Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

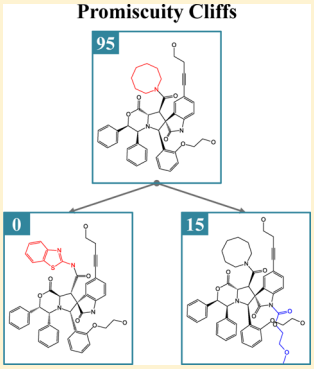
Matched Molecular Pair Analysis of Small Molecule Microarray Data Identifies Promiscuity Cliffs and Reveals Molecular Origins of Extreme Compound Promiscuity

Dilyana Dimova,[†] Ye Hu,[†] 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: The study of compound promiscuity is a hot topic in medicinal chemistry and drug discovery research. Promiscuous compounds are increasingly identified, but the molecular basis of promiscuity is currently only little understood. Utilizing the matched molecular pair formalism, we have analyzed patterns of compound promiscuity in a publicly available small molecule microarray data set. On the basis of our analysis, we introduce "promiscuity cliffs" as pairs of structural analogs with single-site substitutions that lead to large-magnitude differences in apparent compound promiscuity involving between 50 and 97 unrelated targets. No substructures or substructure transformations have been detected that are generally responsible for introducing promiscuity. However, within a given structural context, small chemical replacements were found to lead to dramatic promiscuity effects. On the basis of our analysis, promiscuity is not an inherent feature of molecular scaffolds but can be induced by small chemical substitutions. Promiscuity cliffs provide immediate access to such modifications.



INTRODUCTION

Target promiscuity of small molecules is a much investigated topic in medicinal chemistry, for several reasons. First, the binding behavior of a promiscuous compound might be associated with nonspecific binding events, as exemplified by frequent hitters in biological screens.[1] Second, specific interactions of compounds with multiple (related or unrelated) targets might give rise to polypharmacological behavior[2−5] and also provide a basis for drug repurposing.[6,7] Third, increasing evidence that many bioactive compounds do act on multiple targets is beginning to change the single-target specificity paradigm that has long governed drug discovery and design efforts.[8−11] Previous studies have mostly addressed compound promiscuity through database mining,[5,12,13] for example, by identifying molecular scaffolds that are recurrent in promiscuous compounds,[12] or have focused on polypharmacology by detecting new targets for existing drugs[5] and by studying side effects.[13]

Most information about compound promiscuity is currently obtained from target annotations of bioactive compounds collected from literature resources and stored in major compound data repositories, such as ChEMBL.[14] In addition, promiscuity information might also be obtained by comparing screening libraries across different bioassays available in PubChem,[15] although this information is limited at present and principally confined to screening hits. Compound promiscuity can experimentally be assessed by systematically testing compound collections on arrays of diverse targets. Unfortunately, such compound profiling data is currently rarely available, at least in the public domain. However, there are a few notable exceptions. For example, a data set recently released by a group from Abbott Laboratories contains 1473 compounds with reported activities against 1−122 different kinases from a representative sample of the kinome.[16] While this data set provides an excellent test case for large-scale SAR exploration,[17] it is not suitable for promiscuity analysis beyond kinases. Furthermore, Schreiber and colleagues have reported a small molecule microarray experiment that involved screening of diverse compounds against a total of 100 sequence-unrelated targets.[18] The data released as a part of this investigation are highly attractive for a systematic assessment of compound promiscuity. In their original study, Clemons et al. assembled a total of 15 252 compounds from three different sources including compounds commercially available from medicinal chemistry vendors (CCs), natural products (NPs), and compounds originating from diversity-oriented synthesis (DCs).[18] These compounds were then printed on glass slides through surface chemistry or noncovalent absorption and tested against 100 sequence-unrelated soluble proteins. These proteins were selected to represent a total of 145 different InterPro domain classification types.[19] Purified tacked proteins were incubated on microarrays, and proteins bound to array compounds were detected with labeled monoclonal antibodies. These experiments produced a binary readout of activity, i.e., a compound was classified as active against a target or not. Hence, given the nature of microarray experiments, no exact activity measurements were obtained. However, these data

Received: September 7, 2012
Published: October 10, 2012

dx.doi.org/10.1021/jm301292a | J. Med. Chem. 2012, 55, 10220−10228

reflect binding patterns of compounds across a large array of different targets and are thus suitable for the analysis of compound promiscuity or specificity. Clemons et al. determined the distribution of active compounds and analyzed their data primarily considering measures of stereochemical and shape complexity. They found that NPs generally yielded lower hit rates than synthetic compounds and that both NPs and DCs produced many more specific hits than CCs. Increasing stereochemical and shape complexity generally favored compound specificity, as one might anticipate. However, it was also observed that 16% of CCs and 3% of DCs were promiscuous in nature. Clemons et al. found that a spirooxindole moiety was recurrent in the promiscuous subset of DCs. By contrast, possible structural origins of promiscuity among CCs did not become apparent in the course of the analysis. However, a key finding has been that compounds with apparent target selectivity were clearly enriched among DCs compared to CCs.[18]

We have been interested in exploring compound promiscuity from a structural perspective, encouraged by the microarray analysis efforts of Schreiber and colleagues involving 100 sequence-unrelated targets. For a thorough structural assessment, we have carried out a matched molecular pair (MMP) analysis[20] of all compounds in this data set. We reasoned that MMPs might provide direct access to structural features implicated in promiscuity because compounds forming an MMP are only distinguished by the exchange of a single substructure with limited size. On the basis of our analysis, structural relationships between nonpromiscuous and highly promiscuous compounds were established and substructures were identified that induced large-magnitude promiscuity within a given structural context. MMPs included compounds with very large differences in the number of targets they were active against, leading to the introduction of promiscuity cliffs.

