Molecular Scaffolds with High Propensity to Form Multi-Target Activity Cliffs

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

Received February 8, 2010

In target-dependent activity landscapes of compound series, cliffs are formed by pairs of molecules that are structurally analogous but display significant differences in potency. The detection and analysis of such activity cliffs is a major task in structure−activity relationship analysis and compound optimization. In analogy to activity cliffs, selectivity cliffs can be defined that are formed by structural analogs having significantly different potencies against two targets. The formation of activity cliffs by analogs is generally a consequence of different R-group patterns; e.g., a specific substitution of a given scaffold might increase and another substitution decrease potency. Therefore, activity (or selectivity) cliffs are typically analyzed for a given scaffold representing an analog series, and it has thus far not been explored whether certain scaffolds might display a general tendency to yield compounds forming activity cliffs against different targets. We have exhaustively analyzed scaffolds and associated compound activity data in the ChemblDB and BindingDB databases in order to compare the availability of target-selective scaffolds in these databases and determine whether multi-target activity and multi-target selectivity cliff scaffolds exist. Perhaps unexpectedly, we have identified 143 scaffolds that are represented by multiple compounds and form activity or selectivity cliffs against different targets. These scaffolds have varying chemical complexities and are in part promiscuous binders (i.e., compounds containing these scaffolds bind to distantly related or unrelated targets). However, analogs derived from these scaffolds form steep activity cliffs against different targets. A catalog of scaffolds with high propensity to form activity or selectivity cliffs against multiple targets is provided to help identify potentially promiscuous candidate scaffolds during compound optimization efforts.

INTRODUCTION

Molecular scaffolds (core structures) are of high interest in pharmaceutical research as building blocks or markers of drug-like compounds.[1−5] Scaffolds are often defined in different ways, which makes it difficult to assess and compare studies that explore scaffold distributions or scaffold hopping.[3] For example, scaffolds might be systematically derived by breaking predefined bonds in compounds following a hierarchy or on the basis of retro-synthetic criteria,[6] i.e., by separating groups in molecules according to chemical reactions carried out to synthesize them. Organic ring systems have thus far been a major focal point of scaffold analysis and design.[7,8] However, scaffolds have been analyzed from rather different points of view. For example, scaffold distributions have been determined for screening libraries,[9] large databases of synthetic molecules,[10] or compounds at different pharmaceutical development stages.[11] Frequency analysis has typically been applied to identify molecular scaffolds that are recurrent in synthetic molecules[12] or in compounds active against different targets.[13] In addition, attempts have also been made to systematically organize scaffold populations derived from active compounds on the basis of structural and activity criteria and thereby establish scaffold systems and hierarchies.[14,15] Scaffold analysis has also received much attention in the context of fragment-based drug discovery[16−18] where small weakly active compounds are combined in order to generate potent leads.

In addition to the scaffold analysis schemes described above, the high interest in "privileged substructures"[19] thought to preferentially bind to a given target class has triggered intense scaffold analysis efforts.[20−22] For the evaluation of privileged substructures, frequency analysis has also been carried out to compare the occurrence of proposed privileged substructures in different compound activity classes.[22] Although evidence for a preferential enrichment of certain scaffolds in specific activity classes has been accumulating,[21] the existence of truly privileged substructures has remained controversial.[22]

In order to thoroughly explore the presence of target class-selective molecular scaffolds beyond frequency calculations, we have previously carried out a large-scale analysis of public domain compounds with multiple activity annotations.[23] In this study, target communities were defined via a compound-based target network where individual targets were connected if they shared at least five active compounds. For different target communities, active compounds were collected, and the community selectivity of scaffolds derived from them was explored. This analysis has led to the identification of a total of 206 scaffolds that were selective for one of 18 target communities.[23] On the basis of these findings, we subsequently also searched for target-selective, rather than target class-selective, scaffolds. For this purpose, we modified the network analysis approach and introduced a scaffold-based target network where targets were connected if they shared at least five "active" scaffolds (rather than compounds).[24] Ultimately, we identified 42 scaffolds, each

* To whom correspondence should be addressed. Tel.: +49-228-2699-306. Fax: +49-228-2699-341. E-mail: bajorath@bit.uni-bonn.de.

10.1021/ci100059q
Published on Web 04/02/2010

of which was represented by multiple compounds that were highly selective (at least 100-fold) for a given target over one or more others.[24] However, we also found that currently available selectivity data were in general only sparse. For example, many scaffolds that formally met the criteria for target selectivity were only represented by individual compounds. Our scaffold analyses were based on public domain compound data available in BindingDB (BDB),[25] a major source of activity information of small molecules, in addition to PubChem.[26] From BDB data, a total of 520 pairs of human targets were identified that shared at least five ligands.[23] By contrast, in PubChem confirmatory bioassays, only three target pairs could be identified that met our selection criterion.

