JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Rapid Scanning Structure−Activity Relationships in Combinatorial Data Sets: Identification of Activity Switches

José L. Medina-Franco,*[,†] Bruce S. Edwards,[‡] Clemencia Pinilla,[§] Jon R. Appel,[§] Marc A. Giulianotti,[†] Radleigh G. Santos,[†] Austin B. Yongye,[†] Larry A. Sklar,[‡] and Richard A. Houghten[†,‡]

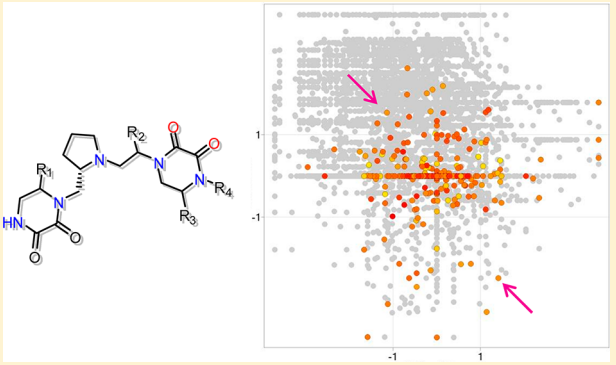
[†]Torrey Pines Institute for Molecular Studies, Port St. Lucie, Florida 34987, United States

[‡]University of New Mexico, Albuquerque, New Mexico 87131, United States

[§]Torrey Pines Institute for Molecular Studies, San Diego California 92121, United States

Supporting Information

ABSTRACT: We present a general approach to describe the structure−activity relationships (SAR) of combinatorial data sets with activity for two biological endpoints with emphasis on the rapid identification of substitutions that have a large impact on activity and selectivity. The approach uses dual-activity difference (DAD) maps that represent a visual and quantitative analysis of all pairwise comparisons of one, two, or more substitutions around a molecular template. Scanning the SAR of data sets using DAD maps allows the visual and quantitative identification of activity switches defined as specific substitutions that have an opposite effect on the activity of the compounds against two targets. The approach also rapidly identifies single- and double-target R-cliffs, i.e., compounds where a single or double substitution around the central scaffold dramatically modifies the activity for one or two targets, respectively. The approach introduced in this report can be applied to any analogue series with two biological activity endpoints. To illustrate the approach, we discuss the SAR of 106 pyrrolidine bis-diketopiperazines tested against two formylpeptide receptors obtained from positional scanning deconvolution methods of mixture-based libraries.



INTRODUCTION

Structure−activity relationship (SAR) analyses of large data sets usually require the application of computational methods, which enable an organized characterization and rapid identification of activity and selectivity cliffs. Systematic identification and quantification of such cases have been the subject of intense research giving rise to the development of "activity landscape modeling" that is extensively reviewed elsewhere.[1−3] Most of the activity landscape methods are applied to diverse data sets using fingerprint-based representations calculated from whole molecular structures. Just recently, substructure-based representations have been explored for activity landscapes using the concept of matched molecular pair (MMP), which is defined as a pair of compounds that only differs at a single site.[4−6]

Combinatorial data sets continue to play a central role in lead identification and drug discovery. For example, screening of highly dense mixture-based libraries[7−9] explores uncovered regions of the medicinally relevant chemical space,[10] increases the potential of identifying activity cliffs, and provides a rapid understanding of the SAR associated with novel leads and targets.[10] Furthermore, in vivo testing of mixture-based libraries offers the possibility of identifying "master key compounds" for multitarget drug discovery (i.e., molecules that may operate on a desired set of "locks"—targets—to gain access to a desired clinical effect).[11] High-density libraries are suitable for lead identification because they facilitate the detection of small structural modifications that contribute to biological activity and selectivity.

Computational methods have been developed to navigate and visualize the SAR of analogue series or combinatorial data sets. Recent examples include the SAR map,[12] SAR matrix,[13] single R-group polymorphisms that identify R-cliffs,[14] and SAR analysis tools recently reviewed by Duffy et al.[15] Although most of these methods are suitable to quickly identify specific R-groups that lead to active, inactive, and selective compounds, it is not straightforward to identify pairs of compounds that have opposing activity outcomes on two or more targets due to specific changes in structure.

Herein, we introduce an approach for the facile visualization and analysis of the SAR of combinatorial data sets. The method is based on systematic pairwise comparisons of the R-groups of all molecular pairs in a data set tested across two biological endpoints. The approach represents an extended application of the dual-activity difference (DAD) maps that were designed to

Received: April 1, 2013

   dx.doi.org/10.1021/ci400192y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

explore the SAR of diverse sets with activity against two targets.[16,17] In previous applications of the DAD maps, the structural relationships were obtained with similarity calculations computed using fingerprint representations. Because combinatorial data sets have, in general, low-structural diversity, fingerprint-based methods do not always result in easily interpretable SAR. In addition, it has been discussed that the activity landscape models largely depend on the variables utilized to represent chemical structures.[18,19] In order to address these issues, we present an intuitive method to compare systematic changes in the substitutions of combinatorial data sets and finding the associations between those changes and the response in biological activity. We also discuss examples of "activity switches" and "selectivity switches",[3] the latter concept defined as a minor structural modification that drastically inverts the selectivity pattern of two compounds. As a case study, we explored the SAR of a series of novel high affinity formylpeptide receptor (FPR) ligands we reported recently.[20] FPRs are a small group of G protein-coupled receptors that are important in host defense and inflammation. Specifically, the two receptors investigated were FPR1, linked to antibacterial inflammation[21] and malignant glioma cell metastasis,[22] and FPR2, linked to chronic inflammation in systemic amyloidosis, Alzheimer's disease, and prion diseases.[23] Using positional scanning deconvolution methods, FPR1 and FPR2 selective ligands with nanomolar binding affinities were identified from mixture-based molecule libraries containing more than 700,000 compounds. The ligands were identified by the screening and deconvolution of the Torrey Pines Institute for Molecular Studies (TPIMS) libraries using a high-throughput screening duplex receptor assay.[20] As noted previously, a number of compounds in this data set showed selective affinities for FPR1 or FPR2 that are the most functionally active reported to date for small molecules in a ligand competition assay format.[20]

METHODS

Data Set. We analyzed the SAR of 106 compounds obtained by screening and deconvoluting the pyrrolidine bis-diketopiperazine library in Figure 1.[20] The library has four

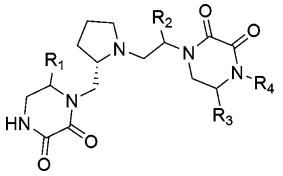


Figure 1. Core scaffold of the 106 pyrrolidine bis-diketopiperazines analyzed in this work.

diversity positions. Each molecule in the data set has reported binding inhibition constants (K_i) that were obtained from competitive ligand displacement assays. The 106 compounds were synthesized and tested as single molecules on the basis of deconvoluted results from primary high-throughput screening of mixture libraries in a duplex flow cytometry binding assay. Of note, most of the R-groups are polar substituents. The purity of all molecules was analyzed by liquid chromatography−mass spectrometry as described elsewhere.[20] The chemical structures and binding activity (K_i) are presented in Table S1 of the Supporting Information. The initial K_i values (in nM) were transformed to pK_i ($-\log_{10} K_i$) values. The activity of compounds with undefined experimental K_i (>10,000 nM) was approximated as 10,000 nM.

