JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

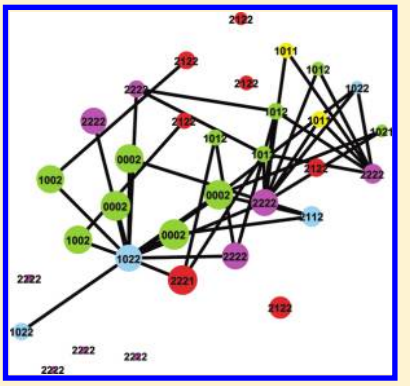
Design of Multitarget Activity Landscapes That Capture Hierarchical Activity Cliff Distributions

Dilyana Dimova, Mathias Wawer, Anne Mai Wassermann, and Jürgen Bajorath*

Department of Life Science Informatics, B-IT, LIMES Program Unit, Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Dahlmannstr. 2, D-53113 Bonn, Germany

Supporting Information

ABSTRACT: An activity landscape model of a compound data set can be rationalized as a graphical representation that integrates molecular similarity and potency relationships. Activity landscape representations of different design are utilized to aid in the analysis of structure−activity relationships and the selection of informative compounds. Activity landscape models reported thus far focus on a single target (i.e., a single biological activity) or at most two targets, giving rise to selectivity landscapes. For compounds active against more than two targets, landscapes representing multitarget activities are difficult to conceptualize and have not yet been reported. Herein, we present a first activity landscape design that integrates compound potency relationships across multiple targets in a formally consistent manner. These multitarget activity landscapes are based on a general activity cliff classification scheme and are visualized in graph representations, where activity cliffs are represented as edges. Furthermore, the contributions of individual compounds to structure−activity relationship discontinuity across multiple targets are monitored. The methodology has been applied to derive multitarget activity landscapes for compound data sets active against different target families. The resulting landscapes identify single-, dual-, and triple-target activity cliffs and reveal the presence of hierarchical cliff distributions. From these multitarget activity landscapes, compounds forming complex activity cliffs can be readily selected.



INTRODUCTION

The concept of activity landscapes provides the basis for a comprehensive analysis of structure−activity relationships contained in large compound sets.[1] For example, activity landscape models aid in the rationalization of global and local SAR features and the selection of compounds for chemical exploration.[1] Activity landscapes are generally defined as representations that integrate structure and potency relationships between compounds having the same biological activity.[2] As such, activity landscapes can be represented in rather different ways, ranging from simple 2-D plots that compare the structural and activity similarity between data set compounds in a pairwise manner[3] and potency-annotated molecular network representations[4,5] to detailed 3-D landscape views.[6] In such 3-D activity landscape models, an interpolated potency surface is added to a 2-D projection of chemical reference space as the third dimension,[6] giving rise to landscapes that are reminiscent of topographical maps.[2,7]

Regardless of the specifics of different activity landscape representations, the assessment of pairwise molecular similarity relationships is a key element of landscape design. It has been shown that chosen molecular representations for similarity evaluation very often influence the topology of landscape models.[6,8] The most prominent features of activity landscapes, however they might be represented, are activity cliffs that are formed by pairs or groups of structurally similar compounds, for example, analog series, with large differences in potency.[2,9] Regions spanning multiple activity cliffs are rich in SAR information content and represent primary focal points of landscape analysis.

Although a common feature of most activity landscape representations reported thus far is that they focus on activity against a single target, there are no principal reasons to limit activity landscape modeling to individual targets. However, only very few studies have considered two biological activities of compounds for the generation of activity landscapes. Recently, an extension of the activity landscape concept has been introduced, where potency ratios for compounds active against two targets have been utilized instead of single-target compound potency values.[10] The use of potency ratios (or logarithmic potency differences) to annotate similarity-based compound networks is straightforward, giving rise to selectivity landscapes and the notion of selectivity cliffs that are formed by similar compounds having significantly different potency against the two targets.[10] Comparisons of dual-target activity landscapes have also been carried out in a pairwise manner for analog series with activity annotations against three targets (i.e., yielding three pairwise landscape representations).[11] In this case, insights into complex SARs of compound series could be obtained by

Received: December 6, 2010
Published: January 28, 2011

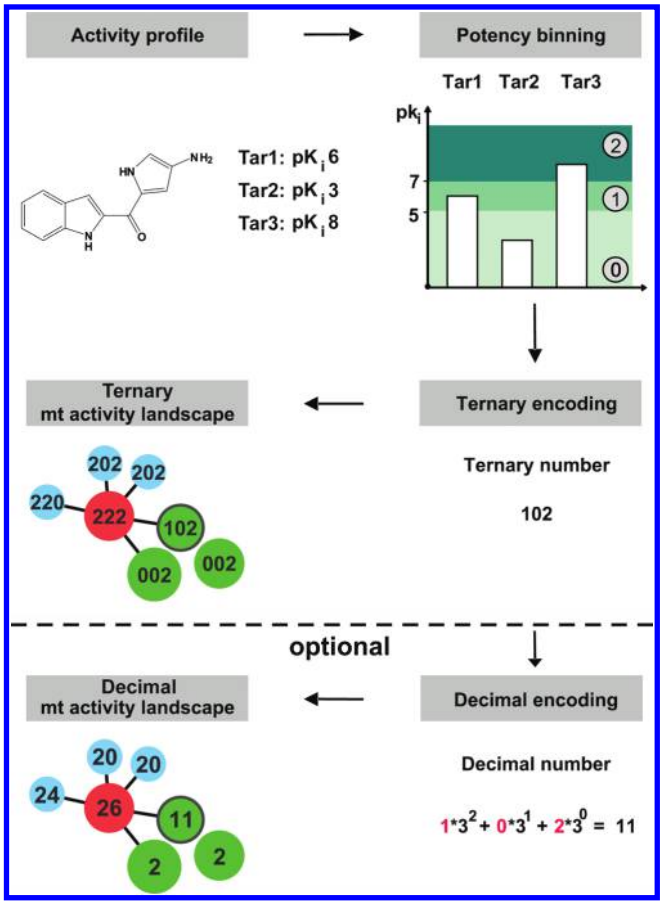


Figure 1. Generation and encoding of compound activity profiles. The schematic illustration summarizes the steps involved in converting compound potencies against multiple targets into activity profiles and representing these activity profiles as ternary numbers (or corresponding optional decimal codes). Ternary (or decimal) codes are used as node labels in the multitarget activity landscape representations.

