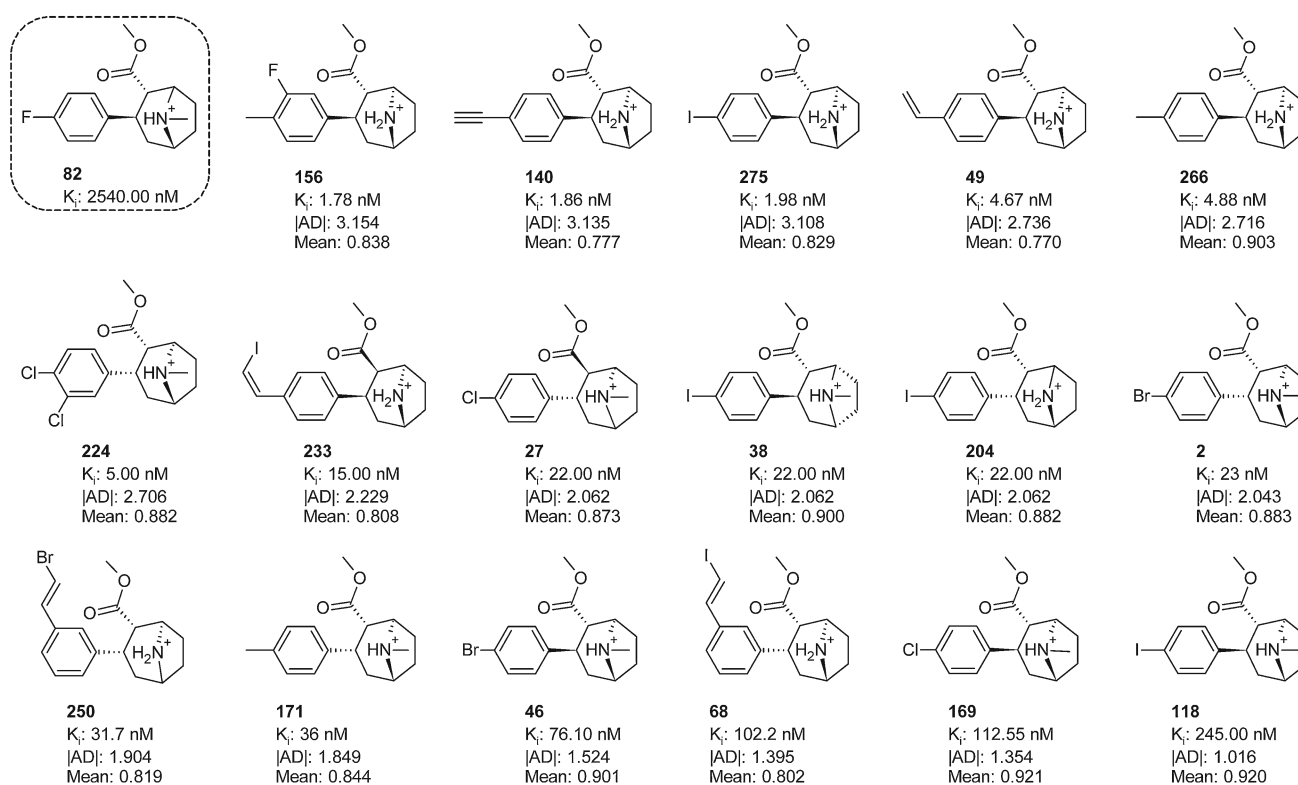
**Table 3** continued

Chemotype	Compound	Act. cliffs	Factor ( $NAC_{x\lambda}/CF_{\lambda}$ )
PP97T	<b>119</b>	62	0.7561
PP97T	<b>485</b>	62	0.7561
PP97T	<b>97</b>	59	0.7195
L2U5P	<b>257</b>	23	0.8846
L2U5P	<b>255</b>	22	0.8462
L2U5P	<b>256</b>	20	0.7692
6CEKT	<b>186</b>	29	0.8056
6CEKT	<b>167</b>	28	0.7778
QZ3TX	<b>52</b>	14	0.8750
PPAR $\delta$ (threshold = 0.5072)			
X1R61	<b>14</b>	11	0.6471
X1R61	<b>135</b>	10	0.5882
A04KM	<b>144</b>	9	0.5625
PPAR $\gamma$ (threshold = 0.4742)			
X1R61	<b>135</b>	15	0.8824

for the same chemotype when the molecules were evaluated against DAT. However, since the number of pairs in SAS maps for each chemotype is related to chemotype frequency, the quantitative characterization by the number of molecular pairs that fall regions I–IV cannot be used for comparisons between different chemotypes.