Herein, we address as of yet unexplored questions in scaffold analysis, namely, whether scaffolds exist that have a high propensity to introduce "activity cliffs"[27] in structure−activity landscapes. Activity cliffs are formed by structurally similar compounds having large differences in potency and are the major source of SAR discontinuity,[28] which provides opportunities for compound optimization but often hinders QSAR modeling and activity predictions.[27,28]

It is indeed difficult to speculate about the question of whether scaffolds might have significantly different abilities to form activity cliffs. This is the case because strong activity cliffs are formed by close structural analogs with large potency differences, i.e., compounds that usually contain the same scaffold. Therefore, the formation of target-specific activity cliffs is primarily attributed to different substitution patterns of the same or very similar scaffolds, and it has thus far not been investigated whether scaffolds might exist that have an intrinsic ability to form activity cliffs across different targets. This analysis can also be extended to study the relationship between scaffolds and "selectivity cliffs" that are formed by structurally similar compounds having different selectivities against two targets.[29]

To investigate these questions, we have systematically explored relationships between active compounds, corresponding scaffolds, activity cliffs, and selectivity cliffs. Recently, the ChemblDB (CDB) database[30] has become available as another major public domain source of compound activity data, in addition to PubChem and BDB. In contrast to PubChem bioassays, we found sufficient CDB compound data for meaningful scaffold-based target network analysis and exploration of target-selective scaffolds. Therefore, we have carried out this analysis also for CDB to compare the results with BDB. The scaffold information extracted from these databases has then been utilized to search for scaffolds forming activity (and selectivity) cliffs for multiple targets.

METHODS

Scaffold Definition and Extraction. Scaffolds were derived from active compounds according to Bemis and Murcko.[1] Following this approach, scaffolds are isolated from synthetic compounds by removing R groups from ring systems but retaining linkers between rings.[1] Thus, these hierarchical scaffolds comprise individual, condensated, or linked ring systems. The Bemis and Murcko approach has been a major origin of systematic analyses of scaffold distributions in drugs.[1] It should also be noted that the information required for ligand-centric scaffold distribution analysis is exclusively obtained from two-dimensional molecular graphs. Hence, aspects of three-dimensional structure or protein−ligand interactions do not play a role in the context of this analysis.

BDB compounds with reported activity against human targets were collected. For compounds with multiple potency measurements against the same target, the geometric mean was calculated as the final potency value. From CDB, compounds active against human targets were selected that had the highest target confidence level (CDB target confidence score 9) for direct interactions (target relationship type "D"). From all selected BDB and CDB compounds, scaffolds were isolated and represented as SMILES strings[31] for further analysis.

Scaffold-Based Target Network. For BDB and CDB scaffolds, separate scaffold-based target networks were generated for comparison. Target pairs were formed if two targets shared multiple active compounds representing at least five unique scaffolds. In such networks, targets are represented as nodes and connected by an edge if the target pair criterion is met. The width of edges is scaled according to the number of shared scaffolds. Network representations were drawn with Cytoscape.[32]

Community- and Target-Selective Scaffolds. Scaffolds that exclusively occurred in compounds active against only one of *n* target communities in the network were determined and termed "community-selective" scaffolds. For each compound active against a target pair, its selectivity ratio (SR) was calculated as follows:

$$\mathrm{SR} = \mathrm{pot}_A(i) - \mathrm{pot}_B(i)$$

Here, $\mathrm{pot}_A(i)$ and $\mathrm{pot}_B(i)$ represent the negative logarithm of potency values of compound i for targets A and B, respectively. Compounds and corresponding scaffolds were classified according to different selectivity levels, i.e., corresponding to at least a 10-fold, 50-fold, or 100-fold potency difference. If all compounds representing a unique scaffold were found to be selective for one particular target over one or more others (e.g., selective for A over B, A over C, etc.), the scaffold was classified as "target-selective".

Classification of Scaffolds. Scaffolds were further classified according to three different criteria reflecting the formation of activity or selectivity cliffs by compounds derived from these scaffolds.

Compound Potency Value Ranges. For each scaffold, the potency range of its compounds against each target was determined, and the maximum potency was recorded. If a scaffold was represented by a single compound, the potency interval was set to 0. If multiple compounds existed that displayed the same potency, the scaffold was assigned to interval [0, 1], i.e., 0 to 1 order of magnitude difference in potency. Scaffolds were assigned to a total of six potency intervals, i.e., 0, [0, 1], [1, 2], [2, 3], [3, 4], and [4, Max]. Max designates the highest potency value range. For example, a scaffold was assigned to interval [2, 3] if the potency range of its compounds spanned 2 to 3 orders of magnitude. For each potency interval, the total number of scaffolds, average number of unique compounds per scaffold, average number of targets, and average maximum potency were calculated.