MATERIALS AND METHODS

Compound Data. The publicly released microarray data set[18] contained 15 252 compounds. Each compound was screened against 100 sequence-unrelated proteins. A total of 3433 compounds were active against 1−97 proteins. Compound structures were examined and standardized using the Molecular Operating Environment[21] and transformed into SMILES strings.[22] Compounds with unique SMILES strings were retained. Following these procedures, 15 042 compounds remained for MMP generation including 6151 CCs, 6437 DCs, and 2454 NPs.

Matched Molecular Pair Analysis. An MMP is defined as a pair of compounds that only differ by a structural change at a single site,[20,23] as illustrated in Figure 1. Compounds forming an MMP are interconverted by the exchange of two substructures, which is termed a chemical transformation.[23] Accordingly, the MMP formalism is descriptor-independent, metric-free, and chemically intuitive. For example, it has been applied to characterize activity cliffs and bioisosteric replacements.[24−26] MMPs were generated using an in-house implementation of the Hussain and Rea algorithm.[23] Following this approach, conserved core structures and variable substituents of MMPs are stored as keys and values in an index table, respectively. The size of an exchanged substructure (value) was limited to maximally 13 non-hydrogen atoms and the size difference between exchanged substructures to maximally eight non-hydrogen atoms. This was done to restrict the size of exchanged fragments to chemically meaningful replacements.[26] In addition, MMP formation was further restricted by the requirement that the core structure of a qualifying compound (key) had to be at least twice the size of each exchanged substructure (value). Application of these size restrictions previously yielded chemically intuitive transformations in an MMP-based study of activity cliffs.[26] Furthermore, if several transformations generated the same MMP, only the transformation comprising the smallest number of atoms was retained. Following this protocol, MMPs were systematically generated for all 15 042 microarray compounds.

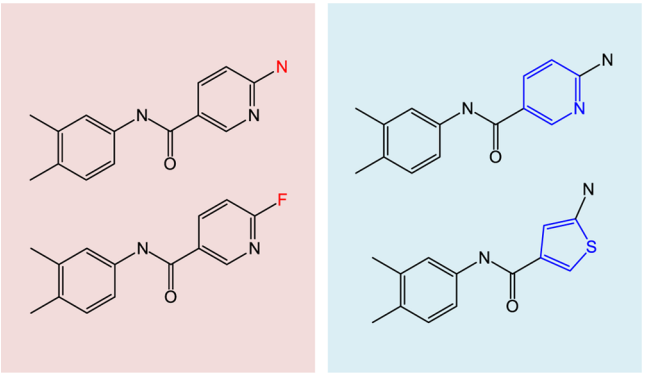


Figure 1. Matched molecular pairs. Two pairs of compounds forming exemplary MMPs are shown. Exchanged fragments are colored in red (left) or blue (right).

All MMP and data-mining calculations were carried out with in-house generated Java programs or KNIME[27] protocols. An MMP-based compound network was drawn with Cytoscape.[28]

Promiscuity Cliff Criteria. On the basis of our analysis, so-called "promiscuity cliffs" were introduced by applying the following criteria:

(1) A compound pair formed a transformation size-restricted MMP (as explained above).
(2) The number of activity annotations of the compounds forming an MMP differed by at least 50 targets, hence indicating large-scale differences in apparent promiscuity.

Accordingly, promiscuity cliffs represented closely related compounds (mostly analogs) with limited structural variations, but large differences in the number of target annotations. These cliffs were systematically explored in the small molecule microarray data set.

RESULTS AND DISCUSSION

MMP Distribution. From the entire compound set, a total of 30 954 nonredundant MMPs were generated that involved a total of 8010 compounds and yielded 7256 different transformations. Most of these transformations were represented by a single MMP or small numbers of MMPs. Differences in the number of target annotations between compounds forming an MMP were evaluated. Therefore, for each MMP, the target profiles of its two compounds were compared. The results are reported in Figure S1 of the Supporting Information. Figure 2 reports the distribution of MMPs over increasing differences in target numbers. Compounds forming 18 251 MMPs (~59%) did not differ in the number of targets they were active against. Only 995 of these MMPs were active against the same number of targets, but different targets (Figure S1 of the Supporting Information). Hence, compounds comprising these 18 251 MMPs displayed the same or comparable levels of promiscuity and were thus of low priority for our analysis.

By contrast, compounds in 829 (~2.7%) and 126 MMPs (~0.4%) differed in their activity by 10 or more and 50 or more targets, respectively, thus revealing structurally similar compounds associated with unexpectedly large differences in apparent promiscuity. As a pinnacle of these trends, 33 MMPs were identified in which compounds differed by 90 or more targets. Taken together, these findings were rather surprising. The 126 MMPs in which activity annotations of compounds differed by 50 or more targets (highlighted in Figure 2) were classified as promiscuity cliffs and subjected to further analysis.