comparing corresponding regions in target-pair landscape models.[11]

Going beyond target-pair landscapes and selectivity cliffs, a currently unsolved problem is the generation of an activity landscape framework for multiple targets. Generating multitarget activity landscapes would be relevant, for example, for the study of ligands active against different members of protein families or polypharmacological targets. However, activity landscape representations for compounds active against three or more targets cannot be obtained on the basis of currently available models, and new design concepts are required. Herein, we report a methodology to derive and visualize multitarget activity landscapes and analyze activity cliff distributions. As an exemplary application, the approach is utilized to characterize compound data sets active against members of different target families.

MATERIALS AND METHODS

Potency Binning and Encoding. A three-level scheme is applied for encoding compound activity profiles.

(1) Potency values are assigned to three different ranges (bins) in order to classify compounds as *weakly potent* ($pK_i \leq 5$), *moderately potent* ($pK_i > 5$ and $pK_i \leq 7$), or *highly potent* ($pK_i > 7$). Hence, a compound is considered weakly potent against a given target if its potency is less than or equal to 10 μM, moderately potent if its potency is greater than 10 μM but lower than or equal to 100 nM, and highly potent if its potency is higher than 100 nM (these potency intervals can be adjusted for different applications).

(2) Weakly potent compounds are assigned the ternary digit "0", moderately potent the digit "1", and highly potent the digit "2". These potency bin values define a ternary numeral system. A ternary numeral system represents numeric values using only the digits 0, 1, and 2. To simplify the notation, we denote a ternary number v of a length n by a sequence of ternary digits of length n, i.e., $v = v_1 \ldots v_n$ with $v_i \in \{0,1,2\}$ for all $1 \leq i \leq n$.

(3) The activity profile of a compound active against n targets is uniquely mapped to a ternary number of length n. For a given ternary number v and number of targets n, we denote by $[v]_i$ the ternary digit at position i in v. This ternary number can also be converted into a decimal code that serves as a compound label defining its multitarget activity profile.

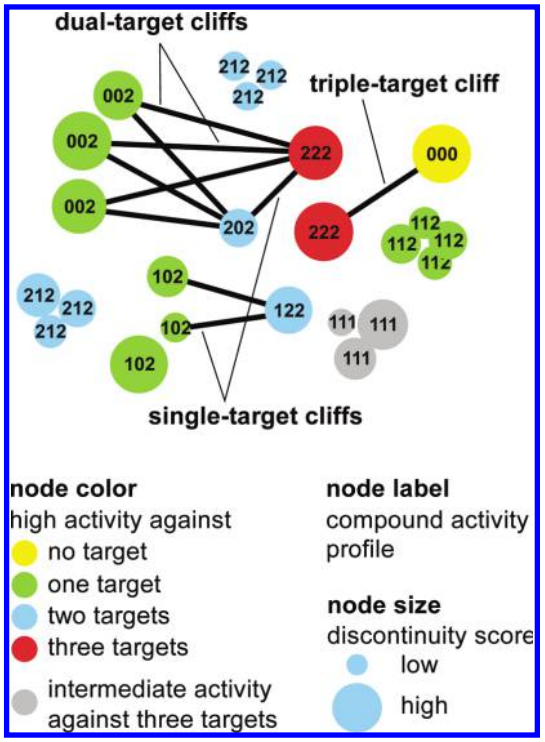


Figure 2. Activity landscape representation. The design of multitarget activity landscapes and the components and information layers of these network-like graphs are schematically illustrated.

Figure 1 illustrates our three-level potency classification and encoding scheme.

Multitarget Graphs. For graphical representations of multitarget activity landscapes, we have modified and further extended the network-like similarity graph (NSG) data structure,[5] a JAVA implementation that is publicly available as part of the SARANEA program suite.[12] For multitarget landscape displays, this similarity-based compound network was annotated with compound activity profiles. For visualization, a layout algorithm[13] is applied that places groups of densely connected vertices in close vicinity, while separating weakly connected regions of the graph from

dx.doi.org/10.1021/ci100477m |J. Chem. Inf. Model. 2011, 51, 258–266

Table 1. Compound Data Sets[a]

set	activity	no. of compounds	no. of targets	targets
AR	adenosine receptor antagonists	342	3	adenosine receptors A1, A2a, A3
MT	monoamine transporter inhibitors	299	3	dopamine (DA), norepinephrine (NE), serotonin (5HT) transporters
OR	opioid receptor antagonists	98	4	δ-, κ-, μ-opioid receptor, nociceptin (O) receptor
CA	carbonic anhydrase (CA) inhibitors	96	4	carbonic anhydrases 1, 2, 9, 12

[a] Four compound data sets were collected from ChEMBL. For each set, the specific activity (activity), number of antagonists or inhibitors (no. of compounds), number of targets (no. of targets), and target names (targets) are reported.

each other. Multitarget graphs are interactively navigated. They can be zoomed and edited, and nodes are graphically associated with compound structures (a structure is displayed when the cursor is placed on a node). The data structure can also be systematically searched for structural relationships or activity cliffs.