Compound Selectivity Ratio Ranges. In analogy to compound potency-based scaffold classification, selectivity ratio ranges for target pairs were determined for compounds representing each scaffold and corresponding selectivity

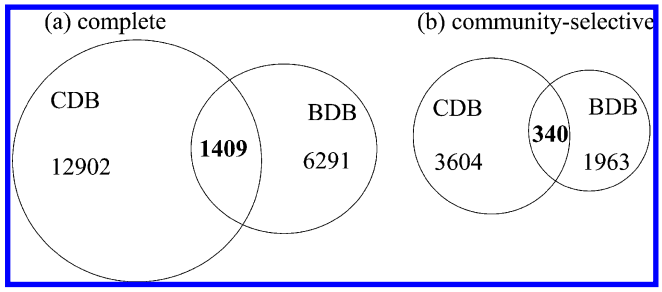

Figure 1. Scaffold overlap between CDB and BDB. Scaffold sets extracted from CDB and BDB are compared in Venn diagrams: (a) all scaffolds, (b) community-selective scaffolds. The number of shared scaffolds is shown in bold.

interval assignments were made, i.e. 0, [0, 1], [1, 2], [2, 3], [3, 4], and [4, Max]. Here, Max designates the highest selectivity value range. For example, a scaffold was assigned to selectivity interval [3, 4] if the potency ratios of its compounds against their target pairs spanned 3 to 4 orders of magnitude. For each selectivity interval, the number of scaffolds, average number of unique compounds per scaffold, average number of target pairs, and average maximum potency ratio were calculated.

Scaffold Discontinuity Scores. A local SAR discontinuity score was originally developed to quantify the SAR contributions of individual compounds in data sets and identify compounds forming activity cliffs.[33] We have adapted this scoring scheme to assess the propensity of scaffolds to yield compounds forming activity or selectivity cliffs. For each scaffold, all compounds were collected for all targets (activity cliff assessment) or target pairs (selectivity cliff assessment),