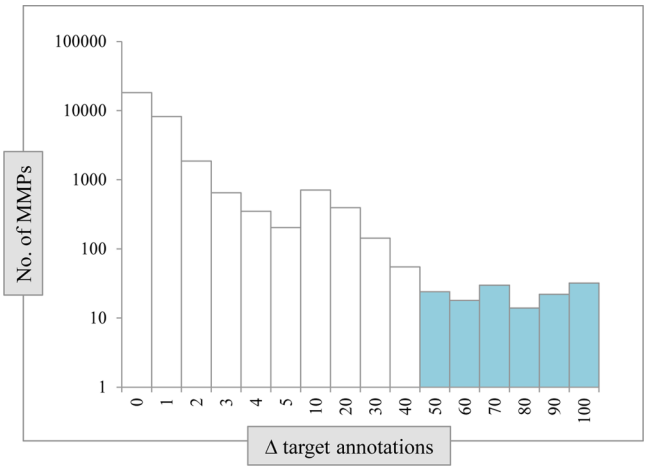


Figure 2. MMPs and target annotations. MMP counts are reported (on a logarithmic scale) for increasing differences in the number of targets MMP-forming compounds were active against. On the horizontal axis, "Δ target annotations" reports binned differences in target numbers. For example, "1", "10", and "100" mean that compounds forming an MMP differed by exactly 1, 6−10, and 91−100 targets, respectively. Sections of the histogram that represent MMPs with a difference of 50 or more targets are highlighted.

Given currently available data, one cannot be certain that binding to a large number of targets might always be specific (in fact, in some instances, this might be unlikely), and which role local concentration effects on arrays might play. Given that compound promiscuity can have several origins and is influenced by multiple factors, as discussed in the Introduction, the analysis of apparent promiscuity on the basis of compound activity profiles takes these factors implicitly into account. On the basis of the original array data analysis reported by Clemons et al., experimental variances were clearly limited to the level expected for microarrays.

We also identified a total of 1146 MMPs that were formed between an inactive compound and an active compound with at least five target annotations, as reported in Figure S2 of the Supporting Information. These MMPs contained compounds active against 5−95 targets. Of these, 58 MMPs qualified as promiscuity cliffs.

Molecular Properties. For 117 compounds involved in the formation of the 126 promiscuity cliffs, four different physicochemical properties were calculated using the Molecular Operating Environment,[21] including molecular weight, octanol/water (o/w) partition coefficient (log *P*), and the numbers of acidic and basic atoms. The distribution of molecular weight is reported in Figure 3a. Compounds that were inactive or active against less than five targets (left region of the plot) covered a broad range, from about 400 to nearly 1000 Da. However, most of the promiscuous compounds (right region) displayed a narrower range. Figure 3b reports the correlation between the changes in molecular weight and promiscuity for individual cliffs. No obvious trends were observed.

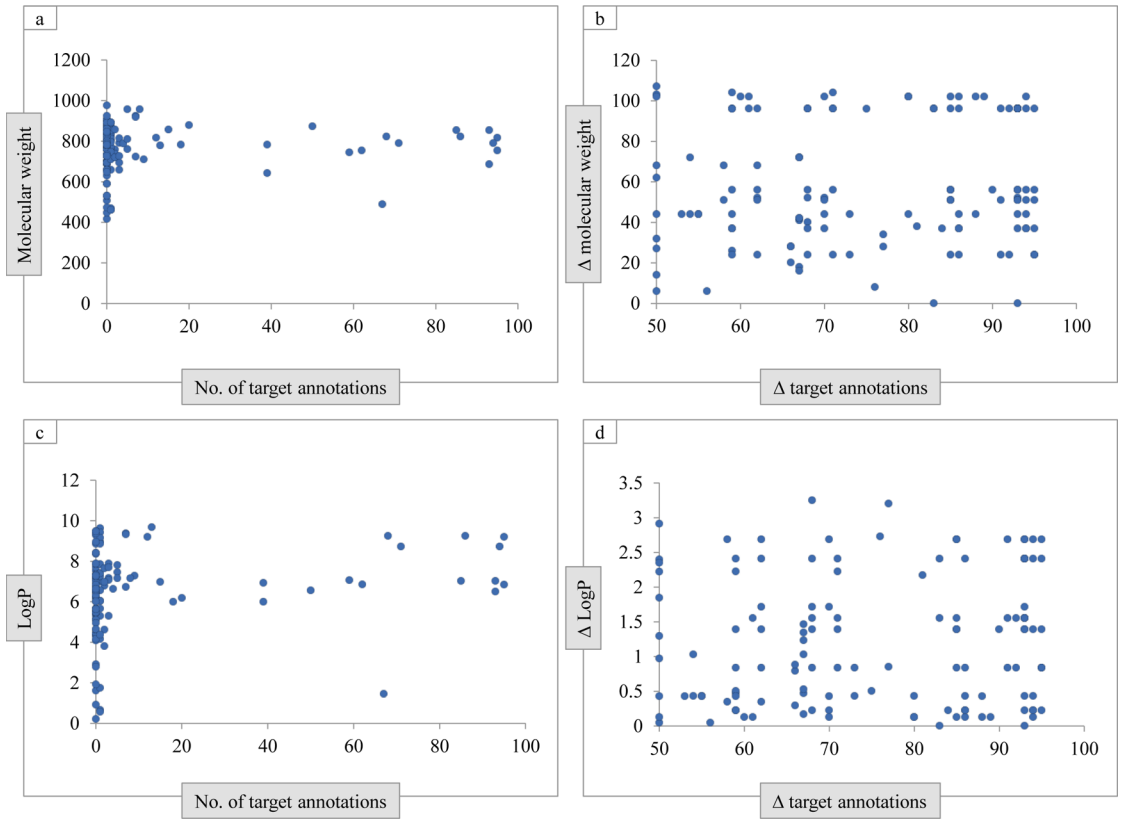
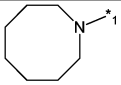
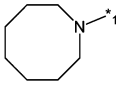
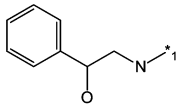
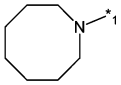
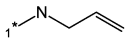
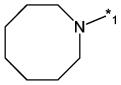
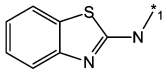
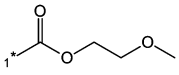
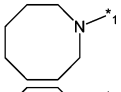
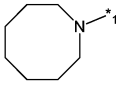
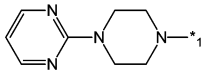
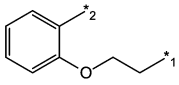
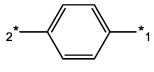
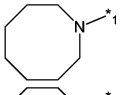
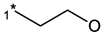
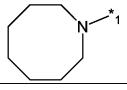