Dual-Activity Difference (DAD) Maps. The SAR of data sets tested with two biological endpoints can be characterized using pairwise comparisons portrayed in DAD maps proposed recently.[16,17,24,25] Given a set of N compounds tested with targets I and II, the DAD map depicts $N(N-1)/2$ pairwise potency differences for each possible pair in the data set against both targets. The potency differences for target T for each molecule pair are calculated with the expression

$$\Delta pK_i(T)_{ab} = pK_i(T)_a - pK_i(T)_b$$

where $pK_i(T)_a$ and $pK_i(T)_b$ are the activities of molecules a and b ($b > a$) against the two targets and T = FPR1, FPR2. Noteworthy, ΔpK_i can have positive or negative values providing information about the directionality of the SAR. Thus, DAD maps are able to differentiate pairs of molecules where the structural change increases the activity for one target but decreases the activity for the other target (see below).[17]

A general form of a DAD map is shown in Figure 2. Vertical and horizontal lines at $\Delta pK_i \pm t$ define boundaries for low/high

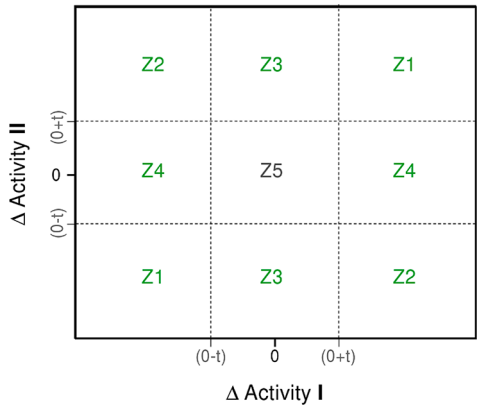


Figure 2. General form of a dual-activity difference (DAD) map for targets I and II. The dashed lines intersect the axes at potency difference values of $0 \pm t$, e.g., $t = 1$ (one log unit). The regions are as follows: Z1, substitution(s) result in a significant decrease or increase of activity in both targets; Z2, substitution(s) increase activity for one target, while decreasing activity for the other target significantly; Z3 and Z4, substitution(s) result in significant changes in activity on one target, but not an appreciable change on the other.

potency difference for targets I and II, respectively. Here, we set $t = 1$, one log unit, so that data points were considered with low potency difference if $-1 \leq \Delta pK_i \leq 1$ for each target. The boundaries define zones Z1 through Z5 in Figure 2. Structural modifications for molecule pairs that fall into zone Z1 (small or a large structural change) have a similar impact on the activity against the two targets (increase or decrease in activity). Therefore, Z1 is associated with similar SAR of the pair of compounds for both targets. In sharp contrast, pairs of compounds that fall into 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 or switch in activity,[3] which increases the activity for one target but decreases the activity for the other target. Thus, activity switches point to structural changes that completely invert the activity pattern. 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

dx.doi.org/10.1021/ci400192y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 1. Number of Pairwise Comparisons for 1−4 Substitutions and Summary of the Distribution of the Molecular Similarity Using Tanimoto and Three Different Fingerprints

			number of substitutions			
		all compounds	1	2	3	4
	no. of pairs	5565	275	896	1863	2531
MACCS	median	0.867	0.947	0.898	0.867	0.844
	U95[a]	0.870	0.946	0.903	0.873	0.851
	mean	0.869	0.941	0.900	0.871	0.849
	L95[a]	0.867	0.936	0.897	0.868	0.847
	std. dev.	0.053	0.039	0.045	0.048	0.048
GpiDAPH3	median	0.773	0.902	0.837	0.778	0.745
	U95	0.778	0.917	0.837	0.779	0.742
	mean	0.776	0.911	0.833	0.776	0.740
	L95	0.774	0.905	0.829	0.774	0.738
	std. dev.	0.074	0.052	0.058	0.060	0.059
radial	median	0.143	0.478	0.248	0.166	0.110
	U95	0.186	0.487	0.289	0.185	0.118
	mean	0.183	0.470	0.282	0.182	0.117
	L95	0.180	0.453	0.276	0.179	0.116
	std. dev.	0.109	0.144	0.094	0.063	0.030
mean similarity[b]	median	0.598	0.774	0.673	0.609	0.568
	U95	0.611	0.782	0.675	0.612	0.570
	mean	0.609	0.774	0.672	0.610	0.568
	L95	0.607	0.767	0.669	0.608	0.567
	std. dev.	0.066	0.063	0.046	0.040	0.033

[a]U95 and L95 represent the upper and lower 95% confidence intervals of the mean, respectively. [b]Computed from the average similarity of MACCS, GpiDAPH3, and radial fingerprints.

activity for the other target (II or I, respectively). Data points in 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. As previously noted, the classification of data points in an activity-difference map is independent of the structure similarity.[16,17,24,25]

Pairwise Comparison of the R-Group Substitutions. Because dual activity-difference maps are based on pairwise comparisons, it is straightforward to incorporate pairwise structure relationships by distinguishing each molecular pair by the number of substitutions (one to four) that differ between the two molecules. The number of different R-groups for each pair of compounds was determined by comparing the text strings of the chemical names of the substituents in Table S1 of the Supporting Information. This is a simple but powerful method to compare combinatorial data sets. Remarkably, the stereochemistry is taken into account such that substituents with "R" and "S" configuration are easily distinguished by the specific name of the R-group (e.g., "R-propyl" vs "S-propyl").

Analysis of the SAR is focused on data points in the DAD maps with one or two substitutions because these examples are straightforward to interpret from the experimental point of view. As noted above, distinguishing molecular pairs based on the different number of R-groups around a core scaffold is a substructure-based approach to represent chemical structures. The substructure-based representation of compounds has been proposed in activity landscape studies to enhance interpretability of the SAR[5] and address the "facts vs artifacts" issue of activity cliffs.[26]

Fingerprint Representations and Structure Similarity. For comparison, we computed pairwise similarity values using molecular access system (MACCS) keys (166-bits), graph-based three-point pharmacophores fingerprints (GpiDAPH3) implemented in MOE,[27] and radial fingerprints implemented in Canvas.[28] These three fingerprints were selected because they have conceptually different designs capturing distinctive aspects of chemical structures.[19] For example, MACCS keys used in this work are a predefined set of 166 structural keys; GpiDAPH3 fingerprints are graph-based three-point pharmacophores employing any set of three possible atom types (pi system, donor, acceptor); and radial fingerprints are equivalent to the extended connectivity fingerprints (ECFPs) and entail a growing a set of fragments radially from each heavy atom over a series of iterations.[29,30] We also computed the average similarity of all three measures as a "consensus" representation discussed previously. Briefly, the consensus or aggregated representation is intended to capture the different aspects covered by the individual molecular representations.[17,31]

