JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

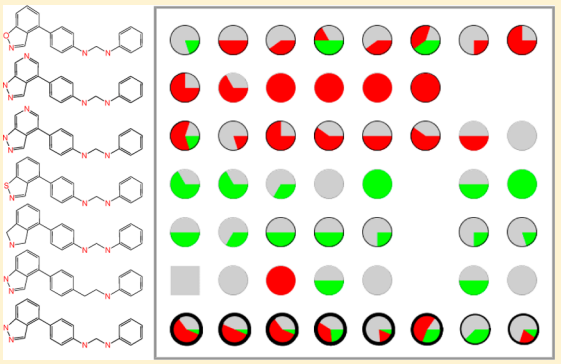
pubs.acs.org/jcim

Identification of Multitarget Activity Ridges in High-Dimensional Bioactivity Spaces

Disha Gupta-Ostermann 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

ABSTRACT: An activity cliff is defined as a pair of structurally similar compounds that have a large difference in potency against a given target. The activity cliff concept has recently been extended in different ways, including the introduction of the activity ridge data structure. An activity ridge consists of two subsets of highly and weakly potent structurally analogous compounds that form all possible pairwise activity cliffs between them. As such, the activity ridge data structure is rich in structure−activity relationship (SAR) information and attractive for SAR analysis. Activity ridges have been detected in various compound data sets. Analogously to single-target activity cliffs, activity rides have thus far only been investigated for individual targets. In this study, we have asked the question whether multitarget activity ridges might also exist. The analysis has been complicated by the limited availability of suitable compound profiling data sets in the public domain. However, in a high-dimensional kinase inhibitor data set recently released by Abbott Laboratories, multitarget activity ridges involving up to 43 different inhibitors and 26 kinase targets were identified. Given the inherently complex architecture of multitarget activity ridges, a new representation format was designed for these ridges based on a scaffold-target matrix. Furthermore, a scoring scheme was developed to identify compounds that were most variably distributed across a multitarget ridge and displayed target differentiation potential. Taken together, our results indicate that multitarget activity ridges represent an attractive data structure for SAR exploration of high-dimensional activity spaces.



INTRODUCTION

Activity cliffs were originally defined as pairs of structurally similar active compounds with large potency differences against their target.[1] In recent years, the activity cliff concept has experienced increasing popularity in compound data mining and medicinal chemistry.[2,3] A major reason for this popularity has been, and continues to be, that activity cliffs often reveal SAR information, given that small chemical changes are associated with large potency effects. Especially when large compound data sets are examined for interpretable SAR patterns, activity cliffs usually become an initial focal point of the analysis.[2,3] However, for a nonambiguous study of activity cliffs, structural similarity and potency difference criteria must be clearly defined.[2] Going beyond the analysis of activity cliffs at the level of compound pairs, it has also been observed that activity cliffs are often formed in a coordinated manner involving multiple compounds.[4] This means that subsets of highly and lowly potent structural analogs often form multiple overlapping cliffs. Such observations have given rise to the introduction of the "activity ridge" concept,[4] a higher-order activity cliff data structure. As illustrated in Figure 1, an activity ridge can be conceptualized to consist of a highly potent "upper" and a lowly potent "lower" compound layer in activity space, with all possible pairwise activity cliffs formed across these layers. The ridge data structure is attractive for SAR exploration because it is much richer in SAR information than activity cliffs studied in isolation. Albeit originally introduced as a hypothetical data structure, activity ridges have indeed been detected in a variety of compound sets. In a survey of 242 compound activity classes, a total of 125 activity ridges were detected in 71 of these data sets.[4] Furthermore, up to 70 active compounds were found to be involved in ridge formation, depending on the activity class.

Although activity cliffs have thus far been mostly considered for individual targets, multitarget activity cliffs have also been observed.[5] In its simplest version, a multitarget cliff is formed by a pair of compounds in which one is consistently highly potent against more than one target and the other consistently lowly potent against these targets.[5] However, multitarget activity cliffs are generally much less frequent than conventional single-target cliffs, as one might expect, because they can in principle only be formed by compounds that share multiple targets. In a systematic survey of activity cliffs in bioactive compounds, it was determined that only ∼4% of all detected activity cliffs were multitarget cliffs.[5]

Having identified single-target activity ridges and multitarget activity cliffs, a logical question has been whether multitarget activity ridges (MTARs) might also exist. However, source data that are suitable for this analysis have been difficult to obtain.