Figure 3. Distribution of molecular properties. For compounds involved in the formation of promiscuity cliffs, the distributions of their molecular weight and o/w partition coefficient (log *P*) are shown in parts a and c, respectively, as a function of the number of target annotations. In these plots, each dot represents a cliff-forming compound. In addition, for promiscuity cliffs, the distributions of the difference in molecular weight and log *P* are shown in parts b and d, respectively, as a function of the difference in the number of target annotations (i.e., differences in the degree of promiscuity). Here, each dot represents a compound pair forming a promiscuity cliff.

dx.doi.org/10.1021/jm301292a | *J. Med. Chem.* 2012, 55, 10220−10228

Table 1. Ranked Transformations[a]

Rank	Transformation		No. of promiscuity cliffs	Total no. of MMPs	Δ target annotations		
					Min	Max	Median
1			11	42	0	95	6.5
2			11	42	0	95	5
3			11	42	0	95	6
4			11	42	0	95	5
5			11	148	0	94	1
6			10	25	0	93	18
7			10	42	0	95	7
8			7	106	0	94	1
9			6	114	0	88	1
10			4	6	0	93	65

[a]The top 10 transformations most frequently found in promiscuity cliffs are listed. The number of promiscuity cliffs and the total number of MMPs containing each transformation are reported. In addition, the minimal (Min) and maximal (Max) differences in the number of target annotations among MMP-forming compounds and median values are given.

Figure 3c shows the distribution of log *P* values. Analogously to the observations made for molecular weight, inactive and nonpromiscuous compounds also covered a broad range of log *P* values. Most of the promiscuous compounds had a much narrower range, i.e., from 6 to 10. On the other hand, in the area of high lipophilicity (upper region of the plot), both nonpromiscuous and promiscuous compounds were found. Although many promiscuous compounds had relatively high log *P* values (as one might expect), there was no detectable correlation between the changes in lipophilicity and the difference in promiscuity, as shown in Figure 3d.

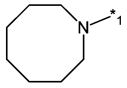
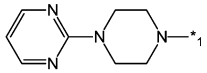
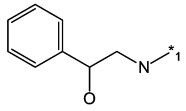
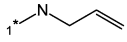
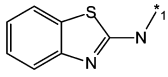
In addition, the protonation states of these compounds were analyzed by counting the numbers of acidic and basic atoms. Nearly all compounds were neutral and only one compound was found to be basic.

Transformations. The 126 MMPs representing promiscuity cliffs encoded 38 unique transformations representing different structural changes (as reported in Table S1 of the Supporting Information). These transformations were ranked according to the number of promiscuity cliffs they occurred in. Table 1 reports the 10 top-ranked transformations. Individual transformations were detected in up to 11 promiscuity cliffs. Notably, eight of the top 10 transformations involved an azocane ring. We calculated the total number of MMPs that represented each of the 38 transformations (including promiscuity cliffs and others). The results for the top 10 transformations are also reported in Table 1. The total number of MMPs ranged from six to 148. We next determined whether these transformations exclusively occurred in MMPs with large target number differences, i.e., whether they represented promiscuity-inducing transformations. Therefore, target number differences in all MMPs representing a given transformation were analyzed. For the top 10 transformations, the minimal and maximal differences in target numbers between MMP-forming compounds and median values are reported in Table 1. For example, for the top-ranked transformation, the median value was 6.5 and MMPs with no target number differences existed. In three other cases, median values of 1 were obtained. Hence, many promiscuity cliff-containing transformations also occurred in MMPs with small target number differences (or no differences). None of the 38 transformations was found to exclusively occur in promiscuity cliffs or other MMPs with large target number differences. Hence, no chemical transformations were detected that consistently induced large-magnitude compound promiscuity.

Substructures. Following the analysis of transformations, we ranked individual substructures involved in these transformations according to the number of promiscuity cliffs in which they occurred (excluding substructures comprising single atoms). Table 2 shows the top five substructures that were

dx.doi.org/10.1021/jm301292a | J. Med. Chem. 2012, 55, 10220−10228

Table 2. Ranked Substructures[a]

Rank	Substructure	No. of promiscuity cliffs	Total no. of MMPs	No. of compounds	No. of annotations		
					Min	Max	Median
1		68	272	86	0	97	6.5
2		14	174	84	0	50	0
3		12	90	1834	0	97	0
4		12	118	625	0	48	0
5		12	90	116	0	28	1.5

[a]The top five substructures most frequently found in promiscuity cliffs are listed. The number of promiscuity cliffs and the total number of MMPs that contain each substructure are reported. In addition, the total number of compounds containing each substructure is reported. Furthermore, the minimal (Min) and maximal (Max) number of target annotations among these compounds and median values are given.

found in more than 10 promiscuity cliffs. Table S2 of the Supporting Information reports all 37 qualifying substructures involved in the formation of cliffs. These substructures were diverse. Corresponding to observations made for transformations, the azocane ring found in 68 promiscuity cliffs was the top-ranked substructure in Table 2. The top five substructures occurred in a total number of 90−272 MMPs and 84−1834 compounds. As also reported in Table 2, the number of targets that compounds containing each substructure were active against greatly varied and also yielded low median values. In three instances, median values of zero were obtained, indicating that at least half of the compounds containing a highly ranked substructure were inactive. As expected on the basis of our transformation analysis, no substructure was found to exclusively occur in promiscuous compounds.