#### Activity cliff enrichment factor for chemotype classes

The concept of activity cliff enrichment factor (*ACEF*) is introduced in this work to normalize the number of activity cliffs in each chemotype taking as a reference the entire database. This factor makes possible to compare the relative number of activity cliffs between chemotype classes for the most important chemotypes in the databases, this is, chemotypes with a considerable frequency useful to extract SAR information. The purpose of ACEF is to identify cyclic systems enriched in activity cliffs for the current databases. Calculated ACEF values for the most frequent chemotypes (frequency  $\geq 10$ ) are shown in Table 2. Note that ACEF for the entire database has a value of one in all databases; therefore, chemotypes with values higher than one are rich in activity cliffs as compared with the entire database. Similarly, chemotypes with value lower than one are poor in activity cliffs, thus representing a smoother landscape as compared to those scaffolds with ACEF  $> 1$ . Chemotype KZHH1 has high ACEF against the three monoamine transporters ( $\geq 6.52$ ). In contrast HQRBX has only high ACEF for DAT (4.28). These measures of ACEF for individual chemotypes along with the visual representations of the SAR in the SAS maps (Fig. 6a, c) can be interpreted in function of SAR discontinuity. Chemotype KZHH1 is characterized by a heterogeneous



**Fig. 8** Chemical structure of compounds that form activity cliffs with the activity cliff generator **82** for NET database. Each compound is labeled with their identification number, biological activity  $K_i$ (nM), activity difference ( $|AD|$ ), and mean similarity

SAR for all monoamine transporters, whereas HQBRX has a flat SAR for NET and SERT. Interesting chemotypes for COX inhibitors with contrasting ACEF values are PP97T and 4ZLWP: while PP97T shows the highest ACEF against both cyclooxygenases ( $\geq 10.66$ ), 4ZLWP is characterized by a flat SAR for both targets. Similar results were found for the PPARs dataset. In this case, X1R61 is the chemotype with the highest ACEF ( $\geq 4.65$ ), whereas X6B2V has a totally flat SAR for the three studied targets.

The ACEF measure can provide useful information when choosing a scaffold for lead optimization. For example, compounds containing cyclic systems with high ACEF are interesting, as it is more likely to have large changes in activity with few structural modifications. On the other hand, compounds having cyclic systems with a low ACEF can be useful to develop predictive models, i.e., QSAR.

### Chemotype-based activity cliff generators

In addition to the characterization of chemotypes enriched with activity cliffs, it is of interest the detection of chemotypes associated with activity cliff generators. It is worth noting that characterization of activity cliff generators in this work was carried employing a slight modification of the original definition reported by Méndez-Lucio et al.

[13]. In this case, we computed the fraction: [number of activity cliffs]/[chemotype frequency] in order to compare results between compounds classified by chemotype classes with different frequencies (see Methods). For NET database (Fig. 7), some compounds classified in chemotype KZHH1 tend to form high number of activity cliffs, e.g., compounds **82**, **140**, **156**, and **275**. Note that this same chemotype also has high ACEF for NET database (Table 2). Similar figures highlighting chemotypes with activity cliff generators for SERT, DAT, COXs, and PPARs can be found in the Supporting Information as Fig. S1–S7. Additionally, Table 3 shows chemotypes that have at least one activity cliff generator, the number of compounds associated with each generator, and the relation  $NAC_{x,\lambda}/CF_\lambda$  (factor). Table 3 also shows that chemotypes KZHH1 and CK1V9 comprise most of the activity cliff generators for MATs. Similarly, chemotype PP97T is rich in activity cliff generators for COXs databases. Furthermore, for COX-2 database, chemotype L2U5P is associated with three activity cliff generators. This result is in agreement with the SAS map presented in Fig. 6e, where several molecular pairs form deep activity cliffs (activity difference  $\geq 2$  log units). For PPARs database, a low number of activity cliff generators were identified. The most important chemotype identified is X1R61, which is present in two activity cliff generators for PPAR $\delta$  and in one generator for PPAR $\gamma$ .