Figure 2 summarizes the design elements of these multitarget graphs. Compounds are represented as nodes that are connected by type-1 edges if their pairwise structural similarity, calculated as Tanimoto similarity for ECFP4 fingerprints,[14] exceeds a threshold of 0.4. Nodes are labeled with ternary (or decimal) codes that represent their activity profiles and are color-coded to highlight selected profiles: compounds highly potent against only one, two, three, or four targets are colored green, blue, red, and purple, respectively, and compounds with intermediate potency against all targets are colored light gray. All compounds with potency profiles different from these four categories (i.e., different combinations of moderate and/or low potencies) are colored yellow. Nodes are scaled in size according to a multitarget discontinuity score (as described below) such that large nodes represent high discontinuity score values. Single-target and multitarget activity cliffs formed by pairs of compounds are identified by type-2 edges, i.e., an edge is drawn between two compounds if they form an activity cliff. Single-target and multiple target activity cliffs are selectively displayed.

Multitarget Discontinuity Score. A multitarget discontinuity score (mtDiscScore) is defined to quantitatively account for the degree of multitarget SAR discontinuity that an individual compound introduces in the activity landscape. This score is a variant of the SAR Index per-compound discontinuity score we previously introduced.[5] For conventional single-target activity landscapes, the per-compound discontinuity scores identifies compounds that have large potency deviations from their immediate structural neighbors. Here, the mtDiscScore is defined in analogy to the per-compound discontinuity score. Thus, the mtDiscScore quantitatively compares potency differences across multiple targets for each data set compound with its structural neighbors in a pairwise manner. Formally, let n be the number of targets. Two compounds are considered similar if their ECFP4 value is greater than 0.4. For every compound c we define the set $N(c)$ as the set of all structural neighbors. Furthermore, let $|N(c)| = m$. For a ligand c, we define $[c]_i$ as the potency value of c against target i. Then the mtDiscScore is defined as

$$\mathrm{mtDiscScore}_{\mathrm{raw}}(c) = \frac{1}{n \cdot m} \sum_{c' \in N(c) \mathrm{sim}(c,\, c') > 0.4} \sum_{i=1}^{n} \left| [c]_i - [c']_i \right| \cdot \mathrm{sim}(c,\, c')$$

The raw scores are standardized on the basis of Z-score calculations including all compounds in a data set and normalized by calculating the cumulative probability for a normal distribution, yielding final scores falling into the range [0,1].

Given this formalism, a compound makes large contributions to multitarget SAR discontinuity (and achieves a high score) if it has many structural neighbors with different potency profiles.

Data Sets. For the introduction of multitarget activity landscapes, we selected four compound data sets from ChEMBL[15] with (antagonistic or inhibitory) activity against three or four targets belonging to four different families, i.e., adenosine receptors (AR), monoamine transporters (MT), opioid receptors (OR), and carbonic anhydrases (CA). The composition of these compound sets is summarized in Table 1. Only potency measurements were selected with the highest target confidence level (i.e., target confidence score 9) for direct interactions (i.e., target relationship type "D"). Potency measurements containing threshold values (i.e., reported as > or <) were not considered. For compounds with multiple potency values reported against the same target, the arithmetic mean was calculated to yield the final potency.

RESULTS AND DISCUSSION

Activity Landscape Design Principles and Applications. The generation of activity landscape models generally requires the integration of compound similarity and potency relationships. Assessing similarity relationships is independent of the number of targets and a constant for different landscape models representing the same data set. Conventional single-target activity landscapes are utilized to extract SAR information from compound data sets and identify prominent activity cliffs, i.e., structurally similar compounds having large potency differences. The major challenge of multitarget activity landscape design is how to combine compound potency relationships for several targets and best represent activity profiles. For two targets, this can be accomplished by assigning potency ratios to compounds instead of individual potency values. However, for more than two targets, this simple approach is no longer feasible and different data structures are required. For this purpose, we introduce herein a multitarget potency encoding scheme that is based on a ternary numeral system.

The representation of multitarget activity landscapes does not only provide a methodological challenge, but is also of practical relevance for pharmaceutical research. For example, in order to selectively optimize compounds against an individual target compared to closely related ones, multitarget landscapes provide an access to multitarget activity cliffs, their most prominent features. A multitarget activity cliff is formed by compounds with differential potency against two or more targets. Compounds forming such cliffs are most likely to provide information about structural modifications that change compound potency against two or more targets in either the same or opposite directions. This information would be valuable to guide selective