RESULTS AND DISCUSSION

Overview the Diversity of the Data Set. In order to assess quantitatively the structural similarity and "high structural density" (low molecular diversity) of the data set, we measured the molecular similarity of the 106 compounds with the pairwise comparisons using the Tanimoto metric and three different fingerprints, namely, MACCS, GpiDPH3, and radial fingerprints. The average of all three Tanimoto/fingerprint similarities was computed as described in the Methods section. Table 1 summarizes the distribution of the molecular similarities of the 5565 pairwise comparisons of the 106 compounds in the data set. Table 1 also summarizes the molecular similarities of the pairwise comparisons of pairs of compounds with one, two, three, and four substitutions (275, 896, 1863, and 2531 pairs of compounds, respectively). The entire data set has, in general, low structural diversity (or high

dx.doi.org/10.1021/ci400192y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

density) as deduced from the median, mean (0.87), and other statistics of the 5565 Tanimoto/MACCS keys similarity values. For comparison, the median MACCS/Tanimoto similarity reported for a general screening collection and a set of approved drugs is 0.32 and 0.30, respectively.[10] The high density of the data set studied in this work can also be deduced from other fingerprints, e.g., the mean GpiDAPH3/Tanimoto similarity of a set of approved drugs is 0.13.[32] It is also remarkable the different ranges of similarity values obtained with the three fingerprints for the same data set. Such dependence has been extensively discussed in the literature[18,33] and emphasizes the importance of using more than one fingerprint representation.

Not surprisingly Table 1 also shows that the 275 pairs of compounds with one substitution have higher similarity than the pairs of compounds with four substitutions. Indeed, as the number of substitutions increases, the molecular similarity decreases. This result is similar for all three fingerprint representations.

To compare the similarities of the potencies of the data set toward FPR1 and FPR2, the distributions of their absolute pairwise potency differences were analyzed, and the results are summarized in Table 2. A total of 5565 pairwise comparisons

Table 2. Distribution of the Absolute Potency Difference for the Two Targets, Corresponding to All 5565 Pairwise Comparisons and All Corresponding Pairwise Comparisons for 1−4 Substitutions

			number of substitutions			
		all compounds	1	2	3	4
	no. of pairs	5565	275	896	1863	2531
FPR1	median	0.900	0.470	0.750	0.860	1.060
	U95[a]	1.077	0.707	0.954	1.031	1.232
	mean	1.055	0.633	0.906	0.994	1.198
	L95[a]	1.033	0.559	0.857	0.957	1.164
	std. dev.	0.841	0.626	0.745	0.810	0.884
FPR2	median	0.970	0.140	0.420	0.820	1.580
	U95	1.215	0.537	0.805	1.121	1.549
	mean	1.185	0.454	0.747	1.072	1.502
	L95	1.155	0.371	0.688	1.024	1.455
	std. dev.	1.139	0.703	0.889	1.072	1.199

[a]U95 and L95 represent the upper and lower 95% confidence intervals of the mean, respectively.

are shown. Considering all pairwise comparisons, the potency difference for FPR1 is lower than the difference of FPR2, as deduced from the median and all other statistics. This suggests that the activity for FPR2 was, overall, more sensitive to the structural changes of this data set. Table 2 also summarizes the distributions of the potency differences of the pairwise comparisons of pairs of compounds with one, two, three, and four changes in their R-groups. Not surprisingly, for both receptors, the potency difference increases from one to four substitutions. However, the activity for FPR1 was more sensitive than FPR2 to one and two changes in R-groups as clearly shown by the higher absolute potency differences (for example, median values of 0.47 and 0.75 for FPR1 vs 0.14 and 0.42 for FPR2, respectively). This result suggests that, for this data set, FPR1 is involved in more activity cliffs than FPR2, i.e., it is expected that a large change in activity will be observed for FPR1 ligands due to one or two substitutions in the R-groups. It remains to be explored if this observation applies for other pyrrolidine bis-diketopiperazines tested with FPR1 and FPR2. The most dramatic cliffs for each receptor are discussed in the next section with emphasis on those cases where a change in structure switches the activity pattern for FPR1 and FPR2.

Dual Activity-Difference Maps. In order to facilitate the interpretation of the SAR, the analysis is mainly focused on the pairwise comparisons of molecules with one and two substitutions in the R-groups. Figure 3 shows DAD maps with pairwise potency differences corresponding to pairs of compounds with one (275 data points) and two (896 data points) substitutions; the table beneath the plots shows the number and percentage of data points in each region of the map for compound pairs with single and double substitutions, respectively. DAD maps for pairs of compounds with three (1863 data points) and four (2531 data points) substitutions are shown in Figure S1 of the Supporting Information. Activity changes associated with three and four changes, although easily mined in the DAD maps, are less informative from the SAR interpretation point of view. Noteworthy, the distribution of the data points in all DAD maps discussed here is independent of the structure similarity.

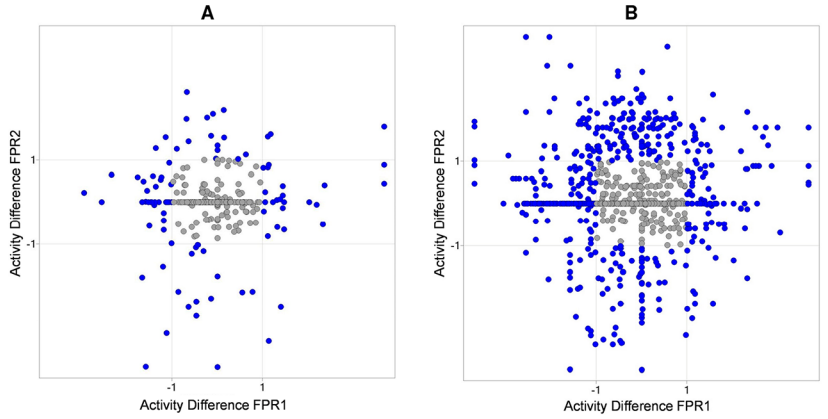
Figure 3A and B show that the majority of the data points in the corresponding DAD maps are in the center, zone Z5, in particular for the DAD map for single substitutions. These results support the notion that "similar compounds have similar activity" as one or two changes around the core scaffold do not have a large impact on the activity for both receptors, e.g., potency difference less than one log unit. As the number of substitutions increases from one to four, the percentage of pairs of compounds in Z5 decreases as shown in Table S2 of the Supporting Information.