Figure 8 shows the chemical structure of a representative activity cliff generator (**82**) with chemotype KZHH1 and molecules that form activity cliffs with this compound for the NET set. All compounds in this figure are arranged in decreasing order of activity difference. Note that in deeper activity cliffs (**82\_156**, **82\_140**, **82\_275**, **82\_49** and **82\_266**), the most active compounds have specific stereochemistry (1*R*,2*S*,3*R*,5*R*) in the 8-azabicyclo[3.2.1]octane nucleus. Changes in this stereochemistry lead to a decrement in activity (compare compounds **204** and **275**). Also, *N*-methylation appears to be unfavorable as is shown by the ‘shallow cliff’ **82\_118** ( $|AD| = 1.016$ ) in comparison with the ‘deep cliff’ **82\_275** ( $|AD| = 3.108$ ) [55]. Interestingly, some subtypes of activity cliffs (based on the classification proposed by Hu and Bajorath) can be observed in Fig. 8, i.e., **82\_156** and **82\_140** are examples of topology and R-group cliffs, respectively [14]. Additional conclusions can be generated by the analysis of different activity cliff generators, chemotype classes, and biological activities.

## Conclusions and perspectives

In this work activity landscape representations employing SAS maps annotated with chemotype information were generated for three structural diverse databases with multiple biological activities. The addition of chemotype information in these representations gave a general overview of the SAR nature related to each cyclic system for the databases analyzed. In addition, it makes feasible to easily quantify activity cliffs related to each chemotype. Abundance of activity cliffs in each chemotype was compared between chemotypes and entire databases employing a new index introduced in this work named activity cliff enrichment factor. Using this approach, some chemotypes were identified as rich in activity cliffs for MATs (i.e., KZHH1), COXs (i.e., PP97T), and PPARs (i.e., X1R61) databases. The ACEF values give information concerning to chemotypes with high and low SAR discontinuity and can provide useful information when choosing a scaffold for lead optimization and predictive studies; whereas discontinuous landscapes are rich in SAR information (i.e., activity cliffs), continuous landscapes can be especially useful in predictive approaches (i.e., QSAR). Also, the activity cliff generator concept was applied to compounds grouped in each chemotype to highlight structural features that lead to high changes in activity. Some chemotypes were identified as cyclic systems that usually form activity cliff generators for MATs (i.e., KZHH1 and CK1V9), COXs (i.e., PP97T), and PPARs (i.e., X1R61) databases. The SAR analysis of activity cliffs leads to specific SAR conclusions for the datasets studied. In this context, the inclusion of chemotype information in activity landscape representations

aids the SAR interpretation and characterization of diverse molecular databases.

**Acknowledgments** The authors would like to express their sincere gratitude to the BindingDB team for providing the studied structure and activity data; to Dr. Mark Johnson for providing the program MEQI; to MayaChemTools for providing the scripts for fingerprint calculations; to VeraChem LLC for providing VConf; to OpenEye Scientific Software, Inc., for providing ROCS (UAM); and to Tableau Software for providing Tableau Public. O. M-L is very grateful to CONACyT (No. 217442/312933) and the Cambridge Overseas Trust for funding. J.L. M-F thanks the National Autonomous University of Mexico (UNAM), grant PAIP 5000-9163, for funding.