Promiscuous Compounds. All 117 compounds involved in the formation of the 126 promiscuity cliffs were used to generate a molecular network in which nodes represented compounds and edges promiscuity cliffs, as shown in Figure 4a. In this network representation, a number of "promiscuity hubs" became apparent, i.e., compounds with a large number of target annotations involved in the formation of multiple cliffs. It should be noted that these compounds were not only highly promiscuous, but also could be transformed into multiple compounds with limited or no promiscuity through small chemical modifications. The five most prominent promiscuity hubs are highlighted in Figure 4a. These hubs were active against more than 90 targets each and involved in the formation of 9−11 promiscuity cliffs. Their structures are shown in Figure 4b. A characteristic feature of all five compounds was that they contained both the azocane ring and spirooxindole rings (the latter identified by Schreiber and colleagues[18] as a single promiscuity marker in DCs; vide supra). Because of the very large number of targets that promiscuity hubs were active against, it is conceivable that they might at least in part also engage in nonspecific interactions (vide supra).

Promiscuity Cliffs. The co-occurrence of the azocane and spirooxindole substructures in many highly promiscuous compounds suggested the possibility that combinations of substructures (rather than individual ones) might be promiscuity determinants. This possibility could be directly explored because the hubs we identified participated in the formation of multiple promiscuity cliffs. Figure 5 shows examples of prominent promiscuity cliffs containing the azocane and spirooxindole substructures (additional examples are provided in Figure S3 of the Supporting Information). Comparison of cliff-forming compounds clearly revealed that co-occurrence of the azocane and spirooxindole moieties was not a major promiscuity determinant. In the promiscuity cliffs in Figure 5a,b, removal of the azocane ring rendered highly promiscuous compounds (with activity against 95 and 94 targets, respectively) inactive. All compounds in these cliffs also contained the spirooxindole moiety. The cliff forming compounds in Figure 5c both contained the azocane and spirooxindole rings. However, a change in the position of an aliphatic substituent from the para to ortho in the phenyl ring at the lower right was sufficient to transform a highly promiscuous compound into an inactive one. In Figure 5d, the compound containing the para-substituted phenyl ring was also highly promiscuous (i.e., active against 93 targets), whereas the presence of a hydroxyl group at the same position dramatically reduced promiscuity to five targets. However, the ortho-substituted phenyl ring in a different structural context, shown in Figure 5e, was highly promiscuous in contrast to the corresponding analog in Figure 5c. Moreover, compounds containing the para-substituted phenyl ring but different substitutions at the spirooxindole moiety displayed very different degrees of promiscuity (Figure 5e). Taken together, these comparisons revealed a strong structural context dependence of chemical modifications, leading to the formation of promiscuity cliffs. There was no individual substructure or transformation that consistently caused large-magnitude