The DAD maps in Figure 3 also show data points in the regions Z1−Z4, which are the most informative from an SAR point of view.[16,17] As discussed in the Methods section, compound pairs in these regions point to one or two R-group replacements that change dramatically (more than log unit) the activity difference for one or both receptors. Because the compounds in the data set were derived from the screening results of a mixture based combinatorial library in positional scanning format, it is expected that pairs of compounds with one or two substitutions will have relatively high structural similarity.

The number and percentage of pairs in each zone are summarized in the table below the DAD maps. The higher percentages of data points in Z4 vs Z3 (Figure 3) indicate that compared to FPR2, FPR1 is more sensitive to changes in activity than FPR2 from single and double substitutions in the data set. This result is in agreement with the results in Table 2.

Mapping Structure Similarity on DAD Maps Filtered by the Number of Substituents. Although the main focus of this work is scanning the SAR from DAD maps showing pairs of compounds with a discrete number of substituents, for reference, the mean similarity values of the molecule pairs was mapped into the DAD maps. Figure S2 of the Supporting Information shows four DAD maps with data points corresponding to one, two, three, and four substitutions. Data points are colored by mean structural similarity using a continuous scale from less similar (green) to more similar (red). The scale was defined based on the distribution of similarity values for all possible pairs of molecules in the data set. This analysis depicted in a visual manner clearly shows that pairs of compounds with only one different R-group are more

dx.doi.org/10.1021/ci400192y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX



Region	Interpretation	Single	Double
		275 pairs total	896 pairs total
Z1	Substitution changes activity for FPR1 and FPR2 (> 1 log potency) in similar magnitude and direction	7 (2.5%)	49 (5.5%)
Z2	Substitution changes activity for FPR1 and FPR2 in similar magnitude but *opposite* direction	4 (1.4%)	49 (5.5%)
Z3	Substitution changes activity for FPR2 (> 1 log potency) but not for FPR1	27 (9.8%)	189 (21.1%)
Z4	Substitution changes activity for FPR1 (> 1 log potency) but not for FPR2	53 (19.3%)	242 (27.0%)
Z5	Substitution does not change significantly (≤ 1 log potency) the activity for FPR1 and FPR2	184 (66.9%)	367 (41.0%)

Figure 3. Dual-activity difference maps for the 106 compounds. Each data point represents a pairwise comparison with (A) one substitution (275 data points total) and (B) two substitutions (896 data points). Data points in the center of each DAD map (zone Z5) with potency difference ≤1 log unit for any target are in gray. The table shows the number and percentage of data points in each region of the map for compound pairs with single and double substitutions, respectively.

similar than compound pairs with two, three, and four substitutions (Figure S2, Supporting Information, and Discussion section). This conclusion reflects the quantitative characterization of the molecular similarity analyzed above (Table 1).

The SAR obtained from zones Z3 and Z4, e.g., single-target activity cliffs, has been broadly discussed in previous applications of DAD maps for other data sets.[16,17] Herein, the discussion of the DAD maps is primarily focused on pairs of compounds in Z1 and more importantly in Z2. It should be recalled that Z1 represents similar SARs, while Z2 indicates inverse SAR. All activity switches (in Z2) are discussed first, followed by representative examples of dual-target activity cliffs with the same directionality of SAR (in Z1).

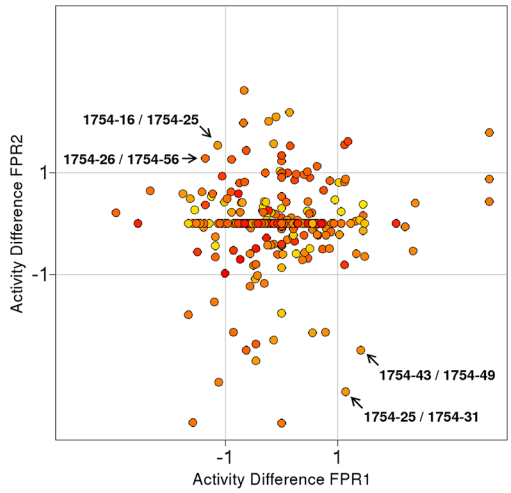
Activity Switches (Z2) with One Substitution. Figure 4 shows a DAD map displaying 275 pairs of compounds with one substitution. In this map, the data points are colored by the mean molecular similarity (distributions summarized in Table 1). As discussed above, most of these points are colored orange-to-red further emphasizing the structural similarity of the pairs of compounds with one substitution (see above). Figure 4 also shows the chemical structures, biological activity, potency difference, and as reference, the structural similarity of the four activity switches in Z2. As discussed below, all compound pairs shown in Figure 4 except 1754-26/1754-56 are in addition "selectivity switches".

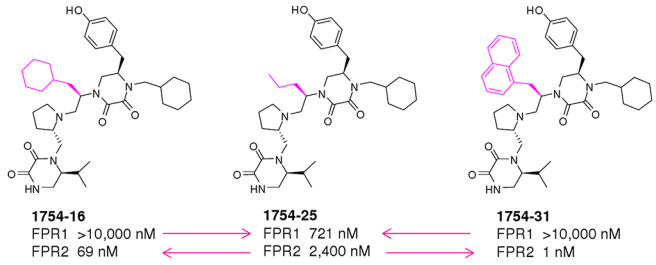
For the four pairs in Figure 4, the change in the R-group (highlighted in magenta) has a large and opposite effect on the activity of FPR1 and FPR2. For example, in the compound pair 1754-43/1754-49 the replacement of an R-propyl with R-2-naphthylmethyl at R_2 dramatically increases the activity for FPR2 by more than 2.48 log units (from K_i = >10,000 to 33 nM), but it greatly decreases the activity for FPR1 by 1.41 log units (from K_i = 90 to 2322 nM). This is also an example of a selectivity switch because 1754-43 is selective for FPR1 whereas 1754−49 is selective for FPR2.

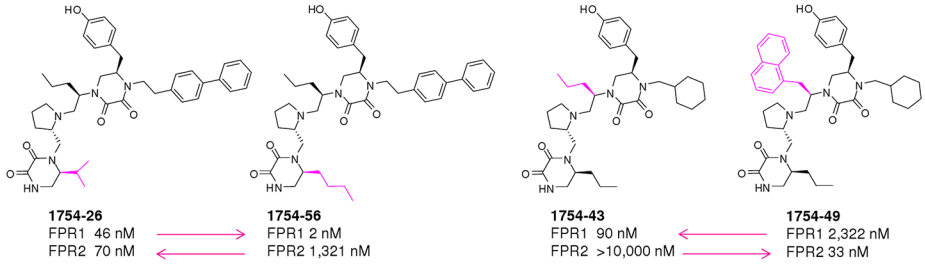
A second notable example is the compound pair 1754-26/1754-56 where the substitution of an S-isopropyl with S-butyl at R_1 increases the activity for FPR1 by 1.36 log units but decreases the activity for FPR2 in 1.28 log units. Notably, as pointed out previously, 1754-26 was the only compound in the data set with a K_i value less than 100 nM for both receptors.[20] The compound pair 1754-26/1754-56, however, is not a selectivity switch because 1754-26 is nearly equipotent with both receptors.