## References

- Maggiora GM (2006) On outliers and activity cliffs: why QSAR often disappoints. *J Chem Inf Model* 46:1535. doi:10.1021/ci060117s
- Cruz-Monteaugudo M, Medina-Franco JL, Pérez-Castillo Y, Nicolotti O, Cordeiro MNDS, Borges F (2014) Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? *Drug Discov Today* 19:1069–1080. doi:10.1016/j.drudis.2014.02.003
- Stumpfe D, Bajorath J (2012) Exploring activity cliffs in medicinal chemistry. *J Med Chem* 55:2932–2942. doi:10.1021/jm201706b
- Pérez-Villanueva J, Medina-Franco JL, Caulfield TR, Hernández-Campos A, Hernández-Luis F, Yépez-Mulia L, Castillo R (2011) Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) of some benzimidazole derivatives with trichomonocidal activity. *Eur J Med Chem* 46:3499–3508. doi:10.1016/j.ejmech.2011.05.016
- Hernández-Vázquez E, Méndez-Lucio O, Hernández-Luis F (2013) Activity landscape analysis, CoMFA and CoMSIA studies of pyrazole CB1 antagonists. *Med Chem Res* 22:4133–4145. doi:10.1007/s00044-012-0418-y
- Iyer P, Wawer M, Bajorath J (2011) Comparison of two- and three-dimensional activity landscape representations for different compound data sets. *Med Chem Commun* 2:113–118. doi:10.1039/c0md00188k
- Shanmugasundaram V, Maggiora GM (2001) Characterizing property and activity landscapes using an information-theoretic approach. Cinf-032. In: 222nd ACS national meeting, Chicago. American Chemical Society, Washington, D. C
- Wawer M, Peltason L, Weskamp N, Teckentrup A, Bajorath J (2008) Structure–activity relationship anatomy by network-like similarity graphs and local structure–activity relationship indices. *J Med Chem* 51:6075–6084. doi:10.1021/jm800867g
- Guha R, Van Drie JH (2008) Structure–activity landscape index: identifying and quantifying activity cliffs. *J Chem Inf Model* 48:646–658. doi:10.1021/ci7004093
- Peltason L, Bajorath J (2007) SAR index: quantifying the nature of structure–activity relationships. *J Med Chem* 50:5571–5578. doi:10.1021/jm0705713
- Bajorath J, Peltason L, Wawer M, Guha R, Lajiness MS, Van Drie JH (2009) Navigating structure–activity landscapes. *Drug Discov Today* 14:698–705. doi:10.1016/j.drudis.2009.04.003
- Kayastha S, Dimova D, Iyer P, Vogt M, Bajorath J (2014) Large-scale assessment of activity landscape feature probabilities of bioactive compounds. *J Chem Inf Model* 54:442–450. doi:10.1021/ci400677b
- Méndez-Lucio O, Pérez-Villanueva J, Castillo R, Medina-Franco JL (2012) Identifying activity cliff generators of PPAR lig-