Received: August 6, 2012
Published: September 24, 2012

ACS Publications © 2012 American Chemical Society

dx.doi.org/10.1021/ci3003683 | J. Chem. Inf. Model. 2012, 52, 2579−2586

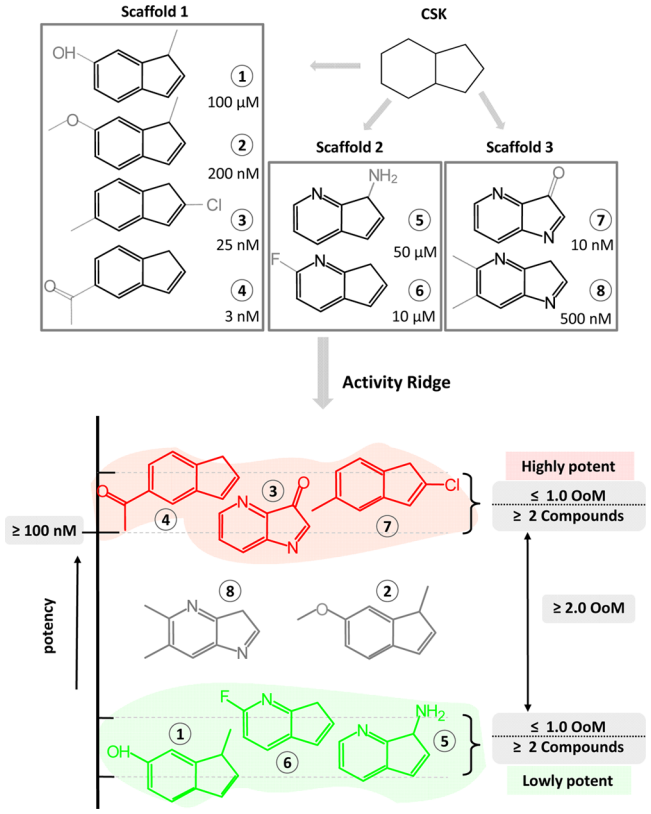


Figure 1. Activity ridge concept. A schematic illustration of the activity ridge concept is provided using a model compound set with hypothetical pK_i values. All compounds yield the same cyclic skeleton (CSK). The set is divided into three subsets of compounds, each of which contains a unique scaffold (black). This set forms an activity ridge that consists of three highly potent (red) and three lowly potent compounds (green). The potency difference between compounds comprising the upper (high potency) and lower (low potency) layer of the ridge must be at least 2 orders of magnitude (OoM). Hence, compounds 2 and 8 (gray), which fall within this potency interval, are not a part of the ridge. Highly and lowly potent ridge compounds form all possible pairwise activity cliffs across the layers.

Compounds in many activity classes are only annotated with single targets or limited numbers of targets, often with little overlap between them, giving rise to low multitarget activity cliff content. Therefore, we have been interested in exploring the potential formation of multitarget activity ridges in high-dimensional bioactivity spaces that typically arise from profiling of compound data sets against target families. Such profiling data sets are currently mostly generated in the pharmaceutical industry and rarely released to the public. Fortunately, a suitable high-dimensional data set has recently been made available by a group from Abbott Laboratories that contained 1496 kinase inhibitors with provided structures and exact activity measurements for varying numbers of 172 different kinases.[6] For individual inhibitors, activity annotations ranged from one to 122 kinase target. We have used this high-dimensional compound data set to search for multitarget activity ridges. The results of our analysis are presented herein.

METHODS AND MATERIALS

Bioactivity Data. From the Abbott data set,[6] 1449 inhibitors were selected, for which exact K_i measurements were available. These inhibitors had reported activities for one to 122 kinases. Inactive compounds are excluded from the analysis because they cannot contribute to ridge formation. If available, equilibrium constants are highly preferred for activity cliff and ridge analysis, as they display a clear tendency to yield more accurate cliff assignments than approximate measurements.[3] However, for the analysis of a single compound profiling experiment, IC_{50} values should also be acceptable because there would be no requirement for comparisons across different assays in this case.

Activity Ridge Criteria. Criteria for single- and multitarget activity ridges were applied essentially following the original activity ridge definition:[4]

1. Structural Similarity Criterion. Compounds with topologically equivalent Bemis and Murcko (BM) scaffolds[7] were considered candidates for activity cliff and ridge formation. Topologically equivalent scaffolds are only distinguished by heteroatom substitutions and bond orders; i.e., they yield the same cyclic skeleton (CSK).[8] This means that all compounds forming a ridge must share the same CSK. Each BM scaffold that participated in the formation of a ridge was individually considered and represented a specific subset of ridge compounds.

This structure-based similarity criterion has been shown to produce chemically intuitive results in activity cliff and ridge analysis[4] and limits potential ambiguities that are often associated with whole-molecule similarity values calculated on the basis of various descriptors.[2,3]

2. Activity Cliff Potency Difference Threshold. Compounds yielding the same CSK were considered to form an activity cliff, if their potency values differed by at least 2 orders of magnitude.

3. Upper Layer Potency Criterion. All compounds forming the upper (high potency) layer of an activity ridge were required to have a potency of at least 100 nM.

4. Upper and Lower Layer Potency Variation. Potency differences between all compounds in both the upper or lower layers had to be less than 1 order of magnitude.

5. Compound Number. A ridge had to contain at least two highly and two lowly potent compounds meeting the criteria specified above.