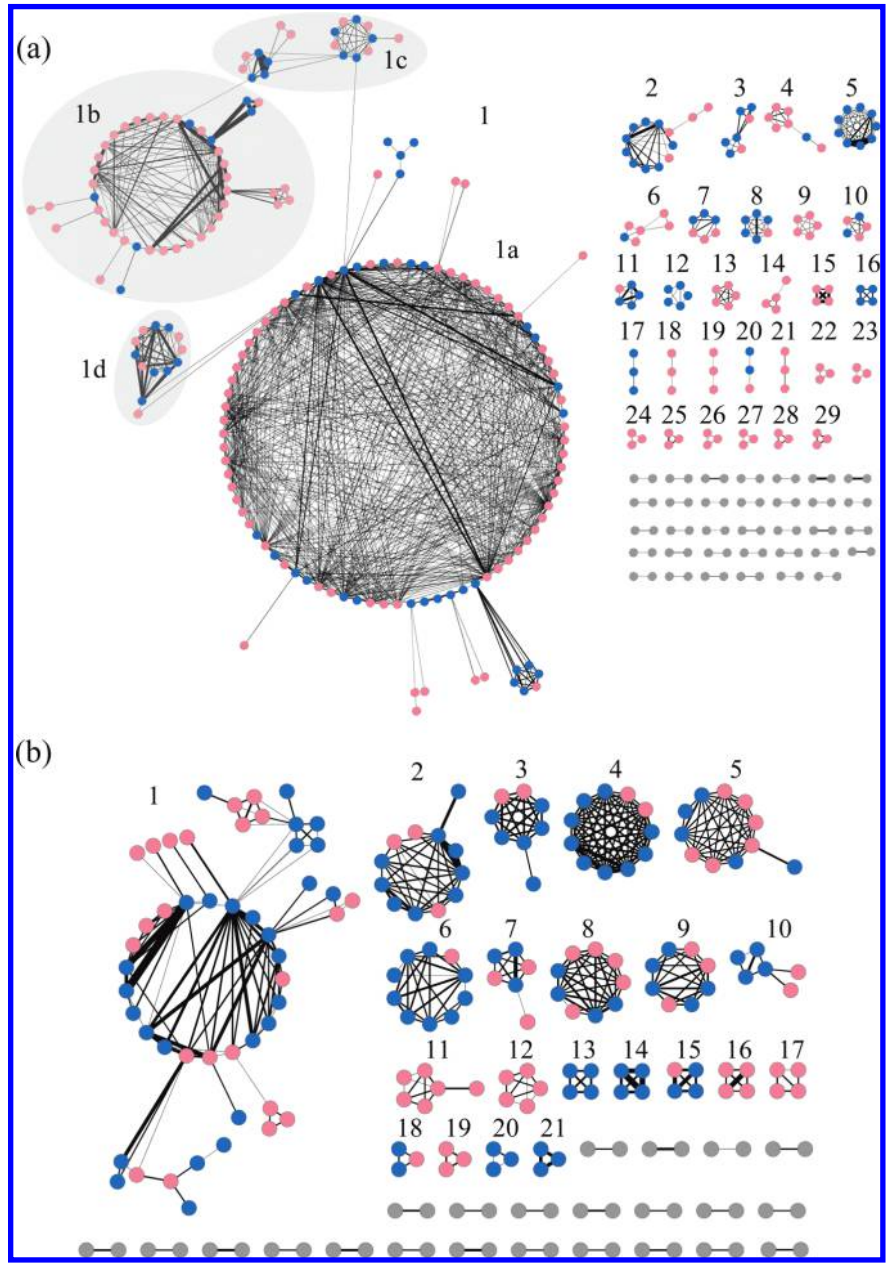

Figure 2. Scaffold-based target networks. Nodes represent targets that are connected by an edge if they share at least five scaffolds. Edge width is scaled according to the number of shared scaffolds. Target communities that contain at least three targets are considered in our analysis and consecutively numbered. Target classes are described in Table 1. Nodes representing targets that do not belong to these communities are colored gray. Network representations are shown for (a) CDB and (b) BDB. Targets common to CDB and BDB are colored blue.

Table 1. Composition of Target Communities[a]

community	target family	number of targets	number of target pairs	number of compounds	number of scaffolds
	(a) CDB				
1a	tyrosine kinases, serine/threonine protein kinases	99	696	1283	605
1b	GPCRs	43	151	2685	1090
1c	GPCRs, cytochrome P450 enzymes	18	45	779	348
1d	matrix metalloproteinases	13	47	577	256
2	serine proteases	13	27	345	202
3	phosphodiesterases	7	12	145	53
4	prostanoid receptors	7	11	120	51
5	carbonic anhydrases	9	36	462	173
6	phosphatases	6	8	35	20
7	dipeptidyl peptidases	6	10	118	67
8	steroid receptors	6	15	399	89
9	GABAA receptors	5	10	48	9
10	sphingosine 1-phosphate (S1P) receptors	5	10	152	39
11	cathepsins	5	9	270	154
12	histone deacetylases	5	7	21	11
13	somatostatin receptors	5	10	75	39
14	cytochrome P450 enzymes	4	4	95	36
15	melanocortin receptors	4	6	317	179
16	caspases	4	6	126	61
17	β-secretases and cathepsin D	3	1	12	6
18	matrix metalloproteinases	3	2	10	10
19	fatty acid binding proteins	3	2	10	6
20	dehydragenases	3	2	8	7
21	vasopressin/oxytocin receptors	3	2	91	32
22	excitatory amino acid transporters	3	3	32	10
23	retinoic acid receptors	3	3	8	6
24	steroid reductases/isomerases	3	3	37	12
25	peroxisome proliferator-activated receptors	3	3	154	65
26	guanine nucleotide-binding protein G	3	3	28	9
27	neuropeptide Y receptors	3	3	10	10
28	adrenergic receptors	3	3	101	46
29	nitric-oxide synthase	3	3	88	42
	(b) BDB				
1	tyrosine kinases and cytochrome P450 enzymes	50	100	2128	782
2	serine proteinases	12	34	545	229
3	protein kinase C	8	22	72	34
4	carbonic anhydrases	11	55	327	87
5	phosphodiesterases	11	39	117	47
6	matrix metalloproteinases	10	24	187	56
7	protein kinase B and serine protein kinases	6	11	109	78
8	caspases	9	31	114	49
9	histone deacetylases	8	22	121	68
10	purinergic receptors	6	7	107	54
11	phosphoinositide 3-kinases (PI3Ks)	6	10	46	26
12	GABAA receptors	5	9	8	7
13	opioid receptors	4	6	84	27
14	cathepsins	4	6	307	152
15	dipeptidyl peptidases	4	6	287	105
16	esterases	4	6	238	110
17	polo-like kinases	4	5	35	21
18	sphingosine 1-phosphate (S1P) receptors	3	3	20	9
19	peroxisome proliferator-activated receptors	3	3	61	16
20	steroid receptors	3	3	35	9
21	β-secretases and cathepsin D	3	3	127	66