- ands using SAS maps. *Mol Inf* 31:837–846. doi:10.1002/minf.201200078
14. Hu Y, Bajorath J (2012) Extending the activity cliff concept: structural categorization of activity cliffs and systematic identification of different types of cliffs in the ChEMBL database. *J Chem Inf Model* 52:1806–1811. doi:10.1021/ci300274c
  15. Jayanthi LD, Ramamoorthy S (2005) Regulation of monoamine transporters: influence of psychostimulants and therapeutic antidepressants. *AAPS J* 7:E728–E738. doi:10.1208/aapsj070373
  16. Torres GE, Gainetdinov RR, Caron MG (2003) Plasma membrane monoamine transporters: structure, regulation and function. *Nat Rev Neurosci* 4:13–25. doi:10.1038/nrn1008
  17. Schneider C, Pozzi A (2011) Cyclooxygenases and lipoxygenases in cancer. *Cancer Metastasis Rev* 30:277–294. doi:10.1007/s10555-011-9310-3
  18. Kirane A, Toombs JE, Ostapoff K, Carbon JG, Zaknoen S, Braunfeld J, Schwarz RE, Burrows FJ, Brekken RA (2012) Aprioxib, a novel inhibitor of COX-2, markedly improves standard therapy response in molecularly defined models of pancreatic cancer. *Clin Cancer Res* 18:5031–5042. doi:10.1158/1078-0432.CCR-12-0453
  19. Moller DE (2001) New drug targets for type 2 diabetes and the metabolic syndrome. *Nature* 414:821–827. doi:10.1038/414821a
  20. Balakumar P, Rose M, Ganti SS, Krishan P, Singh M (2007) PPAR dual agonists: are they opening pandora's box? *Pharmacol Res* 56:91–98. doi:10.1016/j.phrs.2007.03.002
  21. Méndez-Lucio O, Pérez-Villanueva J, Castillo R, Medina-Franco JL (2012) Activity landscape modeling of PPAR ligands with dual-activity difference maps. *Bioorg Med Chem* 20:3523–3532. doi:10.1016/j.bmc.2012.04.005
  22. Dimova D, Wawer M, Wassermann AM, Bajorath J (2011) Design of multitarget activity landscapes that capture hierarchical activity cliff distributions. *J Chem Inf Model* 51:258–266. doi:10.1021/ci100477m
  23. Pérez-Villanueva J, Medina-Franco JL, Méndez-Lucio O, Yoo J, Soria-Arteche O, Izquierdo T, Lozada MC, Castillo R (2012) CASE plots for the chemotype-based activity and selectivity analysis: a CASE study of cyclooxygenase inhibitors. *Chem Biol Drug Des* 80:752–762. doi:10.1111/cbdd.12019
  24. Medina-Franco JL, Yongye AB, Pérez-Villanueva J, Houghten RA, Martínez-Mayorga K (2011) Multitarget structure–activity relationships characterized by activity-difference maps and consensus similarity measure. *J Chem Inf Model* 51:2427–2439. doi:10.1021/ci200281v
  25. Chen X, Lin Y, Gilson MK (2001) The binding database: overview and user's guide. *Biopolymers* 61:127–141. doi:10.1002/1097-0282(2002)61:2lt127:AID-BIP10076>3.0.CO;2-N
  26. Chen X, Lin Y, Liu M, Gilson MK (2002) The binding database: data management and interface design. *Bioinformatics* 18:130–139. doi:10.1093/bioinformatics/18.1.130
  27. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res* 35:D198–D201. doi:10.1093/nar/gkl999
  28. Xu YJ, Johnson M (2001) Algorithm for naming molecular equivalence classes represented by labeled pseudographs. *J Chem Inf Comput Sci* 41:181–185. doi:10.1021/ci0003911
  29. Xu YJ, Johnson M (2002) Using molecular equivalence numbers to visually explore structural features that distinguish chemical libraries. *J Chem Inf Comput Sci* 42:912–926. doi:10.1021/ci0255351
  30. Xu J (2002) A new approach to finding natural chemical structure classes. *J Med Chem* 45:5311–5320. doi:10.1021/jm010520k
  31. Xu J, Gu Q, Liu H, Zhou J, Bu X, Huang Z, Lu G, Li D, Wei D, Wang L, Gu L (2013) Chemomics and drug innovation. *Sci China Chem* 56:71–85. doi:10.1007/s11426-012-4761-0
  32. Gu Q, Yan X, Xu J (2013) Drug discovery inspired by mother nature: seeking natural biochemotypes and the natural assembly rules of the biochemome. *J Pharm Pharm Sci* 16:331–341
  33. Sud M (2012) MayaChemTools: an open source package for computational discovery. *Comp-306*, In 243rd ACS National Meeting, San Diego. American Chemical Society, Washington, D. C
  34. ROCS 3.1.0. OpenEye Scientific Software, Santa Fe. <http://www.eyesopen.com>
  35. Jaccard P (1901) Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat* 37:547–549
  36. Willett P, Barnard JM, Downs GM (1998) Chemical similarity searching. *J Chem Inf Comput Sci* 38:983–996. doi:10.1021/ci9800211
  37. Filimonov D, Poroikov V, Borodina Y, Glorizova T (1999) Chemical similarity assessment through multilevel neighborhoods of atoms: definition and comparison with the other descriptors. *J Chem Inf Comput Sci* 39:666–670. doi:10.1021/ci980335o
  38. Hall LH, Kier LB (1995) Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *J Chem Inf Comput Sci* 35:1039–1045. doi:10.1021/ci00028a014
  39. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. *J Chem Inf Model* 50:742–754. doi:10.1021/ci100050t
  40. Durant JL, Leland BA, Henry DR, Nourse JG (2002) Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci* 42:1273–1280. doi:10.1021/ci010132r
  41. Carhart RE, Smith DH, Venkataraghavan R (1985) Atom pairs as molecular features in structure–activity studies: definition and applications. *J Chem Inf Comput Sci* 25:64–73. doi:10.1021/ci00046a002
  42. Nilakantan R, Bauman N, Dixon JS, Venkataraghavan R (1987) Topological torsion: a new molecular descriptor for SAR applications. Comparison with other descriptors. *J Chem Inf Comput Sci* 27:82–85. doi:10.1021/ci00054a008
  43. Renner S, Fechner U, Schneider G (2006) In pharmacophores and pharmacophore searches, vol 32. Wiley-VCH, Weinheim
  44. Bonachéra F, Parent B, Barbosa F, Froloff N, Horvath D (2006) Fuzzy trivalent pharmacophore fingerprints. 1. Topological fuzzy pharmacophore triplets and adapted molecular similarity scoring schemes. *J Chem Inf Model* 46:2457–2477. doi:10.1021/ci6002416
  45. Chang CE, Gilson MK (2003) Tork: conformational analysis method for molecules and complexes. *J Comput Chem* 24:1987–1998. doi:10.1002/jcc.10325
  46. Vconf v2.0. VeraChem LLC, Germantown 2004. <http://www.verachem.com>
  47. Yongye AB, Byler K, Santos R, Martínez-Mayorga K, Maggiora GM, Medina-Franco JL (2011) Consensus models of activity landscapes with multiple chemical, conformer, and property representations. *J Chem Inf Model* 51:1259–1270. doi:10.1021/ci200081k
  48. Rush TS III, Grant JA, Mosyak L, Nicholls A (2005) A shape-based 3D scaffold hopping method and its application to a bacterial protein–protein interaction. *J Med Chem* 48:1489–1495. doi:10.1021/jm040163o
  49. Sykes MJ, Sorich MJ, Miners JO (2006) Molecular modeling approaches for the prediction of the nonspecific binding of drugs to hepatic microsomes. *J Chem Inf Model* 46:2661–2673. doi:10.1021/ci600221h
  50. Medina-Franco JL, Martínez-Mayorga K, Bender A, Marín RM, Giulianotti MA, Pinilla C, Houghten RA (2009) Characterization of activity landscapes using 2D and 3D similarity methods: consensus activity cliffs. *J Chem Inf Model* 49:477–491. doi:10.1021/ci800379q