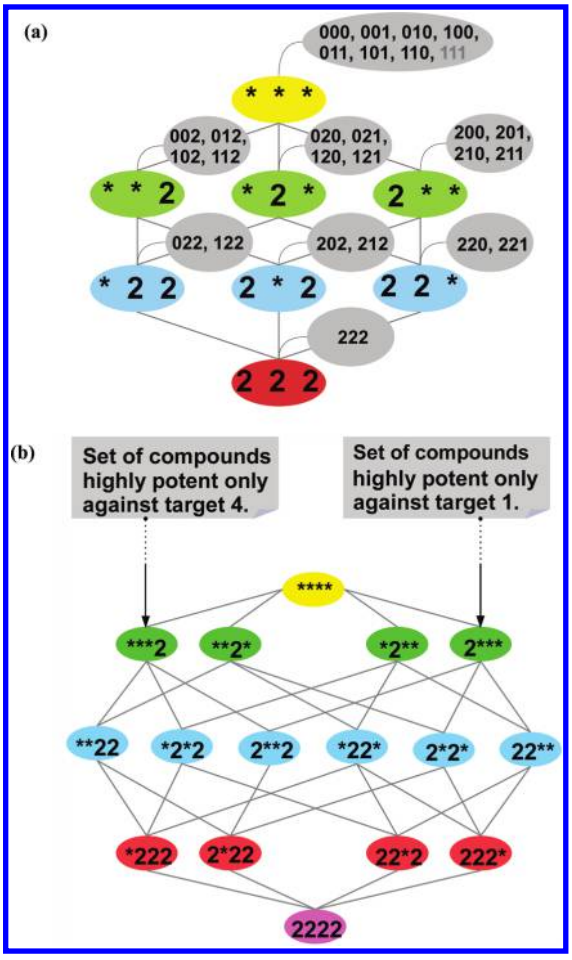


Figure 3. Activity profiles. Shown is the formal organization of activity profiles containing highly potent compounds for (a) three and (b) four targets. Asterisks indicate ternary digits of either 0 or 1. In (a), sets of all ternary numbers covered by the generic profiles are shown as gray tags. The colors correspond to the node coloring scheme introduced for multitarget activity landscapes. This organization scheme provides a basis for the systematic enumeration of all principally possible single-target and multitarget activity cliffs and specification of different activity cliff types using decimal code combinations.

optimization efforts. In multitarget landscapes, as introduced herein, compounds forming such cliffs are readily identified, and their potency profiles can then be directly compared.

Activity Profile Encoding. By classifying compounds as weakly ("0"), moderately ("1"), or highly potent ("2"), we encode multitarget activity profiles as ternary numbers, e.g., "121" for three or "0112" for four targets. Applying this simple formalism, each possible compound activity profile is uniquely encoded. The activity profiles can be systematically organized in different ways. For example, in Figure 3, activity profiles are arranged on the basis of high compound potency against one to three (Figure 3a) or one to four (Figure 3b) targets, which is a prerequisite for the systematic organization of all theoretically possible activity cliffs, as discussed in the following section. The top level in these graphs represents activity profiles of compounds that do not have high potency against any target, the second level compounds with high potency against only one target, the third level compounds with high potency against two targets, and so on.

Activity Cliff Organization. On the basis of our encoding scheme, we define that an activity cliff is formed by any pair of compounds representing a "2−0" potency combination against any target, i.e., a cliff is formed by one compound with high and one with low potency. Hence, for compounds with activity against three targets, we can formally distinguish between single-, dual-, and triple-target activity cliffs, and for compounds active against four targets, quadruple-target activity cliffs can in principle also be formed. In Figure 3a, all possible activity profiles for highly potent compounds and three or four targets are reported that can participate in the formation of activity cliffs. For three and four targets, there are a total of 7 and 15 high-potency profile categories, respectively, that can form activity cliffs. In Figure 3a, the ternary codes of activity profiles representing each type are also provided. These activity profiles can be systematically paired with "0"-containing profiles to yield the theoretically possible numbers of single- and multiple-target activity cliffs. It should be noted that compounds with moderate potency against all targets cannot participate in the formation of activity cliffs. The corresponding activity profiles for three and four targets are "111" and "1111", respectively.

We have calculated the numbers of all principally possible single- and multiple-target activity cliffs that result from unique profile combinations. For three targets, there are 147, 42, and 4 different types of single-, dual-, and triple-target activity cliffs possible, respectively. For four targets, the corresponding numbers are 1372 single-, 588 dual-, and 112 triple-target cliffs. In addition, in this case, 16 types of quadruple-target cliffs could be formed. Each of these potential activity cliffs is identified by a unique code combination. For example, activity profile "222" (representing compounds with high potency against three targets) forms a single-target activity cliff with profile "201".

For three or four targets, ternary numbers might be directly used as node labels. If more target annotations would be available, ternary codes might become too large, and hence, they could be transformed into shorter decimal codes for visualization (as illustrated in Figure 1), for example, using "16" instead of "121" or "14" for "0112". Decimal codes are less intuitive than ternary numbers, but can be interpreted, for example, with the help of a conversion table.

Multitarget Activity Landscapes. In Figure 4, the multitarget graph for the MT compound set is shown. For the description and interpretation of our multitarget activity landscape design, we focus on the MT compound set in the text and provide corresponding representations for the three other data sets in the Supporting Information. Figure S1 of the Supporting Information shows the multitarget graphs for the AR, OR, and CA sets. In addition, for the four-target CA set, a graph with decimal node labels is also shown for comparison to illustrate the decimal coding scheme. Detailed descriptions of the graphs of these three data sets are provided in the Results section of the Supporting Information. In all graph representations, type-1 edges indicating pairwise compound similarity relationships are not displayed for clarity.

The MT inhibitor landscape displays extensive clustering of compound subsets, revealing regions with densely packed nodes that are separated from other clusters. In many cases, compounds in individual clusters have similar activity profiles, i.e., they have the same node color and the same or similar codes. The presence of similar activity profiles would be expected for compounds that are active against closely related targets. Several clusters of structurally similar compounds are observed that include predominantly highly potent (red; code 222) or weakly potent (yellow; code 000) compounds. Furthermore, the graph contains