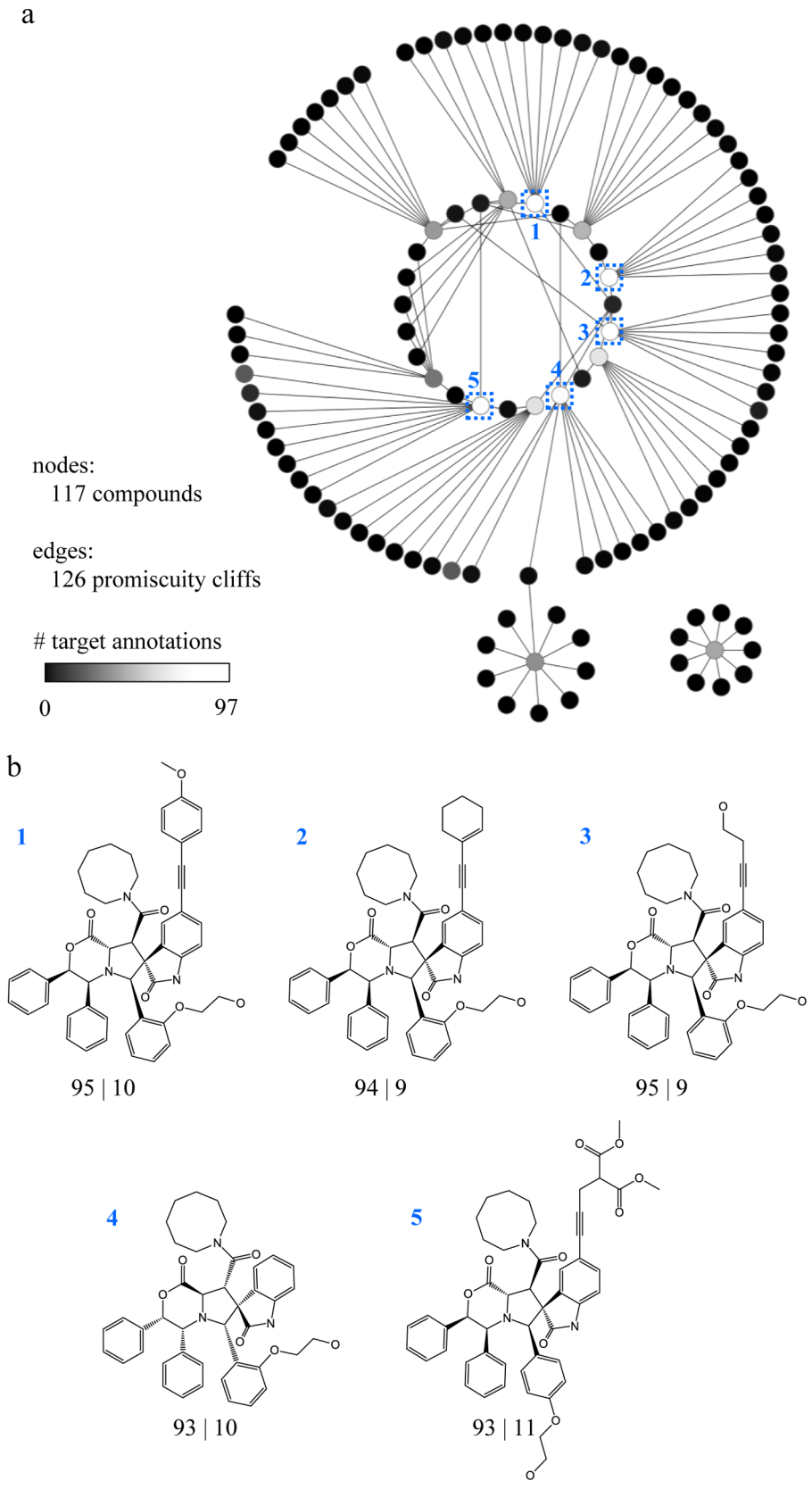


Figure 4. Promiscuity cliff network. (a) MMP-based compound network focusing on promiscuity cliffs. Nodes represent compounds, and edges indicate promiscuity cliffs. Nodes are gray-scaled according to the number of target annotations using a continuous spectrum from black (0 targets; inactive) to white (97 targets; most promiscuous). Five highly promiscuous compounds that were active against more than 90 targets and involved in the formation of 9−11 cliffs are boxed and numbered. Their structures are shown in part b. For each compound, the number of targets it was active against and the number of cliffs it was involved in are reported. For example, "95 | 10" means that the compound was active against 95 targets and involved in the formation of 10 promiscuity cliffs.

dx.doi.org/10.1021/jm301292a | *J. Med. Chem.* 2012, 55, 10220−10228

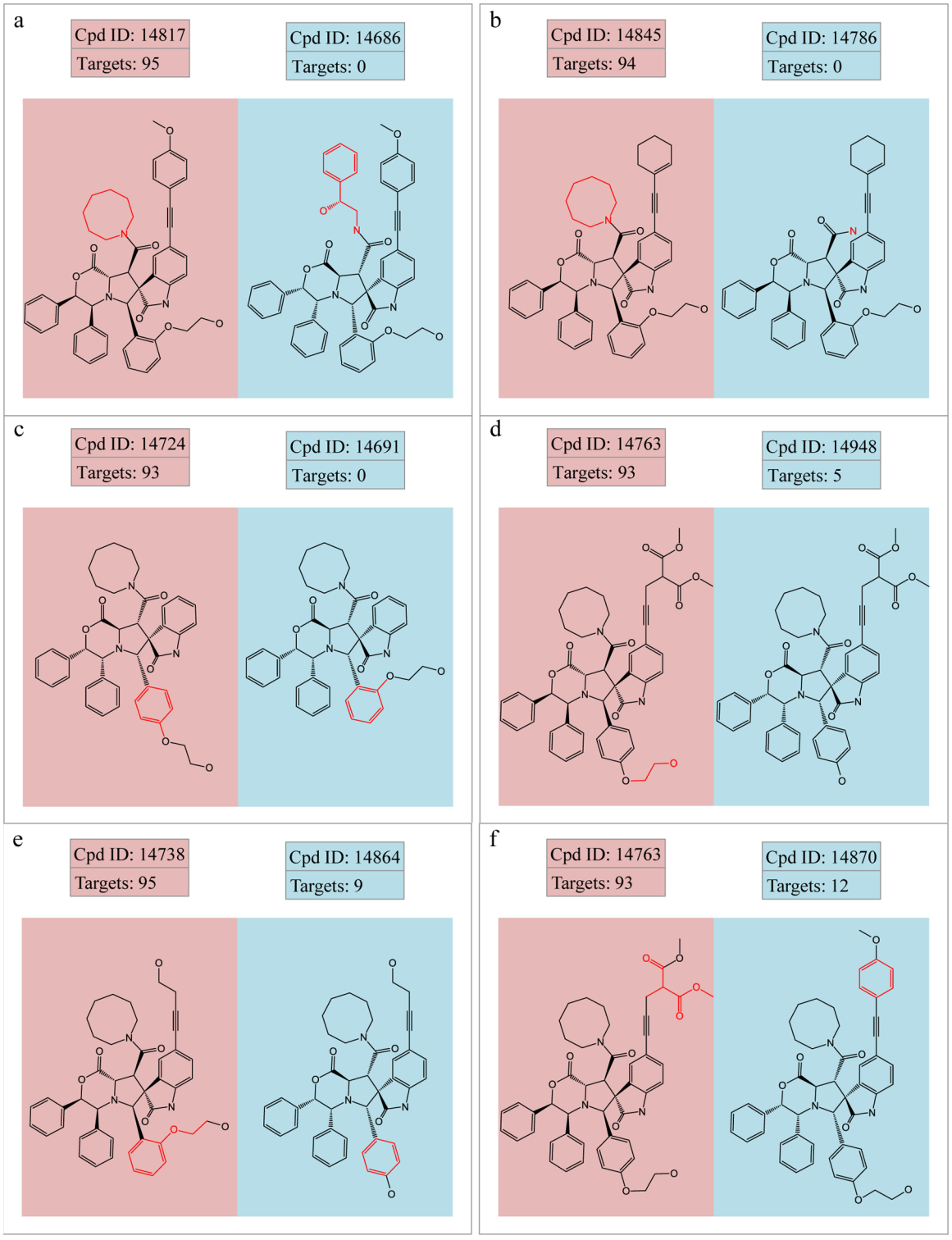


Figure 5. Promiscuity cliffs. Shown are representative MMPs in which activity annotations of compounds differed by more than 80 targets. For each compound, the compound ID and the number of targets it was active against are reported. The promiscuous compound of each cliff is shown on the left and the exchanged fragments are colored red.