51. Chen B, Mueller C, Willett P (2010) Combination rules for group fusion in similarity-based virtual screening. *Mol Inf* 29:533–541. doi:[10.1002/minf.201000050](https://doi.org/10.1002/minf.201000050)
52. Medina-Franco JL (2012) Scanning structure–activity relationships with structure–activity similarity and related maps: from consensus activity cliffs to selectivity switches. *J Chem Inf Model* 52:2485–2493. doi:[10.1021/ci300362x](https://doi.org/10.1021/ci300362x)
53. Pérez-Villanueva J, Santos R, Hernández-Campos A, Giulianotti MA, Castillo R, Medina-Franco JL (2011) Structure–activity relationships of benzimidazole derivatives as antiparasitic agents: dual activity-difference (DAD) maps. *Med Chem Commun* 2:44–49. doi:[10.1039/c0md00159g](https://doi.org/10.1039/c0md00159g)
54. Medina-Franco JL, Petit J, Maggiora GM (2006) Hierarchical strategy for identifying active chemotype classes in compound databases. *Chem Biol Drug Des* 67:395–408. doi:[10.1111/j.1747-0285.2006.00397.x](https://doi.org/10.1111/j.1747-0285.2006.00397.x)
55. Pérez-Villanueva J, Santos R, Hernández-Campos A, Giulianotti MA, Castillo R, Medina-Franco JL (2010) Towards a systematic characterization of the antiprotozoal activity landscape of benzimidazole derivatives. *Bioorg Med Chem* 18:7380–7391. doi:[10.1016/j.bmc.2010.09.019](https://doi.org/10.1016/j.bmc.2010.09.019)