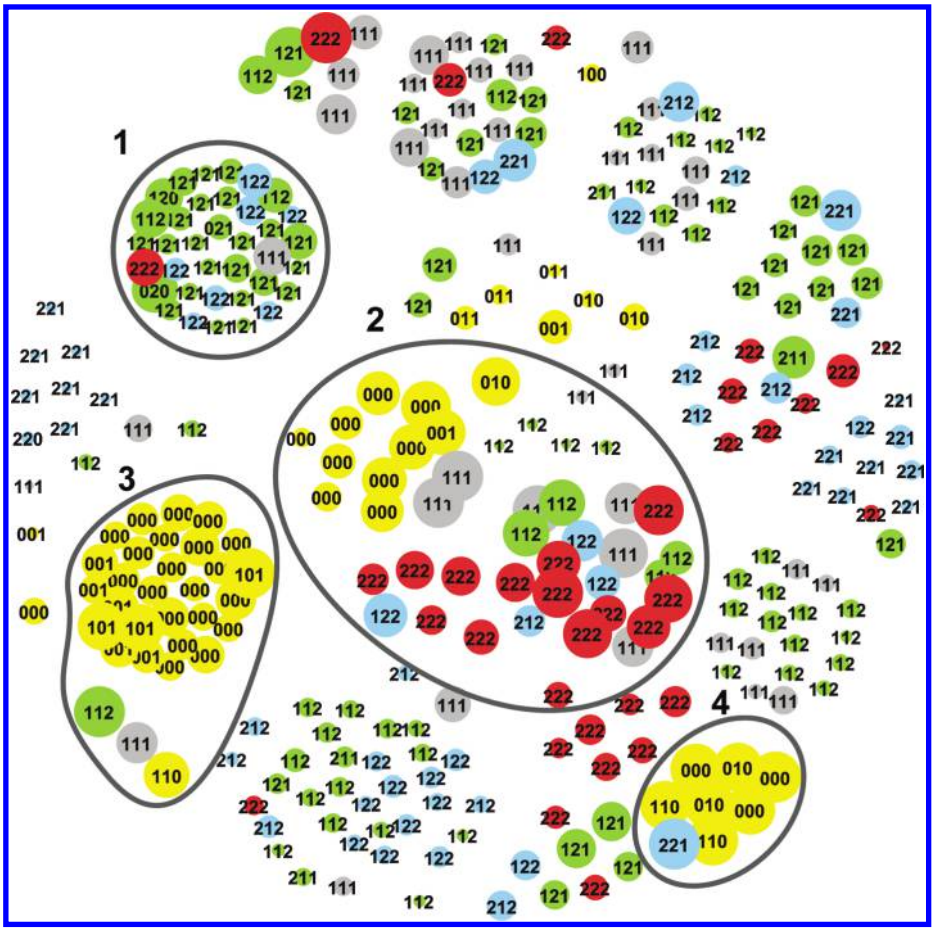


Figure 4. Multitarget graph. Shown is the multitarget activity landscape representation for the MT compound data set. Selected clusters are encircled and shown in detail in Figure 5.

Table 2. Activity Cliff Statistics[a]

				activity cliff distribution		
set	stc	dtc	ttc	type	count	degree
AR	121	9	0	212−002	84	single
				222− 102	15	single
				222−002	9	dual
MT	32	5	0	112−000	9	single
				122−120	7	single
				122−020	7	single
OR	54	4	0	2112−0001	26	single
				2112−0011	19	single
				2122−0001	4	dual
CA	49	16	15	2222−1012	14	single
				2222−0002	11	triple
				2122−0002	8	dual

[a] Four each compound data set, the number of single- (stc), dual- (dtc), and triple-target activity cliffs (ttc) is reported. In addition, "activity cliff distribution" reports the three most frequently occurring activity cliff types for each data set, and "degree" identifies single-, dual-, or triple-target cliffs.

numerous gray nodes (code 111), representing compounds of consistently moderate potency against multiple targets that cannot form activity cliffs according to our classification scheme.

However, the MT activity landscape also reveals clusters that are rich in differently colored nodes representing compounds with different activity profiles. Selected clusters characterized by the presence of rather different activity profiles are encircled in Figure 4. Such regions of heterogeneous node composition are prime candidates for activity cliff formation. Furthermore, the graph contains many differently sized nodes that indicate different compound contributions to multitarget SAR discontinuity, as discussed below.

Activity Cliff Distribution. We next identified activity cliffs contained in all four data sets. The activity cliff distributions are reported in Table 2. In the AR, MT, and OR sets, single- and dual-target cliffs were detected, and in the CA set, single-, dual-, and triple-target cliffs were identified. The activity cliffs were unevenly distributed. For each data set, they involved different target combinations, and not all targets were involved in the formation of cliffs. For AR, MT, OR, and CA, 121, 32, 54, and 49 single-target cliffs were found and 9, 5, 4, and 16 dual-target cliffs, respectively. In addition, for CA, 15 triple-target cliffs were identified. Thus, multitarget activity cliffs were more sparsely distributed than single-target cliffs in these compounds sets directed at closely related members of different protein families. We also found that certain activity profiles were more frequently involved in activity cliff formation than others, and consequently, some activity cliff types were preferentially formed. Table 2 reports the three most frequently occurring activity cliff types for