Ridge Search and Selection. A search routine was implemented in Java[9] using the OpenEye toolkit[10] to identify all compound series meeting the structural similarity criterion that contained compound subsets forming activity ridges (satisfying criterion 1−5) against multiple targets. A previously reported[4] two-step search procedure was then applied to automatically select the largest possible ridges from qualifying compound series. In the first step, highly and lowly potent compound subsets with at least 100-fold potency difference were assembled such that the product of their compound numbers was maximal. In the second step, the largest possible numbers of highly and lowly potent compounds were selected from these subsets whose potency values were within 1 order of magnitude in each ridge layer. When multiple selection sets of equal size were obtained, the one with the largest potency difference between highly and lowly potent compounds was retained. Following the two-step selection procedure, ridge formation was consistently assessed, and large ridges were prioritized.

Ridge Visualization. For the representation of MTARs, a scaffold-target matrix was designed with annotated nodes representing scaffold-based subsets of compounds active against individual targets. Nodes in the matrix were graphically

dx.doi.org/10.1021/ci3003683 | J. Chem. Inf. Model. 2012, 52, 2579−2586

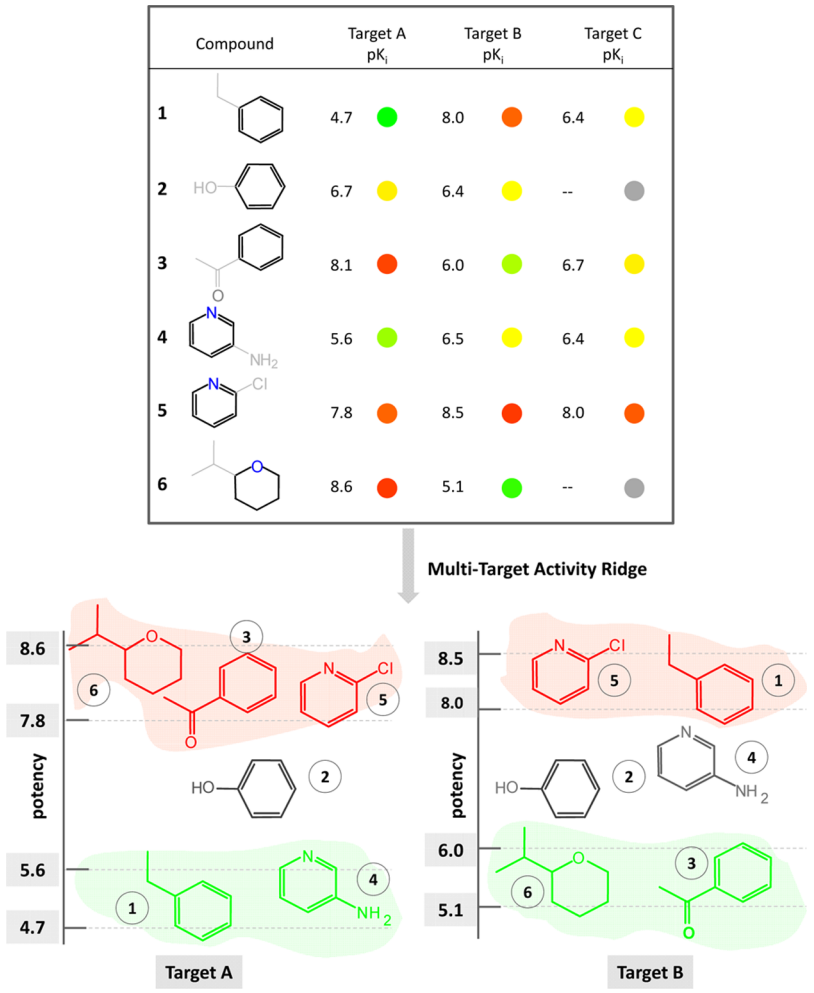


Figure 2. Multitarget activity ridge. A simple model data set sharing the same (cyclohexane) CSK is used to illustrate the MTAR concept. Heteroatoms that distinguish between different scaffolds are colored blue. The compounds are active against two or three targets A, B, and C. The pK_i values are given as colored nodes using a continuous color spectrum from green (low) to red (high potency). The node array represents the activity profile of the compound set. A gray node indicates that a compound is not active against a given target. For targets A and B, a total of five and four analogs, respectively, are involved in the formation of individual activity ridges (represented according to Figure 1). The formation of activity ridges of compound subsets of a series against two or more different targets leads to a multitarget activity ridge.

associated with the corresponding scaffold structures and the number of analogs contained in each subset. The graphical representation was also implemented in Java[9] using the OpenEye toolkit.[10] It is described in detail in the Results and Discussion section.

Potency Range Distribution Score. In order to quantitatively assess the potency range distribution of compounds across an MTAR, a potency range distribution (PRD) score was defined as follows:

$$\underset{i \in w \| h}{\mathrm{PRD}(i)} = \frac{\mathrm{freq}_w - \mathrm{freq}_h}{\mathrm{freq}_w + \mathrm{freq}_h}$$

Here, i denotes a compound in upper (h) or lower (w) layers of the MTAR, and freq is the frequency of occurrence of i in these layers across the different targets. The resulting PRD score ranges from −1 to 1. A score of −1 indicates that a compound exclusively occurs in the upper layers of the ridges of all targets and a score of 1 that it exclusively occurs in lower ridges. A score of 0 indicates that the compound is equally distributed across the upper and lower layers, i.e., scores close to 0 identify compounds that are highly and lowly potent against different targets.