promiscuity effects and exclusively occurred in promiscuous compounds.

What Do We Learn about Promiscuity from a Medicinal Chemistry Perspective? On the basis of the data available to us, it is not possible to conclude with certainty to what extent highly promiscuous compounds engage in specific and/or nonspecific interactions with targets. It is of course unlikely that a compound might form specific interactions with 90 or more diverse targets, even if the interactions were clearly detectable under the given experimental conditions. Hence, it is appropriate to consider promiscuity from a phenotypic point of view in the context of our analysis, given the requirement to analyze the data at face value and avoid overinterpretation. However, it should be

dx.doi.org/10.1021/jm301292a | J. Med. Chem. 2012, 55, 10220−10228

noted that only a small fraction of the array compounds were promiscuous in nature and that the formation of promiscuity cliffs was a rare event, thus indicating that the microarray data were suitable for a systematic analysis of promiscuity effects. As we have shown, only a small fraction of MMPs generated from the entire microarray data set combined compounds with notable differences in the number of targets they were active against. Taking this into account, the detection of cliffs in which structurally similar compounds differed in their activity by 50 or more targets is considered a striking finding, regardless of underlying molecular mechanisms.

For medicinal chemistry, a number of findings reported herein are of immediate relevance. It is evident that the MMP-based approach provides a direct and chemically intuitive access to small structural modifications, leading to large-magnitude promiscuity effects. Previously, a number of structural frameworks have been identified that were highly recurrent in promiscuous compounds across different target families.[12] However, it has remained largely unclear from a medicinal chemistry perspective thus far whether certain molecular frameworks carry an intrinsic likelihood of promiscuity and/or might have frequent hitter character. After all, promiscuity is determined for compounds, not their frameworks. Importantly, the findings presented herein do not promote a framework-centric view of promiscuity. Thus, for the evaluation and prioritization of compound series for medicinal chemistry, frameworks should not primarily be considered as an intrinsic source of promiscuity and potential lack of compound specificity. Rather, we demonstrate that small chemical modifications can trigger large-magnitude promiscuity effects. Importantly, these effects depend on the specific structural environment in which these modifications occur. On the basis of our analysis, substitutions that induce promiscuity in any structural environment were not identified. Thus, in medicinal chemistry, it is important to evaluate promiscuity for individual compounds in series that are preferred from an SAR perspective; observed specificity of certain analogs within a series does not guarantee that others are not highly promiscuous. Taken together, these findings further extend our view of molecular origins of promiscuity, putting strong emphasis on the context-dependence of promiscuity-inducing structural modifications. The analysis of compounds in cliff forming MMPs provided a focal point for the identification of such chemical changes that might have otherwise not been detected.

■ CONCLUSIONS

Herein, we have analyzed compound promiscuity on the basis of small molecule microarray data involving ~15 000 compounds and 100 sequence-unrelated targets. These microarray data provide a binary readout of compound activity and are likely influenced, for example, by variance and local concentration effects associated with printing of compounds on solid surfaces by different mechanisms. Nevertheless, as clearly indicated by the results of Clemons et al., who conducted the microarray experiments, the data revealed meaningful binding patterns and systematic trends concerning compound selectivity and, as demonstrated in our study, promiscuity. In the current analysis, we have focused on identifying closely related compounds with large difference in promiscuity, leading to the introduction of promiscuity cliffs. From these compound pairs, chemical modifications at individual sites have become apparent that led to promiscuous binding behavior. Chemical changes were identified that caused large-magnitude promiscuity effects. We have shown that no individual substructure or transformation involved in these effects exclusively occurred in promiscuous compounds. Rather, they were distributed across compounds with different levels of promiscuity or no apparent promiscuity. On the basis of currently available data, promiscuity is not an inherent feature of certain structural frameworks. However, we have shown that chemical modifications could trigger promiscuity within specific structural contexts. Exemplary promiscuity cliffs have revealed that similar substitutions in different structural environments can lead to promiscuity effects of different magnitude, or even opposite effects (i.e., increase vs reduction in target numbers). On the basis of our analysis, small structural modifications of nonpromiscuous compounds can lead to substantial promiscuity. However, these effects are structural-context-dependent.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Figure S1 reports the comparison of target annotations for compound pairs forming MMPs, Figure S2 shows the distribution of the number of targets for compounds that were active against at least five targets and formed MMPs with inactive ones, Figure S3 shows further examples of promiscuity cliffs, and Tables S1 and S2 report transformations encoded by promiscuity cliffs and substructures involved in these transformations, respectively. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

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

Author Contributions

[†]The contributions of these authors should be considered equal.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS USED

CC, commercial compound; DC, compound from diversity-oriented synthesis; MMP, matched molecular pair; NP, natural product; SAR, structure−activity relationship

■ REFERENCES

(1) Feng, B. Y.; Shelat, A.; Doman, T. N.; Guy, R. K.; Shoichet, B. K. High-Throughput Assays for Promiscuous Inhibitors. *Nat. Chem. Biol.* **2005**, *1*, 146−148.