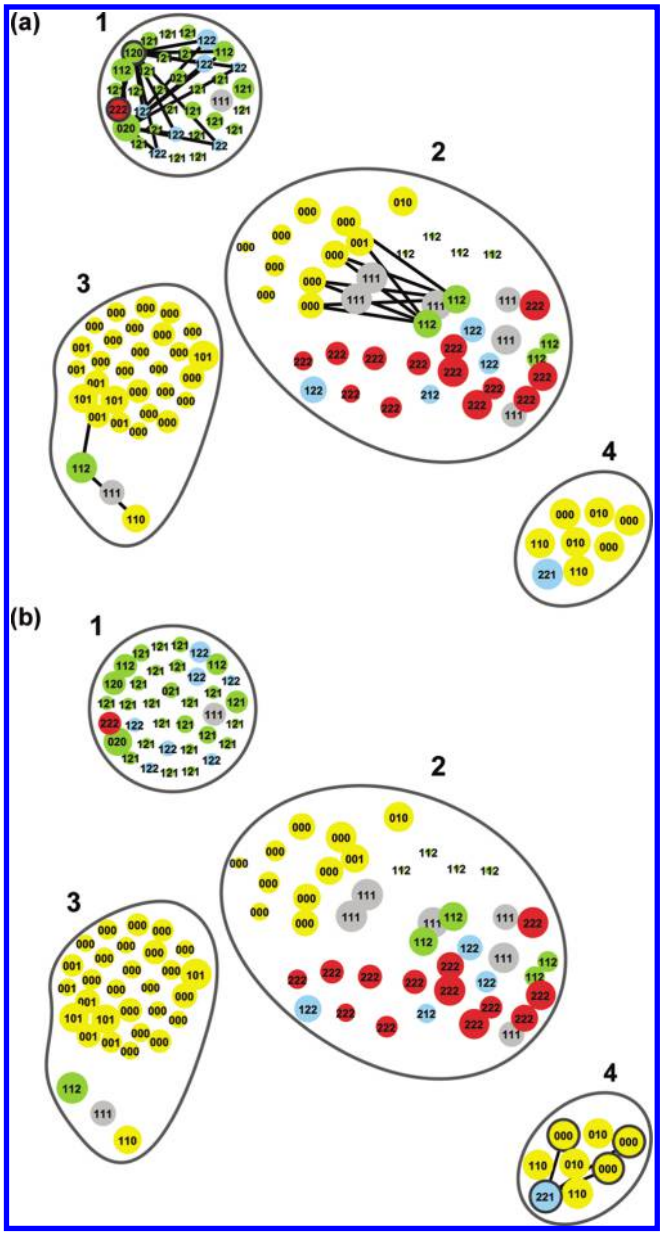


Figure 5. Compound clusters forming activity cliffs. For the MT data set, compound clusters are displayed where prominent single- or dual-target cliffs are formed. In (a), single-target cliffs are reported for the dopamine transporter, and in (b), dual-target cliffs are reported for the norepinephrine and serotonin transporter. Selected activity cliffs are highlighted, and the structures of the corresponding compounds and further details are shown in Figure 6.

each data set. Certain activity cliff types occurred with higher frequency than others, and frequent activity cliffs varied in a data set-specific manner.

Activity Cliff Patterns. We then identified compound clusters in the different data sets where activity cliffs mostly occurred. Figure 5 shows enlarged clusters from the multitarget graph of the MT data set (encircled in Figure 4) that contain single- (Figure 5a) and dual-target cliffs (Figure 5b). In addition, Figure S2 of the Supporting Information shows activity cliff-containing clusters for the other compound sets. In these representations, type-2 edges are displayed, each of which marks an activity cliff. For each data set, single-target activity cliffs involving an exemplary target are shown, and activity cliff views involving the remaining targets are provided in Figure S3 of the Supporting Information. For the AR, OR, and CA sets, detailed descriptions are given in the Results section of the Supporting Information. We generally observe that compounds have very different node sizes according to multitarget discontinuity scoring. In conventional single-target landscapes, compounds with large nodes typically have a potency that significantly differs from their structural neighbors, introduce local SAR discontinuity, and are most frequently involved in the formation of activity cliffs.[5] However, in multitarget landscapes, SAR discontinuity is a much more complex phenomenon because of the many different potency relationships that can result from comparisons of multitarget activity profiles. A characteristic feature of the activity cliff distributions in Figure 5 and Figure S2 of the Supporting Information is that compounds with large nodes (i.e., compounds that introduce notable multitarget SAR discontinuity) are often not involved in the formation of well-defined activity cliffs, but that cliffs are also formed by compounds having rather different node sizes. Thus, in multitarget activity landscapes, the introduction of SAR discontinuity and the formation of large-magnitude activity cliffs do not necessarily correlate because of the complexity of potency relationships between compounds active against multiple targets. However, similar to single-target landscapes, key compounds also emerge that make large contributions to multitarget SAR discontinuity and form multiple activity cliffs. For example, in the most densely connected region of the CA data set, the only compound set where triple-target cliffs were detected, overlapping yet distinct subsets of nodes form dual- (Figure S2f of the Supporting Information) and triple-target activity cliffs (Figure S2g of the Supporting Information). Most triple-target cliffs are formed by three large purple nodes and one large red node (code 2221) that also participate in the formation of dual-target cliffs. These compounds form prominent activity cliffs and, together with surrounding green nodes (code 0002 and 1002), make large contributions to SAR discontinuity. Furthermore, in the MT data set, a cluster is identified (cluster 3 in Figure 5b) where all compounds make large contributions to SAR discontinuity. Here, a compound represented by a blue node (code 221) forms three dual-target cliffs with different weakly potent compounds (code 000).

Such compounds that make large contributions to multitarget SAR discontinuity and also form prominent activity cliffs are a key component of multitarget activity landscapes and prime candidates for the exploration of multitarget SAR determinants. The analysis of structural features that distinguish these activity cliff markers and their activity profiles from each other, as discussed below, is of high relevance for practical applications.

Interpretation of Exemplary Activity Cliffs. Compounds forming exemplary single- and dual-target activity cliffs in the MT data set are shown in Figure 6, and compounds forming prominent activity cliffs in the other compound sets are reported in Figure S4 of the Supporting Information. In Figure 6a, exemplary MT single- and dual-target activity cliffs are displayed. The compound pair on the left forms a single-target cliff. Comparing the structures, it is evident that the removal of a carboxyl group renders a compound highly potent against all three transporters and hence nonselective. By contrast, the analogue containing the carboxyl group is selective for 5HT over DA and, to a lesser extent, NE. The compound pair on the right forms a dual-target cliff. Here, the replacement of an N-substituted piperidine ring with a chemically more complex seven-membered heteroaliphatic ring system leads to a significant change in the activity profile