[a] Target communities extracted from scaffold-based target networks are reported for (a) CDB and (b) BDB. For each community, the target family annotation, the number of targets and target pairs, and the number of active compounds and corresponding scaffolds are reported.

and the potency-based scaffold discontinuity score (PScS) or selectivity-based score (SScS) was calculated as follows:

$$\mathrm{PScS}(s) = \frac{\sum (|p_i - p_j| \times \mathrm{sim}(i,j))}{|ij|}$$

$$\mathrm{SScS}(s) = \frac{\sum (|\mathrm{SR}_i - \mathrm{SR}_j| \times \mathrm{sim}(i,j))}{|ij|}$$

Here, $|p_i - p_j|$ and $|\mathrm{SR}_i - \mathrm{SR}_j|$ indicate the absolute potency value and selectivity ratio difference of compounds i and j represented by scaffold s, respectively, $\mathrm{sim}(i,j)$ is the structural similarity of compounds i and j, assessed by MACCS[34] Tanimoto similarity,[35] and $|ij|$ is the number of all compound pairs. Scores were normalized with respect to scaffold scores. Raw scores were first transformed into conventional z scores and then mapped to a cumulative probability function assuming a normal value distribution in order to obtain final scores between 0 and 1.[33] Scaffolds were initially ranked on the basis of these scores that reflect their general propensity to form activity and/or selectivity cliffs.

For each scaffold representing at least three compounds active against more than one target, the score calculations over all targets described above were then repeated for each individual target using the compounds active against the target. These calculations identify activity/selectivity cliffs on a per target basis.

Scaffold analysis and classification was carried out with in-house generated Perl and Pipeline Pilot[36] programs.

RESULTS AND DISCUSSION

Comparison of BDB and CDB Scaffolds. Given our selection criteria for compounds active against human targets, 17 745 compounds with activity annotations against 433 human targets were taken from BDB. These compounds produced 6291 unique scaffolds. From CDB, 32 848 compounds active against 671 human targets were selected yielding 12 902 unique scaffolds. There was limited compound and scaffold overlap between BDB and CDB; only 3589 compounds and 1409 scaffolds were shared by both databases (Figure 1). Hence, the scaffold information in both databases was complementary and a total of 47 004 unique compounds and 17 784 unique scaffolds were available for further analysis.

Scaffold-Based Target Network. The CDB and BDB scaffold sets were used to build scaffold-based target networks in order to establish target communities (classes) for the analysis of community- and target-selective scaffolds. In these network representations, targets (nodes) are connected if they share active compounds yielding at least five unique scaffolds. The scaffold-based target networks are shown in Figure 2, and the resulting communities are designated in Table 1. The CDB network in Figure 2a displays a total of 29 separate communities each consisting of at least three targets. The major network component 1 can be further subdivided into four distinct target communities (1a−1d), hence yielding a total of 32 target communities. However, the network is clearly dominated by community 1a, representing tyrosine kinases, and, to a lesser extent, by community 1b, representing G protein coupled receptors (GPCRs). Thus, kinase inhibitors and GPCR antagonists account for much of the information contained in CDB. Target communities in the corresponding BDB network in Figure 2b are more evenly distributed. A total of 21 communities with at least three targets are formed. Here, the largest community 1 is formed by kinases and cytochrome P450 isoforms (that share many active compounds in BDB). However, this community is much smaller than community 1a in the CDB network that contains much more kinase (but no cytochrome P450) information. There is significant overlap between a number of multi-target communities in CDB and BDB (as indicated by blue nodes in Figure 2), but both networks also contain several distinct small communities. A particularly noteworthy case of complementarity between these databases is provided by the GPCR community 1b in the CDB network, its second largest community. GPCR information is clearly under-represented in BDB (see communities 10 and 13) where GPCR ligands correspond to a total of fewer than 100 unique scaffolds, whereas 1090 GPCR ligand scaffolds are found in CDB (Table 1). Moreover, there are also relative differences between target and scaffold information in CDB and BDB, the most striking case again being the kinase communities. In CDB, community 1a represents 696 target pairs that are connected by 605 scaffolds. In BDB, the combined kinase/cytochrome P450 community 1 only represents 100 target pairs that are, however, connected by 782 scaffolds. Hence, for kinases, CDB contains more target and BDB more scaffold/chemical information. Similar observations are made for other corresponding target communities.