The mean structural similarity of the activity switches in Figure 4 is high, relative to the mean similarity of the entire data set (for all 5565 pairs of compounds, Table 1). For example, all four molecule pairs have mean similarity equal or greater than 0.73, which is higher than the U95 of the mean similarity for the whole data set, 0.61. These results indicate that the switches discussed in this figure would have been identified following a fingerprint similarity-based approach (because of its "high" fingerprint-based structural similarity). However, as discussed in the literature, using a substructure-based method to classify molecular structures is more intuitive

dx.doi.org/10.1021/ci400192y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX







Pair	ΔpK_i FPR1	ΔpK_i FPR2	Mean Similarity	MACCS	GpiDAPH3	radial
1754-16 / 1754-25	-1.14	1.54	0.736	0.917	0.936	0.354
1754-25 / 1754-31	1.14	-3.3	0.731	0.918	0.834	0.440
1754-26 / 1754-56	-1.36	1.28	0.787	0.952	0.957	0.451
1754-43 / 1754-49	1.41	-2.48	0.745	0.949	0.840	0.447

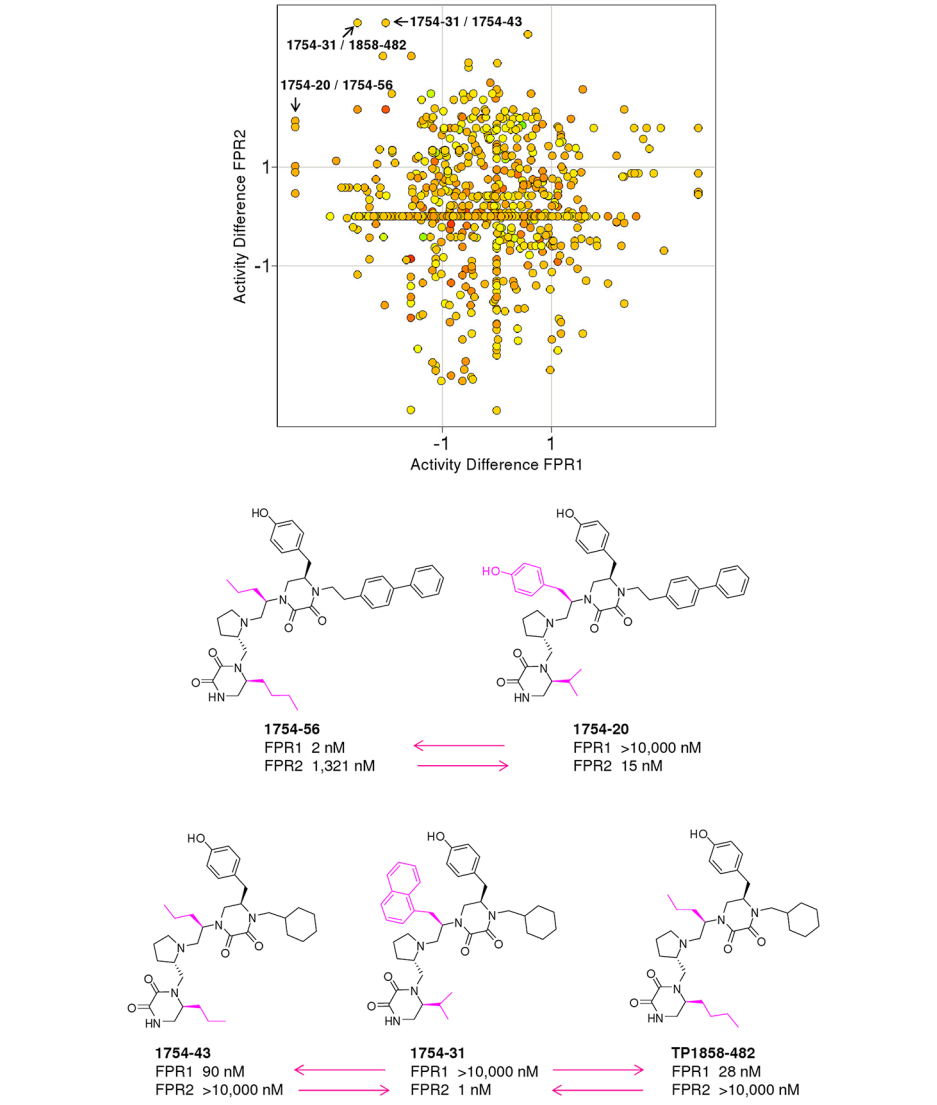
Figure 4. Activity switches for single substitutions. Data points are colored by the mean structure similarity (Figure S2, Supporting Information). The switches are readily identified in zone Z2 of the DAD maps. The structural change in each pair is highlighted in magenta. The table summarizes the potency difference and fingerprint-based similarity values for each pair.

and easy to interpret than using fingerprint representations in activity landscape studies.[5]

Activity Switches (Z2) with Two Substitutions. Figure 5 presents a DAD map showing 896 pairs of compounds with two substitutions. Data points are further distinguished by the mean molecular similarity. Comparison of this figure in the color pattern of the DAD map in Figure 4 clearly shows the overall lower structural similarity of the pairs of compounds with two substitutions (see also Table 1). Figure 5 also presents the chemical structures of three representative activity switches in Z2 (selected from 49 total) along with the biological activity, potency difference, and structural similarity. The changes in the R-groups are highlighted in magenta. We choose these examples because they have one potent compound ($K_i \leq 90$ nM) for FPR1 or FPR2 in the pair. The complete set of 49 activity switches is summarized in Table S3 of the Supporting Information.

For the three examples in Figure 5, the two changes in the R-groups have a large and opposite effect on the activity for FPR1 and FPR2. Notably, in the molecule pair 1754-31/1754-43, the replacement of an S-isopropyl to an S-propyl in R_1 and the substitution of an R-2-naphthylmethyl to R-propyl in R_2 increases the activity for FPR1 by more than two log units (from K_i = >10,000 to 90 nM), but it decreases dramatically the activity for FPR2 by four log units (from K_i = 1 to >10,000 nM). Two other remarkable examples of activity switches with two R-group replacements are given by the pairs of compounds 1754-20/1754-56 and 1754-31/1858-482. The three activity switches in Figure 5 are also selectivity switches because the corresponding replacement in the R-groups has not only an

dx.doi.org/10.1021/ci400192y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX



Pair	ΔpK_i FPR1	ΔpK_i FPR2	Mean Similarity	MACCS	GpiDAPH3	radial
1754-20 / 1754-56	-3.7	1.93	0.699	0.934	0.835	0.329
1754-31 / 1754-43	-2.04	3.92	0.673	0.885	0.814	0.319
1754-31 / 1858-482	-2.56	3.92	0.654	0.902	0.800	0.260

Figure 5. Representative activity switches with double substitutions (selected from 49 pairs in total). The switches are readily identified in the zone Z2 of DAD maps. The structural changes in each pair are highlighted in magenta. The table summarizes the potency difference and fingerprint-based similarity values for each pair.

opposite effect on the activity of the two receptors but also on the selectivity.