dx.doi.org/10.1021/ci100477m |J. Chem. Inf. Model. 2011, 51, 258–266

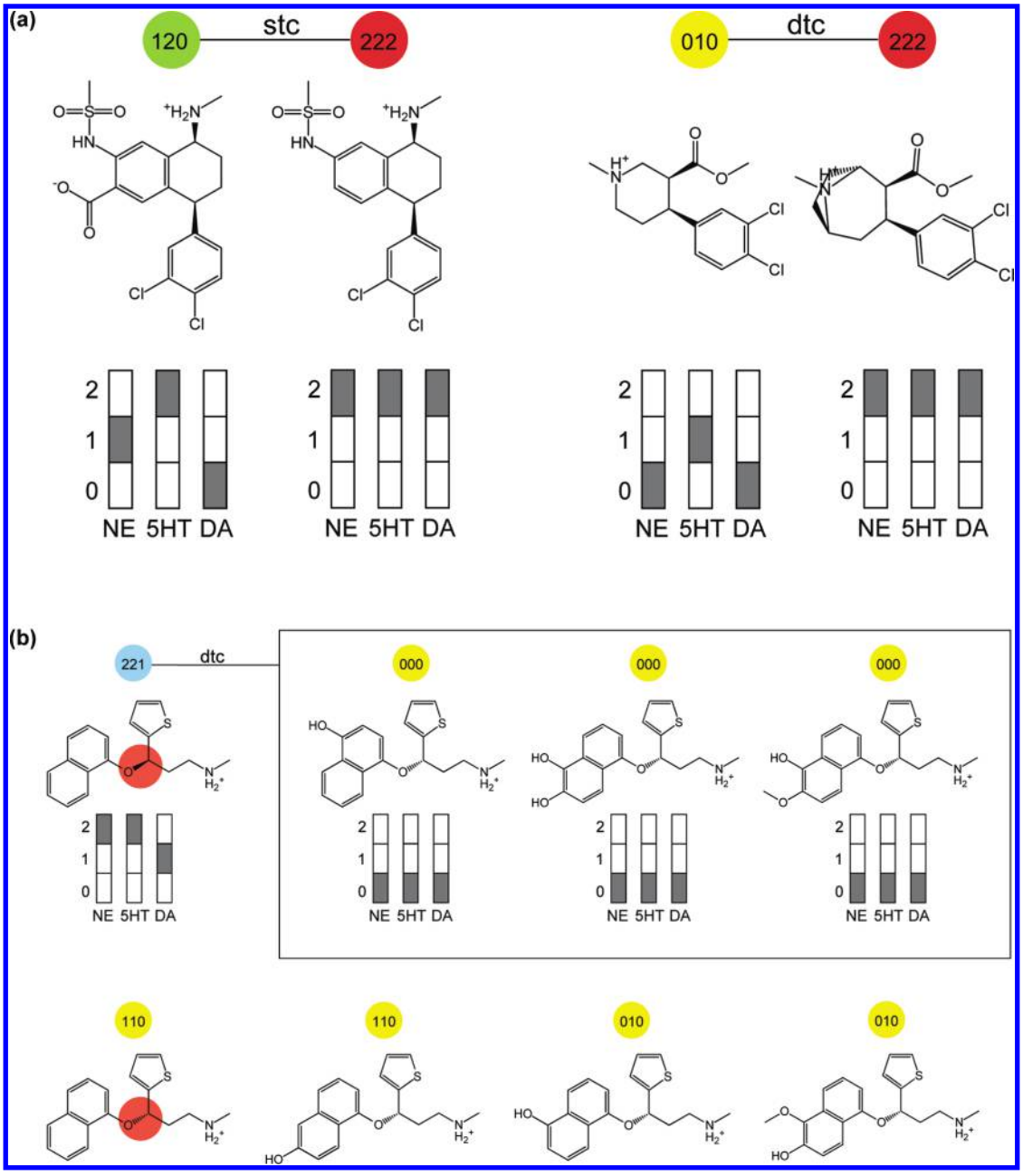


Figure 6. Exemplary activity cliffs. (a) Structures, node combinations, and activity profiles are shown for compounds forming representative single- and dual-target activity cliffs in the MT compound set. Target and activity cliff degree abbreviations are according to Table 1 and Table 2, respectively. (b) A compound involved in the formation of three dual target activity cliffs and its partner compounds are shown (see cluster 4 in Figures 4 and 5b). In addition, structures of compounds in the vicinity of these cliffs are also shown. The stereocenters of the two enantiomers with codes "221" and "110" are highlighted.

and is responsible for high potency against all three transporters. In Figure 6b, the key compound from cluster 4 in Figure 5b is shown (blue node, code 221) that is involved in the formation of three dual-target activity cliffs together with its cliff partners. Also shown are the four remaining compounds from this cluster that are moderately potent against NE and 5HT or only 5HT and not involved in the formation of activity cliffs. This compound series contains a conserved thiophene ring and a differently substituted naphthalene moiety. Comparing the structures of analogs involved in the formation of dual-target cliffs, it is evident that the weakly potent compounds differ from the highly potent one by two features, including single or dual hydroxyl (or hydroxyl and ether) substituents at the naphthalene ring and, in addition, different chirality of a carbon atom of the "ether bridge" connecting the naphthalene and thiophene moieties. Thus, by only comparing these four analogs, it cannot be concluded with certainty which structural changes might be responsible for the dramatic reduction in potency against all three transporters, leading to the formation of dual-target cliffs. However, by inspecting the structures of compounds in the vicinity of these activity cliffs it becomes clear that the chiral center plays a major role. All consistently weakly, or weakly and moderately, potent compounds from this cluster display the same chirality, and one of these compounds is the exact enantiomer of the potent key

dx.doi.org/10.1021/ci100477m |J. Chem. Inf. Model. 2011, 51, 258–266

compound represented by the blue node. Comparing the activity profiles of these enantiomers and other compounds in the cluster, further conclusions can be drawn. The code of the highly potent enantiomer is "221", and the code of the weakly potent one is "110". Thus, it appears that the stereochemical switch leads to a general reduction in potency against all three transporters, whereas hydrophilic substitutions at the naphthalene ring further reduce potency in a transporter-selective manner. Thus, the analysis of this activity cliff region alone in the MT data set provides helpful information about the multitarget structure−activity relationships of these compounds. We would conclude that hydrophilic substitutions of the high-potency "221" enantiomer might be a promising route to generate analogs with differentiated MT activity profiles. The identification of key compounds that provide interpretable structure−activity profile relationship information represents a prime application for multitarget activity landscape models, as introduced herein.