RESULTS AND DISCUSSION

Multitarget Activity Ridge. In Figure 2, the MTAR concept is illustrated. An MTAR represents the combination of multiple single-target ridges. The composition of an MTAR might be highly complex because overlapping yet distinct subsets of active compounds might occur in upper or lower layers of ridges for different targets, depending on the compound activity profiles, as illustrated in Figure 2. Although individual ridges in high-dimensional activity space share the same architectural features, they can differ significantly in the distribution and magnitude of activity cliffs they contain.

MTAR Search. Given the low propensity of multitarget activity cliffs in standard compound activity classes, it was uncertain whether multitarget activity ridges might be formed. Furthermore, the profiling matrix of inhibitors with disclosed structures in the kinase data set was only sparsely populated; i.e., many compounds had reported activities against very different numbers of kinases, and their activity profiles only partly overlapped. However, in our systematic search for single-

dx.doi.org/10.1021/ci3003683 | J. Chem. Inf. Model. 2012, 52, 2579−2586

and multitarget ridges in this data set, a total of 17 activity ridges were detected, as reported in Table 1. Taken together,

Table 1. Activity Ridge Composition

ridge no.	number of targets	number of scaffolds	number of compounds
1	1	1	5
2	1	5	4
3	1	6	7
4	1	7	6
5	1	10	6
6	2	3	10
7	2	4	7
8	2	14	18
9	3	2	13
10	3	5	11
11	5	5	22
12	6	3	18
13	7	15	25
14	8	7	43
15	16	16	40
16	17	5	21
17	26	10	37

the formation of these ridges involved a total of 293 inhibitors that corresponded to 118 distinct BM scaffolds (and 17 CSKs) and a total of 66 different kinases. Interestingly, five of these activity ridges were single-target ridges (numbers 1−5 in Table 1), and 12 were MTARs (numbers 6−17). Hence, in this high-dimensional data set, MTARs occurred with higher frequency than single-target ridges. The formation of the 12 MTARs involved two to 26 different kinases with 7−43 compounds per ridge (with a mean of ∼22) and varying numbers of scaffolds, as reported in Table 1. Seven MTARs involved between five and 26 different kinases, which we considered a rather unexpected finding. The composition of these MTARs was often complex, including scaffold-based compound subsets of different sizes and degrees of overlap for many targets.

MTAR Visualization. Given this complexity, a suitable visualization scheme had to be designed to represent MTARs in an informative manner. The design principles of the visualization method are summarized in Figure 3. The visualization is designed to convey MTAR information at the levels of BM scaffolds (not individual compounds). The basic data structure for representation is a scaffold-target matrix (Figure 3a) in which each of the topologically equivalent scaffolds of an MTAR is assigned to a row of the matrix. Accordingly, the CSK-based compound set forming the MTAR is divided into scaffold-based subsets. Furthermore, each target involved in the formation of the MTAR is assigned to a column of the matrix. As an alternative to the scaffold-based compound subset organization introduced here, which is tailored toward SAR information extraction, one might also consider organizing compounds on the basis of their target activity profiles. However, for MTAR analysis, a structural organization scheme would still be required. For different targets, the composition of a scaffold-based compound subset that they are active against might differ because not all inhibitors are active against all targets, and the potency of inhibitors against different targets might vary, giving rise to the complexity of the MTAR. Therefore, scaffold-based compound subsets are represented as nodes that reveal the assignment of compounds to the upper (red) and lower (green) layer of the ridge of each individual

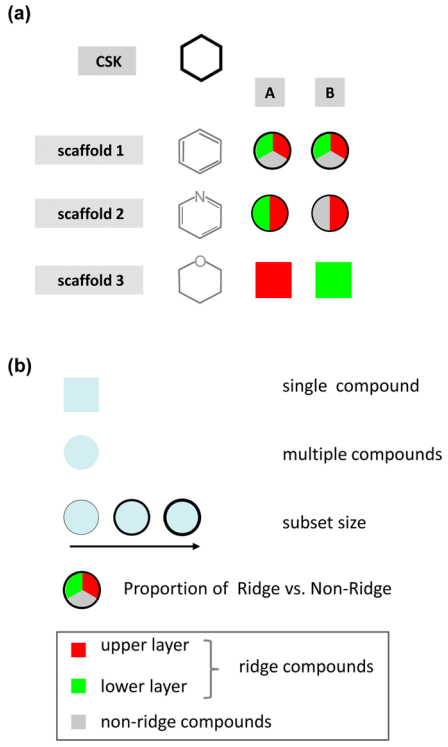


Figure 3. MTAR display. Given the potential complexity of MTARs, they are graphically represented as a scaffold-target matrix, as schematically illustrated in part a. All compounds a scaffold represents are combined into a scaffold-based compound subset. Each scaffold (1−3) is assigned to a row of the matrix, and targets (A and B) are assigned to columns. Scaffold-based compound subsets that are active against each target are represented as nodes whose design is illustrated in part b. Single compounds are shown as squares and multiple compounds as circular nodes. Each node contains a color-coded pie chart representation that reveals the proportion of ridge vs nonridge compounds and compounds in the upper or lower layer of the ridge. The thickness of node borders is scaled according to the relative size of a scaffold-based compound subset.