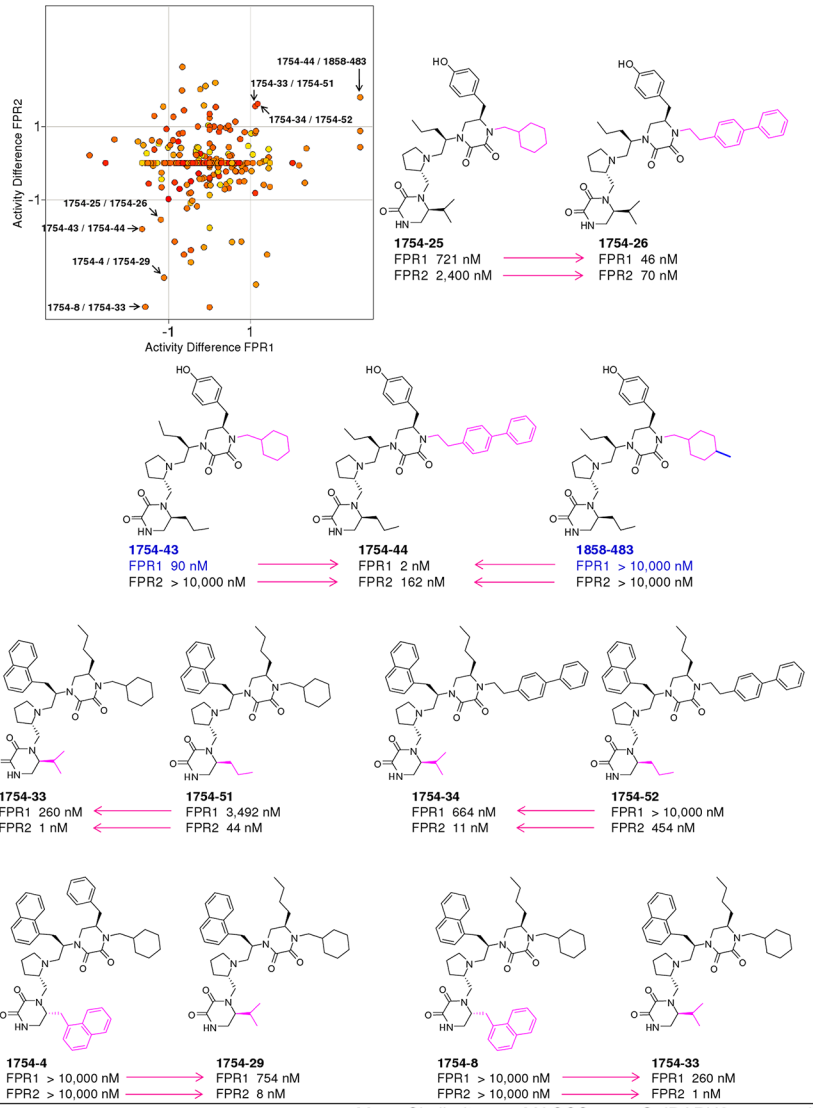
The molecular similarity of the three compound pairs in Figure 5 (e.g., mean structure similarity equal to or greater than 0.65) is still higher than the U95 mean similarity (0.61) of all pairs of compounds in the data set. Thus, they would be considered structurally similar. Although these examples would have been retrieved from a fingerprint similarity-based approach, it is clear that these pairs of compounds represent borderline cases in activity landscape studies based on fingerprint-based molecular similarity.

The previous examples illustrate that the SAR of pairs of compounds with two substitutions can be rapidly analyzed in a systematic manner in DAD maps. However, the interpretation of the SAR for compounds with two (or more) substitutions is more difficult than pairs of compounds with only one substitution. In the following subsections, the discussion is focused on representative examples of dual- and single-target activity cliffs with one R-group replacement. These cases can be regarded as R-cliffs.[14]

Dual-Target Activity Cliffs (Z1). Figure 6 shows a DAD map with pairs of compounds with one substitution. The seven Z1 data points are labeled. As clearly shown in the figure, the change in the R-group for all seven pairs simultaneously increases (or decreases) the activity for both targets by more than a log unit. The pair of compounds 1754-44 and 1858-483 illustrate that the replacement of a 2-biphenyl-4-yl-ethyl to a 4-methyl-1-cyclohexyl-methyl in R_4 decreases the activity for FPR1 and FPR2. Similar analysis can be performed for the other six pairs of compounds.

Single-Target Activity Cliffs. Figure 7 shows the three single-target activity cliffs with a very large potency difference,

dx.doi.org/10.1021/ci400192y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX



Pair	ΔpK$_i$ FPR1	ΔpK$_i$ FPR2	Mean Similarity	MACCS	GpiDAPH3	radial
1754-4 / 1754-29	-1.12	-3.11	0.765	0.889	0.917	0.488
1754-8 / 1754-33	-1.58	-3.91	0.774	0.912	0.926	0.483
1754-25 / 1754-26	-1.2	-1.54	0.764	0.967	0.882	0.443
1754-33 / 1754-51	1.12	1.55	0.829	0.947	0.983	0.556
1754-34 / 1754-52	1.18	1.61	0.840	0.948	0.985	0.587
1754-43 / 1754-44	-1.66	-1.79	0.773	0.966	0.900	0.452
1754-44 / 1858-483	3.7	1.79	0.756	0.934	0.887	0.447

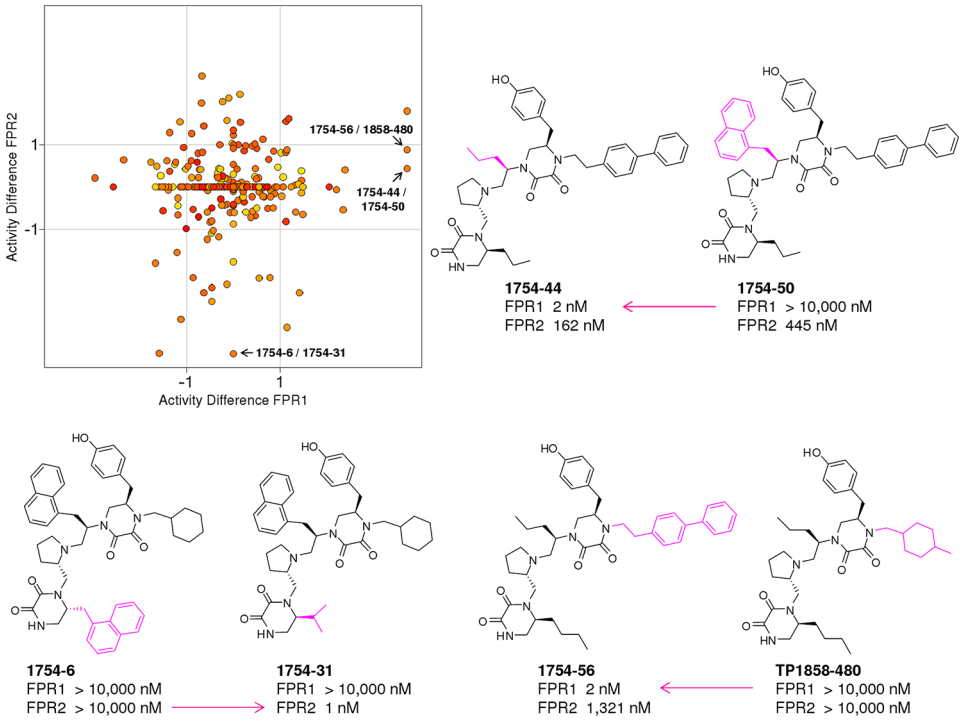
Figure 6. Dual-target activity cliffs for single substitutions with the same direction, e.g., increases or decreases the activity for the two targets. The seven activity cliffs with direct SAR are readily identified in zone Z1 of the DAD maps. The structural changes in each pair are highlighted in magenta. The table summarizes the potency difference and fingerprint-based similarity values for each pair.