Table 2. Comparison of Target and Scaffold Numbers[a]

		CDB	BDB
network	targets	371	220
	target pairs	1188	428
	scaffolds	4167	2467
communities	number	32	21
	targets	303	174
	target pairs	1154	405
	community-selective scaffolds	3604	1963
target-selective scaffolds	10-fold	695 (112)	472 (100)
	50-fold	343 (43)	250 (55)
	100-fold	244 (24)	191 (42)
selectivity patterns	10-fold	184 (104)	78 (50)
	50-fold	114 (64)	49 (31)
	100-fold	88 (51)	42 (23)

[a] The numbers of targets, community- and target-selective scaffolds, and selectivity patterns are reported for CDB and BDB. Target-selective scaffolds and selectivity patterns are provided at different selectivity levels. The numbers of target-selective scaffolds and selectivity patterns that are represented by multiple compounds are given in parentheses. Selectivity patterns are target relationships evolving around specific targets that are formed by multiple target-selective scaffolds.

Community- and Target-Selective Scaffolds. We determined the number of community-selective scaffolds for each database and the number of target-selective scaffolds at different selectivity levels. The results are reported in Table 2. Of 4167 CDB and 2467 BDB scaffolds that were extracted from the scaffold-based target networks, 3658 and 1991, respectively, were found to be community-selective (i.e., the compounds represented by each scaffold were only active against targets in a single community). However, only 340 community-selective scaffolds were common to CDB and BDB, and hence the overlap was limited (Figure 1). We then compared target-selective scaffolds in CDB and BDB. Previously, we reported that a total of 191 target-selective scaffolds were available in BDB at a 100-fold selectivity level but that only 42 of those were represented by multiple compounds.[24] Similar results were also obtained for CDB. The number of target-selective scaffolds declined from 695 at the 10-fold selectivity level to 244 at the 100-fold level. However, at the 50- and 100-fold selectivity level, only 43 and 24 of these CDB scaffolds, respectively, were represented by multiple compounds, i.e., fewer than for BDB (Table 2). Processing the entire CDB only added 23 unique multiple-compound scaffolds at the 100-fold selectivity level to the 42 previously identified BDB scaffolds, hence supporting the conclusion that public domain selectivity data is currently sparse.

Regardless of compound numbers, target-selective scaffolds are in general interesting from another perspective because they often form "selectivity patterns" around individual targets, i.e., inter-target relationships constituted by

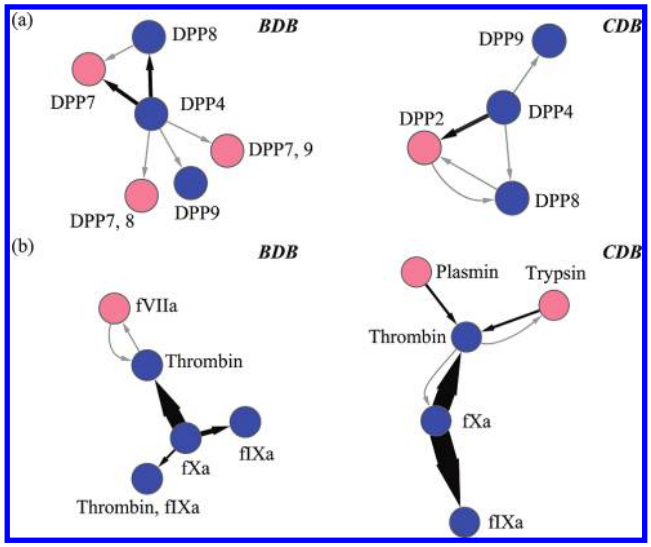

Figure 3. Comparison of selectivity patterns. Shown are four representative selectivity patterns at the 50-fold selectivity level for two target families, (a) dipeptidyl peptidases and (b) serine proteases. For each family, the pattern derived from BDB is shown on the left and the corresponding CDB pattern on the right. Selectivity patterns are displayed in a directed network representation where nodes represent targets that are connected by an arrow if they share at least one target-selective scaffold. The arrow points from the selective target to the non-selective target, thus representing a "selective over" relationship. The width of the arrows is scaled according to the number of shared target-selective scaffolds. Arrows representing selectivity relationships formed by a single scaffold are colored gray. Nodes are annotated with target names. "DPP" stands for dipeptidyl peptidase and "f" for factor. Nodes of targets shared between BDB and CDB are colored blue.

multiple target-selective scaffolds. As examples, Figure 3 shows corresponding BDB and CDB selectivity patterns that evolve around dipeptidyl peptidase 4 and factor Xa, respectively. In such selectivity patterns, target-selective scaffolds establish different selectivity relationships between related targets. Table 2 shows that many selectivity patterns can be extracted from CDB and BDB. As can be seen in Figure 3, selectivity patterns for given targets often differ in CDB and BDB and can be combined to further increase their information content, which again points at a notable degree of complementarity between these databases. The analysis of scaffold-based selectivity patterns is of practical relevance because these patterns that can be exploited, for example, in the design of compounds that are target-selective, have inverse selectivity, or display a desired selectivity profile for a group of related targets.