(2) Paolini, G. V.; Shapland, R. H. B.; van Hoorn, W. P.; Mason, J. S.; Hopkins, A. L. Global Mapping of Pharmacological Space. *Nat. Biotechnol.* **2006**, *24*, 805−815.

(3) Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating Protein Pharmacology by Ligand Chemistry. *Nat. Biotechnol.* **2007**, *25*, 197−206.

(4) Hopkins, A. L. Network Pharmacology: The Next Paradigm in Drug Discovery. *Nat. Chem. Biol.* **2008**, *4*, 682−690.

(5) Keiser, M. J.; Setola, V.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijer, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. Predicting New Molecular Targets for Known Drugs. *Nature* **2009**, *462*, 175−181.

(6) Ashburn, T. T.; Thor, K. B. Drug Repositioning: Identifying and Developing New Uses for Existing Drugs. *Nat. Rev. Drug Discovery* **2004**, *3*, 673−683.

(7) Chong, C. R.; Sullivan, D. J. New Uses for Old Drugs. *Nature* **2007**, *448*, 645−646.
(8) Mestres, J.; Gregori-Puigjané, E. Conciliating Binding Efficiency and Polypharmacology. *Trends Pharmacol. Sci.* **2009**, *30*, 470−474.
(9) Merino, A.; Bronowska, A. K.; Jackson, D. B.; Cahill, D. J. Drug Profiling: Knowing Where It Hits. *Drug Discovery Today* **2010**, *15*, 749−756.
(10) Koutsoukas, A.; Simms, B.; Kirchmair, J.; Bond, P. J.; Whitmore, A. V.; Zimmer, S.; Young, M. P.; Jenkins, J. L.; Glick, M.; Glen, R. C.; Bender, A. From in Silico Target Prediction to Multi-Target Drug Design: Current Databases, Methods and Applications. *J. Proteomics* **2011**, *74*, 2554−2574.
(11) Xie, L.; Xie, L.; Kinnings, S. L.; Bourne, P. E. Novel Computational Approaches to Polypharmacology as a Means to Define Responses to Individual Drugs. *Annu. Rev. Pharmacol. Toxicol.* **2012**, *52*, 361−379.
(12) Hu, Y.; Bajorath, J. Polypharmacology Directed Compound Data Mining: Identification of Promiscuous Chemotypes with Different Activity Profiles and Comparison to Approved Drugs. *J. Chem. Inf. Model.* **2010**, *50*, 2112−2118.
(13) Campillos, M.; Kuhn, M.; Gavin, A. C.; Jensen, L. J.; Bork, P. Drug Target Identification Using Side-Effect Similarity. *Science* **2008**, *321*, 263−266.
(14) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107.
(15) Wang, Y.; Xiao, J.; Suzek, T. O.; Zhang, J.; Wang, J.; Zhou, Z.; Han, L.; Karapetyan, K.; Dracheva, S.; Shoemaker, B. A.; Bolton, E.; Gindulyte, A.; Bryant, S. H. PubChem's Bioassay Database. *Nucleic Acids Res.* **2012**, *40*, D400−D412.
(16) Metz, J. T.; Johnson, E. F.; Soni, N. B.; Merta, P. J.; Kifle, L.; Hajduk, P. J. Navigating the Kinome. *Nat. Chem. Biol.* **2011**, *7*, 200−202.
(17) Iyer, P.; Dimova, D.; Vogt, M.; Bajorath, J. Navigating High-Dimensional Activity Landscapes: Design and Application of the Ligand−Target Differentiation Map. *J. Chem. Inf. Model.* **2012**, *52*, 1962−1969.
(18) Clemons, P. A.; Bodycombe, N. E.; Carrinski, H. A.; Wilson, J. A.; Shamji, A. F.; Wagner, B. K.; Koehler, A. N.; Schreiber, S. L. Small Molecules of Different Origins Have Distinct Distributions of Structural Complexity that Correlate with Protein-Binding Profiles. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 18787−18792.
(19) *Interpro Sequence Analysis & Classification*, www.ebi.ac.uk/interpro/ (Accessed August 14, 2012).
(20) Kenny, P. W.; Sadowski, J. Structure Modification in Chemical Databases. In *Chemoinformatics in Drug Discovery*; Oprea, T. I., Ed.; Wiley-VCH: Weinheim, Germany, 2004; pp 271−285.
(21) *Molecular Operating Environment (MOE)*, 2011.10; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2011.
(22) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.
(23) Hussain, J.; Rea, C. Computationally Efficient Algorithm To Identify Matched Molecular Pairs (MMPs) in Large Data Sets. *J. Chem. Inf. Model.* **2010**, *50*, 339−348.
(24) Wassermann, A. M.; Bajorath, J. Chemical Substitutions That Introduce Activity Cliffs across Different Compound Classes and Biological Targets. *J. Chem. Inf. Model.* **2010**, *50*, 1248−1256.
(25) Wassermann, A. M.; Bajorath, J. Large-Scale Exploration of Bioisosteric Replacements on the Basis of Matched Molecular Pairs. *Future Med. Chem.* **2011**, *3*, 425−436.
(26) 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.
(27) Tiwaria, A.; Sekhar, A. K. T. Workflow Based Framework for Life Science Informatics. *Comput. Biol. Chem.* **2007**, *31*, 305−319.
(28) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498−2504.