i.e., more than three log units. These examples can be regarded as "deep activity cliffs".[31] It is clear from the figure that the corresponding R-group modification greatly changes the activity for only one target, either FPR1 (pairs 1754-44/1754-50; 1754-56/1858-480) or FPR2 (1754-6/1754-31). However, the replacement of the R-group does not have a major impact (<1 log unit) on the activity of the other target.

Interestingly, the chemical structures and K_i values of compounds 1754-43 and 1858-483 in Figure 6 exemplify a single-target activity cliff for FPR1 (the pair 1754-43/1858-483 is not labeled in the DAD map in figure 6 that illustrates dual-target cliffs). The only difference in the chemical structure of this pair of compounds is the methyl group in the R_4 substituent highlighted in blue, i.e., cyclohexyl-methyl vs 4-methyl-1-cyclohexyl-methyl. This subtle modification changes the activity for FPR1 from K_i = 90 nM (1754-43) to K_i > 10,000 nM (1858-483). However, this modification has no impact on the activity of FPR2.

SAS Maps. DAD maps are highly related to the structure−activity similarity (SAS) maps that are two-dimensional plots representing the relationship between the potency difference for one target (typically plotted on the *Y*-axis) and the structural similarity (usually plotted on the *X*-axis).[3] It is possible to generate SAS maps for FPR1 and FPR2 and further distinguish the data points based on the number of R-group changes (Figures S3 and S4, Supporting Information).

dx.doi.org/10.1021/ci400192y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX



Pair	ΔpK$_i$ FPR1	ΔpK$_i$ FPR2	Mean Similarity	MACCS	GpiDAPH3	radial
1754-44 / 1754-50	**3.7**	0.44	0.767	0.967	0.860	0.474
1754-56 / 1858-480	**3.7**	0.88	0.761	0.951	0.900	0.432
1754-6 / 1754-31	0	**-3.92**	0.773	0.898	0.924	0.498

Figure 7. Activity cliffs for FPR1 and FPR2 (deep cliffs with >3 log units in potency difference for single substitutions). The structural changes in each pair are highlighted in magenta. The table summarizes the potency difference and fingerprint-based similarity values for each pair.

Although it is straightforward to identify single-target activity cliffs from these maps, the directionality of the SAR is lost; therefore, it is impossible to identify selectivity switches and dual-target cliffs with similar SAR.

CONCLUSIONS

We report an intuitive substructure-based approach for the systematic and rapid scanning of the SAR in combinatorial data sets using the concept of activity landscape modeling. The general method herein introduced enables the quick and methodological identification on large changes in biological activity associated with one, two, or more replacements in the R-groups of a common scaffold. The larger and more structural diverse the R-groups (e.g., measured by a fingerprint-based method) will increase the "applicability domain" of the DAD maps. The approach captures the substructure relationships of all possible pairs in the data set, and it is reminiscent of the MMP concept. However, while the MMP and MMP-cliffs consider structural changes at a single R-group, the current approach considers structural changes at one, two, or more groups. Of course, changes at one or two groups are the ones that are easier to interpret for a medicinal chemist. The substructure relationships can be readily mapped and visualized in DAD maps. DAD maps are based on pairwise comparisons of potency differences making it straightforward to represent changes in the R-groups and to systematically explore single- and dual-target activity R-cliffs. In particular, DAD maps enable the rapid identification of all "activity switches", defined as pairs of compounds where a small change in the structure (e.g., replacement of one or two R-groups) completely inverts the biological response for two targets, namely, increases the activity for one target but decreases the activity for the second target. Several "activity switches" identified in this work were also "selective switches" defined as structural changes that completely invert the selectivity pattern of similar compounds against two biological endpoints. Activity and selectivity cliffs and switches are of value for the medicinal chemist because they point to specific molecules and substitutions that can be further modified for improved activity and/or selectivity. However, as currently stands, the DAD maps presented in this work are focused on the description of the SAR. Efforts to conduct prospective applications of DAD maps are ongoing. Of note, the number of data points analyzed in the DAD maps are selected solely based on the number of R-group substitutions of the core scaffold. This approach is in sharp contrast with previous applications of the DAD maps where data points are selected based on the fingerprint-based similarity values. To illustrate the rapid detection of activity switches in combinatorial data sets using DAD maps, we discuss the activity switches, single- and dual-target activity cliffs of a novel and relevant set of 106 pyrrolidine bis-diketopiperazines tested with two formylpeptide receptors.[20] Most of the R-group substituents of the main scaffold are hydrophobic groups, and we did not observe cliffs that involve highly polar groups. This data set represents one example of several combinatorial data sets obtained from positional scanning deconvolution methods of mixture-based libraries.[7,9]

dx.doi.org/10.1021/ci400192y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

■ ASSOCIATED CONTENT

Supporting Information

SMILES representation and biological activity of the 106 compounds analyzed in this work (Table S1); number of pairs of compounds in Z5 (Table S2); 49 activity switches with two substitutions (Table S3); consensus SAS maps for FPR1 (Figure S3); and consensus SAS maps for FPR2 (Figure S4). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

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

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Jacob Waddell for writing scripts used to generate the DAD maps. This work was supported by NIH Grants U54MH074425 (L.A.S.), U54MH084690 (L.A.S.), R01HG005066 (B.S.E.), and 1R01DA031370 (R.A.H.) and the University of New Mexico Shared Flow Cytometry and High Throughput Screening Resource (supported in part by UNM Cancer Center and NIH Grants P30 CA118100 and U54 RR026083). We also appreciate funding from the State of Florida, Executive Office of the Governor's Department of Economic Opportunity.

■ REFERENCES

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

(2) Bajorath, J. Modeling of activity landscapes for drug discovery. *Expert. Opin. Drug Discov.* **2012**, *7*, 463−473.