Potency and Selectivity Ranges. We next classified the complete CDB and BDB scaffold sets according to the potency and selectivity ranges of the compounds they represent. The potency- and selectivity-based classifications are reported in Table 3a and b, respectively. Increasing potency ranges of compounds representing a given scaffold provide an indication of activity cliff potential. Table 3a reveals some clear trends for both CDB and BDB scaffolds. With increasing potency ranges, the number of scaffolds decreases, as one would expect, but the number of compounds per scaffold increases and also the number of targets the compounds are active against. For the largest potency ranges of more than 4 orders of magnitude, 278 CDB scaffolds are found that are, on average, represented by ∼17 compounds active against ∼7 targets and 165 BDB scaffolds each represented by ∼18 compounds active against ∼5 targets. Equivalent trends are observed for the selectivity-based classification in Table 3b. Here, the number of scaffolds also decreases for increasing selectivity ranges, but the number of compounds they represent and the number of target pairs these compounds are active against increase. For the largest target selectivity ranges, 52 CDB and 42 BDB scaffolds exist that are, on average, represented by ∼18 and ∼9 compounds active against ∼209 and ∼31 target pairs, respectively. There are generally fewer scaffolds for selectivity- than potency-based classification because selectivity results from activity against a minimum of two targets, which only applies to a subset of compounds and scaffolds. Many of the prioritized scaffolds have already been explored rather extensively, as suggested by, in part, large numbers of compounds corresponding to individual scaffolds and multiple targets they have been tested against. This suggests that more extensive chemical exploration of other scaffolds might further increase the number of scaffolds yielding compound potency differences of more than 4 orders of magnitude. Taken together, these findings strongly indicated that several hundred scaffolds already exist in public domain compound data that generate compounds with differential activity

Table 3. Scaffold Classification Based on Potency and Selectivity Ranges[a]

(a) potency-based

potency range	# scaffolds	# cpds	# targets	max pot
		CDB		
0	6431	1.0	1.0	6.82
[0, 1]	2433	2.1	1.7	7.07
[1, 2]	1881	3.2	2.2	7.62
[2, 3]	1212	4.5	2.8	8.17
[3, 4]	667	7.2	3.6	8.68
[4, 11.72]	278	16.9	6.9	9.30
		BDB		
0	2823	1.0	1.0	6.59
[0, 1]	1232	1.9	1.7	6.96
[1, 2]	989	3.1	2.2	7.47
[2, 3]	658	4.8	2.7	7.90
[3, 4]	424	8.0	3.2	8.48
[4, 8.29]	165	17.7	5.4	9.18

(b) selectivity-based

selectivity range	# scaffolds	# cpds	# tps	max |sr|
		CDB		
0	1642	1.0	1.0	1.11
[0, 1]	948	2.2	2.8	0.93
[1, 2]	861	2.6	4.5	1.85
[2, 3]	560	4.0	11.8	2.70
[3, 4]	253	4.9	26.0	3.61
[4, 7.13]	52	15.6	208.6	4.63
		BDB		
0	1033	1.0	1.0	1.17
[0, 1]	491	2.6	2.4	1.14
[1, 2]	459	3.6	5.0	1.95
[2, 3]	295	4.7	6.9	2.70
[3, 4]	106	6.1	15.8	3.51
[4, 6.83]	42	9.2	31.3	4.74

[a] Scaffolds extracted from CDB and BDB are classified into six sets based on (a) potency and (b) selectivity value ranges. Reported are the number (#) of scaffolds, average number of compounds (cpds) and targets or target pairs (tps) per scaffold, and the average maximal logarithmic potency (max pot) or selectivity ratio (max |sr|).

Table 4. Scaffolds from CDB with High Potency-Based Discontinuity Scores[a]