target (Figure 3b). In addition, each scaffold-based subset displayed in the matrix contains all compounds active against a given target that are not part of the ridge (gray). Thus, the pie chart representation of nodes also reveals the proportion of MTAR compounds among all active compounds in the ridge of each individual target. Furthermore, the thickness of the node borders of scaffold-based compound subsets is scaled according to the size of each subset relative to the total number of compounds in the CSK-based set. As illustrated in the following, the annotated scaffold-target matrix format provides a consistent and intuitive representation of MTARs, regardless of their composition and complexity.

Exemplary MTARs. In Figure 4, two of the MTARs we identified are displayed. Figure 4a shows an MTAR involving eight different kinase targets and a total of 43 inhibitors corresponding to seven topologically equivalent scaffolds. The targets are arranged in alphabetical order. In Figure 4b, a corresponding representation is shown in which targets are ordered according to the Tanimoto similarity of their compound activity profiles (i.e., reflecting nearest neighbor relationships). For the display of complex matrices, this target organization might also be useful. Figure 4c shows representative compounds that participated in the formation of this MTAR. It is evident that a number of different (albeit

dx.doi.org/10.1021/ci3003683 | J. Chem. Inf. Model. 2012, 52, 2579−2586

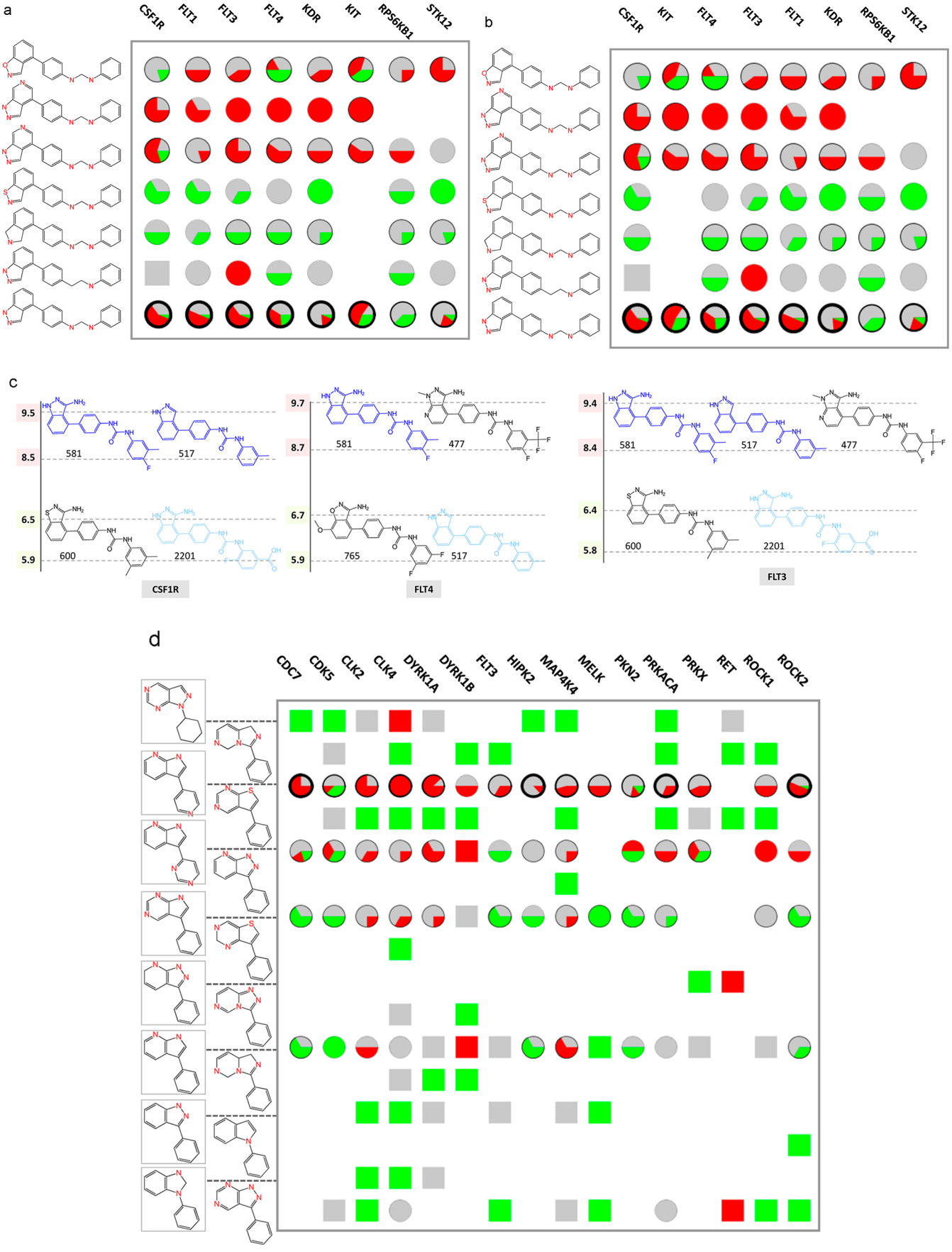


Figure 4. continued

dx.doi.org/10.1021/ci3003683 | J. Chem. Inf. Model. 2012, 52, 2579−2586

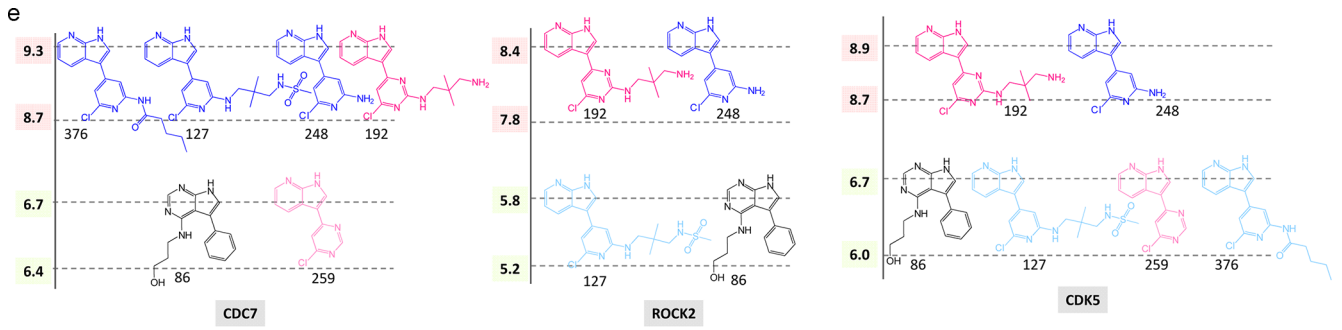


Figure 4. Exemplary MTARs. (a) MTAR involving eight different kinase targets (abbreviated according to ref 6 in the column header). This MTAR is formed by compounds having seven unique scaffolds and is represented according to Figure 3. Heteroatoms that distinguish between these scaffolds are colored red. Targets are arranged in alphabetical order. In part b, a corresponding representation is shown in which targets are ordered according to the similarity of their compound activity profiles. (c) Representative highly and lowly potent compounds from the MTAR in part a are shown for three kinases (CSFR1, FLT4, and FLT3). The set ID of each compound is given. Highly potent and weakly potent inhibitors sharing the same scaffold are colored dark blue and light blue, respectively. Other compounds are shown in black. Potency ranges of compounds forming the upper and lower layers of each ridge are reported (pK_i values). (d) MTAR involving 16 kinase targets formed by compounds with 16 unique scaffolds. Heteroatoms that distinguish between these scaffolds are colored red. (e) Representative highly and lowly potent compounds from the MTAR in part c are shown for three kinases (CDC7, ROCK2, and CDK5). Here, two different scaffolds represent both highly and lowly potent compounds that are shown in dark/light blue and dark/light pink, respectively.