Concluding Remarks. Activity landscape models of compound data sets are obtained on the basis of systematic comparisons of structural and potency relationships. Landscape models are useful computational tools for the study of SAR features contained in compound sets and the identification of key compounds that determine local or global SAR characteristics.[1,2] For compounds active against an individual target, the generation and analysis of activity landscapes is straightforward, and different types of 2-D and 3-D representations of varying complexity have been introduced.[2] For compounds active against two targets, selectivity landscapes have also been generated. However, the design of multitarget activity landscapes is difficult to conceptualize and has remained an unsolved problem as of yet. Herein, we have introduced a first approach to construct multitarget activity landscapes that is based on a numerical encoding scheme of compound activity profiles derived from potency values against multiple targets. On the basis of systematic activity profile comparisons, we have derived a generally applicable formal organization of single-target and multitarget activity cliffs. The multitarget landscapes are displayed using a modified and extended version of network-like similarity graphs. In these representations, single-target and multitarget activity cliffs are easily identified. They are also formally defined by ternary code signatures that can be utilized for systematic mining of the data structure. Our multitarget activity landscape approach has been applied to characterize four compound data sets directed against three or four members of different target families. Compounds with confirmed activity annotations against more than four targets were difficult to find in public domain sources. As one might expect, compounds active against closely related protein family members often had similar activity profiles, and consequently, multitarget activity cliffs were generally more sparsely distributed than single-target cliffs. However, dual- or triple-target activity cliffs were readily identified in the activity landscapes of different data sets. In a number of instances, multitarget activity cliffs were centered on small sets of compounds that formed complex cliff patterns in multitarget graphs. The multitarget activity landscape approach and activity cliff hierarchy introduced herein is thought to provide a basis for the analysis of complex activity landscapes.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information. Figure S1 shows multitarget activity landscapes for the AR, OR, and CA data sets. Figure S2 shows selected clusters that form activity cliffs in the AR, OR, and CA compound sets. Figure S3 shows graph representations of single-target activity cliff distributions for all four data sets. Figure S4 shows exemplary activity cliffs for AR, OR, and CA. Results contain descriptions of multitarget activity landscape representations and activity cliff patterns for these three compound sets. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +49-228-2699-306; fax: +49-228-2699-341; e-mail: bajorath@bit.uni-bonn.de.

■ ACKNOWLEDGMENT

M.W. is supported by Boehringer-Ingelheim Pharma, Biberach, Germany.

■ REFERENCES

(1) Bajorath, J.; Peltason, L.; Wawer, M.; Guha, R.; Lajiness, M. S.; Van Drie, J. H. Navigating structure−activity landscapes. *Drug Discovery Today* **2009**, *14*, 698–705.

(2) Wassermann, A. M.; Wawer, M.; Bajorath, J. Activity landscape representations for structure−activity relationship analysis. *J. Med. Chem.* **2010**, *53*, 8209–8223.

(3) Shanmugasundaram, V.; Maggiora, G. M. Characterizing Property and Activity Landscapes Using an Information−Theoretic Approach. Proceedings of the 222nd American Chemical Society National Meeting, Division of Chemical Information, Chicago, IL, August 26−30, 2001; American Chemical Society: Washington, DC, 2001; Abstract No. 77.

(4) Guha, R.; Van Drie, J. H. Assessing how well a modeling protocol captures a structure−activity landscape. *J. Chem. Inf. Model.* **2008**, *48*, 1716–1728.

(5) Wawer, M.; Peltason, L.; Weskamp, N.; Teckentrup, A.; Bajorath, J. Structure−activity relationship anatomy by network-like similarity graphs and local structure−activity relationship indices. *J. Med. Chem.* **2008**, *51*, 6075–6084.

(6) Peltason, L.; Iyer, P.; Bajorath, J. Rationalizing three-dimensional activity landscapes and the influence of molecular representations on landscape topology and the formation of activity cliffs. *J. Chem. Inf. Model.* **2010**, *50*, 1021–1033.

(7) Maggiora, G. M.; Shanmugasundaram, V.; Lajiness, M. S.; Doman, T. N.; Schulz, M. W. A Practical Strategy for Directed Compound Acquisition. In *Chemoinformatics in Drug Discovery*; Oprea, T. I., Ed.; Wiley-VCH: Weinheim, Germany, 2005; pp 317−332.

(8) 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.

(9) Maggiora, G. M. On outliers and activity cliffs: Why QSAR often disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535–1535.

(10) Peltason, L.; Hu, Y.; Bajorath, J. From structure−activity to structure−selectivity relationships: Quantitative assessment, selectivity cliffs, and key compounds. *ChemMedChem* **2009**, *4*, 1864–1873.

(11) Wassermann, A. M.; Peltason, L.; Bajorath, J. Computational analysis of multi-target structure−activity relationships to derive preference orders for chemical modifications toward target selectivity. *ChemMedChem* **2010**, *5*, 847–858.

(12) Lounkine, E.; Wawer, M.; Wassermann, A. M.; Bajorath, J. SARANEA: A freely available program to mine structure−activity and structure−selectivity relationship information in compound data sets. *J. Chem. Inf. Model.* **2010**, *50*, 68–78.

(13) Fruchterman, T. M. J.; Reingold, E. M. Graph drawing by force-directed placement. *Software: Pract. Exper.* **1991**, *21*, 1129–1164.
(14) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
(15) ChEMBL. http://www.ebi.ac.uk/chembl (accessed July 1, 2010).