(3) Medina-Franco, J. L. Scanning structure−activity relationships with with structure−activity similarity and related maps: From consensus activity cliffs to selectivity switches. *J. Chem. Inf. Model.* **2012**, *52*, 2485−2493.

(4) Kenny, P. W.; Sadowski, J. *Chemoinformatics in Drug Discovery*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2005.

(5) Hu, X.; Hu, Y.; Vogt, M.; Stumpfe, D.; Bajorath, J. MMP-Cliffs: Systematic identification of activity cliffs on the basis of matched molecular pairs. *J. Chem. Inf. Model.* **2012**, *52*, 1138−1145.

(6) Dossetter, A. G.; Griffen, E. J.; Leach, A. G. Matched molecular pair analysis in drug discovery. *Drug Discovery Today* **2013**, DOI: 10.1016/j.drudis.2013.03.003.

(7) Houghten, R. A.; Pinilla, C.; Appel, J. R.; Blondelle, S. E.; Dooley, C. T.; Eichler, J.; Nefzi, A.; Ostresh, J. M. Mixture-based synthetic combinatorial libraries. *J. Med. Chem.* **1999**, *42*, 3743−3778.

(8) Pinilla, C.; Appel, J. R.; Borras, E.; Houghten, R. A. Advances in the use of synthetic combinatorial chemistry: Mixture-based libraries. *Nat. Med.* **2003**, *9*, 118−122.

(9) Houghten, R. A.; Pinilla, C.; Giulianotti, M. A.; Appel, J. R.; Dooley, C. T.; Nefzi, A.; Ostresh, J. M.; Yu, Y. P.; Maggiora, G. M.; Medina-Franco, J. L.; Brunner, D.; Schneider, J. Strategies for the use of mixture-based synthetic combinatorial libraries: Scaffold ranking, direct testing, in vivo, and enhanced deconvolution by computational methods. *J. Comb. Chem.* **2008**, *10*, 3−19.

(10) López-Vallejo, F.; Giulianotti, M. A.; Houghten, R. A.; Medina-Franco, J. L. Expanding the medicinally relevant chemical space with compound libraries. *Drug Discovery Today* **2012**, *17*, 718−726.

(11) Medina-Franco, J. L.; Giulianotti, M. A.; Welmaker, G. S.; Houghten, R. A. Shifting from the single to the multitarget paradigm in drug discovery. *Drug Discovery Today* **2013**, *18*, 495−501.

(12) Kolpak, J.; Connolly, P. J.; Lobanov, V. S.; Agrafiotis, D. K. Enhanced SAR maps: Expanding the data rendering capabilities of a popular medicinal chemistry tool. *J. Chem. Inf. Model.* **2009**, *49*, 2221−2230.

(13) Wassermann, A. M.; Haebel, P.; Weskamp, N.; Bajorath, J. SAR Matrices: Automated extraction of information-rich SAR tables from large compound data sets. *J. Chem. Inf. Model.* **2012**, *52*, 1769−1776.

(14) Agrafiotis, D. K.; Wiener, J. J. M.; Skalkin, A.; Kolpak, J. Single R-group polymorphisms (SRPs) and R-cliffs: An intuitive framework for analyzing and visualizing activity cliffs in a single analog series. *J. Chem. Inf. Model.* **2011**, *51*, 1122−1131.

(15) Duffy, B. C.; Zhu, L.; Decornez, H.; Kitchen, D. B. Early phase drug discovery: cheminformatics and computational techniques in identifying lead series. *Bioorg. Med. Chem.* **2012**, *20*, 5324−5342.

(16) 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. Comm.* **2011**, *2*, 44−49.

(17) Medina-Franco, J. L.; Yongye, A. B.; Pérez-Villanueva, J.; Houghten, R. A.; Martínez-Mayorga, K. Multitarget structure−activity relationships characterized by activity-difference maps and consensus similarity measure. *J. Chem. Inf. Model.* **2011**, *51*, 2427−2439.

(18) 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.

(19) 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.

(20) Pinilla, C.; Edwards, B. S.; Appel, J. R.; Yates-Gibbins, T.; Giulianotti, M. A.; Medina-Franco, J. L.; Young, S. M.; Santos, R. G.; Sklar, L. A.; Houghten, R. A. Selective agonists and antagonists of formylpeptide receptors: Duplex flow cytometry and mixture-based positional scanning libraries. *Mol. Pharmacol.* **2013**, in press.

(21) Le, Y.; Murphy, P. M.; Wang, J. M. Formyl-peptide receptors revisited. *Trends Immunol.* **2002**, *23*, 541−548.

(22) Zhou, Y.; Bian, X.; Le, Y.; Gong, W.; Hu, J.; Zhang, X.; Wang, L.; Iribarren, P.; Salcedo, R.; Howard, O. M. Z.; Farrar, W.; Wang, J. M. Formylpeptide receptor FPR and the rapid growth of malignant human gliomas. *J. Natl. Cancer Inst.* **2005**, *97*, 823−835.

(23) Le, Y.; Yazawa, H.; Gong, W.; Yu, Z.; Ferrans, V. J.; Murphy, P. M.; Wang, J. M. Cutting edge: The neurotoxic prion peptide fragment PrP(106−126) is a chemotactic agonist for the G protein-coupled receptor formyl peptide receptor-like 1. *J. Immunol.* **2001**, *166*, 1448−1451.

(24) Pérez-Villanueva, J.; Medina-Franco, J. L.; Méndez-Lucio, O.; Yoo, J.; Soria-Arteche, O.; Izquierdo, T.; Lozada, M. C.; Castillo, R. Case plots for the chemotype-based activity and selectivity analysis: A case study of cyclooxygenase inhibitors. *Chem. Biol. Drug Des.* **2012**, *80*, 752−762.

(25) Méndez-Lucio, O.; Pérez-Villanueva, J.; Castillo, R.; Medina-Franco, J. L. Activity landscape modeling of PPAR ligands with dual-activity difference maps. *Bioorg. Med. Chem.* **2012**, *20*, 3523−3532.

(26) Medina-Franco, J. L. Activity cliffs: Facts or artifacts? *Chem. Biol. Drug Des.* **2013**, *81*, 553−556.

(27) *Molecular Operating Environment (MOE)*, version 2011.10; Chemical Computing Group, Inc.: Montreal, Quebec, Canada. http://www.chemcomp.com (accessed May 31, 2013).

(28) *Canvas*, version 1.5; Schrödinger, LLC: New York, 2012.

(29) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model* **2010**, *50*, 742−754.

(30) 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.

(31) 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.
(32) Singh, N.; Guha, R.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A.; Medina-Franco, J. L. Chemoinformatic analysis of combinatorial libraries, drugs, natural products, and molecular libraries small molecule repository. *J. Chem. Inf. Model.* **2009**, *49*, 1010−1024.
(33) 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.