chemically closely related) scaffolds mostly represented compounds that were either highly or lowly potent against different kinases. In this case, scaffold-based compound subsets that contained lowly potent inhibitors were often enriched with non-MTAR compounds. However, the scaffold-based compound subset in the top row and especially the bottom row of the matrix displayed significant variations in the proportion of upper layer, lower layer, and nonridge compounds, depending on the targets. Hence, these scaffolds represented compounds with differentiation potential across different kinases. The scaffold-based subsets in the bottom row were also the largest within this MTAR. Hence, this scaffold was extensively explored through analog design, probably because initially obtained compounds might have acted differently against their kinase targets, more so than other MTAR compounds. The trends described above are also reflected in Figure 4c. Most inhibitors consistently appeared in either the upper or lower layer of individual ridges. However, there were exceptions. For example, compound 517 switched between upper and lower ridge layers, depending on the target. Hence, this compound was able to differentiate between these kinases.

In Figure 4d, another MTAR is displayed that involved 16 different kinases and a total of 37 inhibitors corresponding to 16 scaffolds. Figure 4e shows representative compounds belonging to this MTAR. The matrix of this MTAR was more sparsely populated than the one shown in Figure 4a. A number of scaffolds had only been little explored. In many instances, a scaffold-based subset contained only a single compound with similar potency against different kinases. Here, a number of scaffold-based compound subsets that contained upper layer compounds were enriched with nonridge analogs, different from the MTAR in Figure 4a. Only the scaffold in the third row of the matrix was extensively explored. For a number of kinases, its scaffold-based compound subsets also displayed significant variations in the proportion of upper layer, lower layer, and nonridge compounds, which was indicative of target differentiation potential. The layer distribution of exemplary MTAR compounds in Figure 4e was similar to that in Figure 4c. Most compounds also exclusively occurred in the upper or lower layer of individual ridges and displayed no differentiation potential for these kinases. The exception was compound 127, which switched between upper and lower ridge layers of different kinases.

Potency Range Distribution. In order to quantify the observations made above for all MTAR compounds, PDR scores were systematically calculated for these inhibitors. The distribution is reported in Figure 5a. The majority of compounds were positioned in the upper (scores close to −1) or lower layers (scores close to 1) of MTARs, and scaffolds represented mostly highly or lowly potent compounds against different kinases. However, there also were notable exceptions. For example, the score interval [0, 0.20] in Figure 5a contained ∼25 compounds. These compounds had scores close to 0. Thus, they were almost equally distributed across upper and lower ridge layers. Within this data set, these compounds represented some of the most interesting inhibitors because they might be used as chemical probes to differentiate between kinase targets. Figure 5b shows the potency distribution of inhibitors taken from the −0.25 to 0.25 score range against selected kinases, which reveals the presence of comparable potency differences of inhibitors against different kinases. In addition, it also shows that two of these compounds (431 and 569) have a potency difference of nearly 3 orders of magnitude against an individual kinase (CSNK1G1). Taken together, the differentiation potential of these compounds is evident. Finally, in Figure 5c, the target distribution of compounds with consistently low (green) and high (red) potency is compared to the distribution of compounds within the −0.33 to 0.33 score range (blue), i.e., inhibitors with most variable potency distribution in this data set. The majority of lowly potent compounds were only active against individual targets. Highly potent compounds were often annotated against larger numbers of targets. More than 10 MTAR compounds that were consistently highly potent acted against four to six different kinases. Furthermore, 30 inhibitors with the most variable potency were active against two to three targets. However, 10 other inhibitors with variable potency were annotated with four to nine targets. Hence, these compounds represented the most interesting candidates for molecular probes among the 265 MTAR compounds we identified in the kinase data set.

dx.doi.org/10.1021/ci3003683 | J. Chem. Inf. Model. 2012, 52, 2579−2586

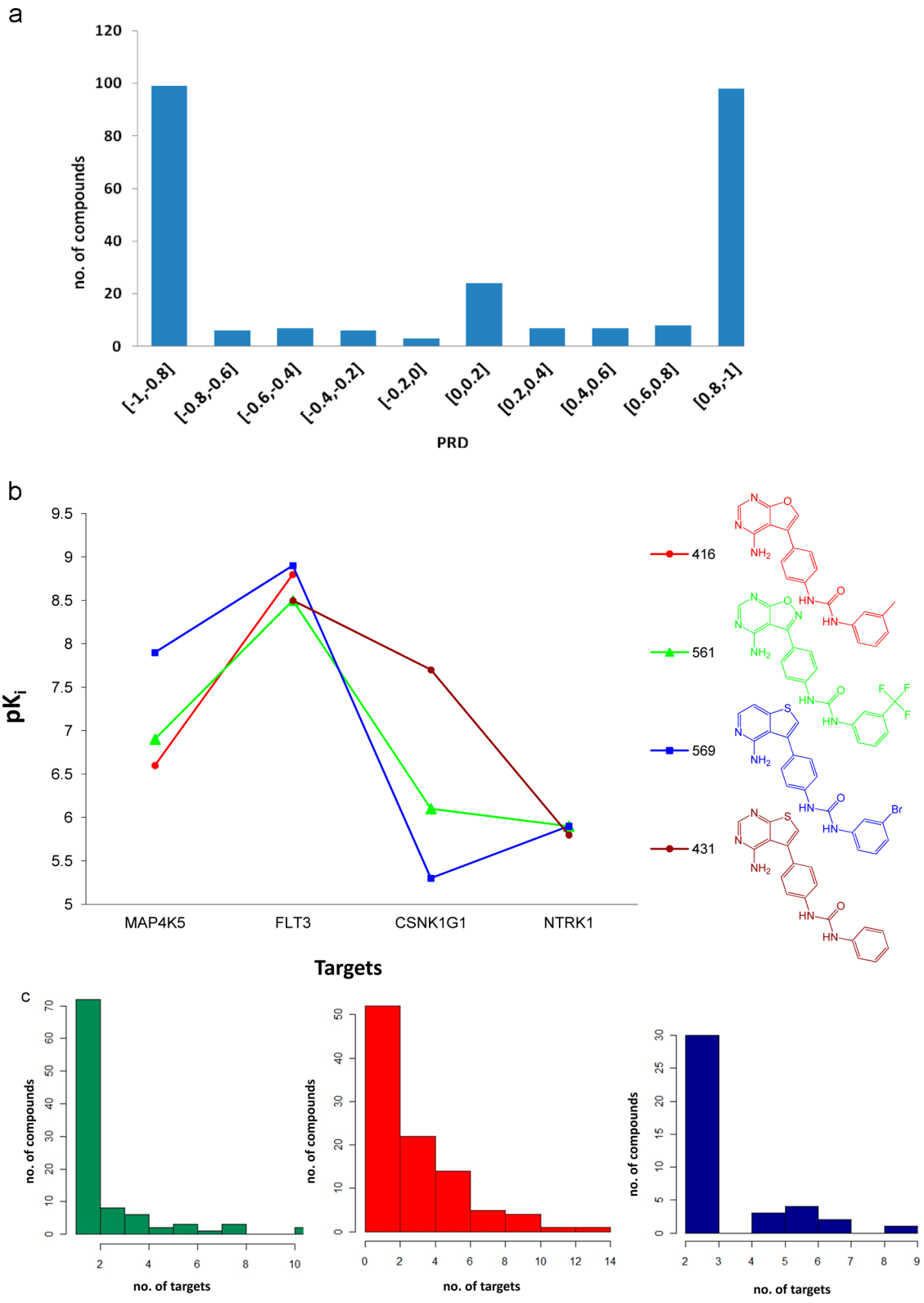


Figure 5. Potency range distribution scores. (a) The frequency of PRD scores (divided into score intervals) is reported for the kinase data set. (b) pK_i values of four compounds having a PRD score of −0.25 to 0.25 are compared for four kinases. (c) Shown is the target distribution of compounds with consistently low (green; PRD score range [0.80, 1]), high (red; [−1,-0.80]), or varying potency (blue; [-0.33, 0.33]).

MTAR Information and SAR Analysis. The methodology introduced herein makes it possible to systematically search for MTARs in compound profiling data sets. As stated above, MTARs are particularly rich in SAR information because activity cliffs are systematically formed between groups of structurally similar compounds with different potency. Hence, from an activity ridge, structural modifications can be extracted that result in high compound potency. Such modifications can then be explored in the structural context of related molecular scaffolds. Extending the activity ridge concept to high-

dx.doi.org/10.1021/ci3003683 | J. Chem. Inf. Model. 2012, 52, 2579−2586

dimensional activity spaces, which result from compound profiling experiments, has further application potential for multitarget SAR analysis. In addition to studying activity determinants for individual targets, it is now also possible to identify compounds that are variably distributed across potency layers of MTARs and thus have target differentiation potential. By comparing such compounds with structural analogs in different layers of MTARs, it might be possible to identify structural changes that render compounds selective for individual targets over others. Such apparent differentiation and/or selectivity determinants can then also be explored on the basis of different structural frameworks, thereby utilizing results of retrospective SAR analysis for compound design.

CONCLUSIONS

Herein, we have introduced the multitarget activity ridge data structure. A search procedure was implemented to automatically extract MTARs from high-dimensional bioactivity spaces. In our analysis, we focused on a publicly available kinase inhibitor profiling data set. Although this set comprised ~1500 compounds with reported activities against varying numbers of up to 122 different kinases, it represented an overall only sparsely populated compound profiling matrix. However, we identified 12 MTARs in this data set (in addition to five single-target activity ridges). The formation of these MTARs involved ~18% of all inhibitors in the set and 66 different kinases. On the basis of these findings, it is very likely that MTARs are also found in other high-dimensional data sets. MTARs can generally be considered as an information-rich data structure for the study of multitarget SARs. However, given their potential complexity, MTARs are not straightforward to visualize and study, and we therefore have designed a special representation scheme for MTARs based on an annotated scaffold-target matrix. Using this format, the composition of MTARs can be readily visualized and compared. Furthermore, a scoring function was introduced to identify compounds with the most variable distribution across upper and lower MTAR layers. Such compounds have target differentiation potential and are thus prime candidates for further analysis. Taken together, the MTAR data structure represents a further extension of the activity cliff concept that should be of interest for mining high-dimensional bioactivity spaces and exploring multitarget SARs.

AUTHOR INFORMATION

Corresponding Author

*Tel: +49-228-2699-306. Fax: +49-228-2699-341. E-mail: bajorath@bit.uni-bonn.de.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to thank Ye Hu, Anne Mai Wassermann, and Preeti Iyer for helpful discussions.

REFERENCES

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

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

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

(4) Vogt, M.; Huang, Y.; Bajorath, J. From Activity Cliffs to Activity Ridges: Informative Data Structures for SAR Analysis. *J. Chem. Inf. Model.* **2011**, *51*, 1848–1856.

(5) Wassermann, A. M.; Dimova, D.; Bajorath, J. Comprehensive Analysis of Single- and Multi-Target Activity Cliffs Formed by Currently Available Bioactive Compounds. *Chem. Biol. Drug Des.* **2011**, *78*, 224–228.

(6) Metz, J. T.; Johnson, E. F.; Soni, N. B.; Merta, P. J.; Kifle, L.; Hajduk, P. J. Navigating the Kinome. *Nature Chem. Biol.* **2011**, *7*, 200–202.

(7) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.

(8) Xu, Y.-J.; Johnson, M. Using Molecular Equivalence Numbers to Visually Explore Structural Features that Distinguish Chemical Libraries. *J. Med. Chem.* **2002**, *42*, 912–926.

(9) Java Universal Network/Graph Framework, version 2.0.1. http://jung.sourceforge.net/ (accessed Jan 11, 2012).

(10) *OEChem TK*, version 1.7.4.3; OpenEye Scientific Software Inc.: Santa Fe, NM, 2010.