rank	scaffold ID	PScS	#cpds	#targets	#targets ⟨MultiCpds⟩	#TargetCliffs
1	6445	1	30	24	**10**	**4**
2	9793	1	5	28	**1**	**1**
3	2269	1	4	3	**3**	**1**
4	1901	1	4	6	**1**	**1**
5	3866	1	3	9	**3**	**2**
16	2623	0.98	24	9	**2**	**1**
17	5821	0.98	9	3	**3**	**3**
27	5749	0.97	15	2	**1**	**1**
33	3659	0.96	27	3	**3**	**2**
34	11047	0.96	8	10	**4**	**2**
39	10707	0.95	46	22	**11**	**7**
40	8996	0.95	13	2	**2**	**1**
41	1927	0.95	10	8	**7**	**3**
48	5794	0.94	52	6	**3**	**3**
59	347	0.92	28	3	**3**	**2**
60	10775	0.92	14	2	**2**	**1**
61	8115	0.92	9	3	**3**	**3**
68	4196	0.90	16	15	**6**	**1**
73	2712	0.89	27	4	**4**	**2**
84	11159	0.87	21	2	**2**	**2**
85	9992	0.87	19	4	**4**	**2**
90	10539	0.86	24	10	**7**	**1**
91	3062	0.86	11	9	**1**	**1**
92	1105	0.86	9	3	**3**	**2**
94	2306	0.85	18	7	**1**	**1**
95	10235	0.85	12	2	**1**	**1**
96	12449	0.85	12	5	**2**	**3**
97	10483	0.85	12	3	**3**	**2**
98	2831	0.85	9	4	**2**	**1**
106	5285	0.84	10	3	**3**	**1**
111	3355	0.83	27	2	**2**	**2**
112	9749	0.83	9	5	**4**	**3**
117	6783	0.82	14	5	**4**	**1**
118	9066	0.82	14	3	**3**	**2**
126	5314	0.81	20	3	**3**	**2**
127	2849	0.81	19	3	**2**	**1**

[a] A total of 36 CDB scaffolds with PScS greater than 0.8 that represent more than two compounds that are active against more than one target are listed. For each scaffold, the score-based rank position (rank), discontinuity score (PScS), the number of unique compounds it represents (#cpds), the total number of targets (#targets), the number of targets with multiple active compounds (#targets <MultiCpds>), and the number of targets for which it forms activity cliffs (#TargetCliffs) are reported.

Table 5. Scaffolds from BDB with High Potency-Based Discontinuity Scores[a]

rank	scaffold ID	PScS	#cpds	#targets	#targets ⟨MultiCpds⟩	#TargetCliffs
1	1990	1	6	2	**1**	**1**
2	413	1	4	2	**2**	**2**
3	455	1	4	2	**2**	**2**
4	1161	1	3	2	**2**	**2**
5	851	0.99	32	6	**2**	**2**
7	363	0.98	17	4	**3**	**2**
11	312	0.97	9	3	**3**	**3**
14	1348	0.96	8	8	**7**	**3**
15	274	0.95	13	5	**5**	**5**
16	1304	0.95	13	4	**1**	**1**
20	1144	0.94	9	3	**3**	**1**
27	1506	0.92	13	4	**2**	**2**
28	1922	0.92	8	4	**1**	**1**
31	1120	0.91	13	2	**2**	**1**
35	204	0.90	31	2	**1**	**1**
36	606	0.90	24	5	**5**	**1**
41	819	0.89	9	3	**1**	**1**
42	2389	0.88	60	22	**7**	**4**
46	857	0.87	8	5	**3**	**3**
48	1106	0.86	8	5	**4**	**3**
51	1457	0.85	14	2	**2**	**1**
52	39	0.85	12	4	**2**	**1**
54	1169	0.84	10	6	**6**	**4**
56	1257	0.83	148	7	**6**	**2**
57	1109	0.83	40	9	**4**	**3**
58	193	0.83	35	2	**1**	**1**
59	972	0.83	27	2	**2**	**1**
60	833	0.83	14	2	**2**	**2**
61	1385	0.83	14	3	**2**	**1**
62	2363	0.83	11	2	**2**	**2**
63	1152	0.82	63	12	**7**	**4**
64	810	0.82	13	2	**2**	**1**
65	1425	0.82	9	3	**2**	**1**
68	1851	0.81	29	2	**1**	**1**
69	720	0.81	20	2	**2**	**1**
70	787	0.81	15	2	**2**	**1**

[a] A total of 36 BDB scaffolds with PScS greater than 0.8 that represent more than two compounds that are active against more than one target are listed. For each scaffold, the score-based rank position (rank), discontinuity score (PScS), the number of unique compounds it represents (#cpds), the total number of targets (#targets), the number of targets with multiple active compounds (#targets < MultiCpds>), and the number of targets for which it forms activity cliffs (#TargetCliffs) are reported.

against several targets and the tendency to produce activity (or selectivity) cliffs.

Cliff-Forming Scaffolds. The search for activity cliff-forming scaffolds was further refined by calculating a discontinuity score for each scaffold. This calculation involves systematic pairwise similarity and potency comparison of compounds containing the scaffold. High discontinuity scores approaching 1 indicate the presence of significant activity cliffs within the compound set. This scoring formalism is also applicable to selectivity-based score calculation because selectivity is expressed as a pairwise potency ratio.

Scaffold discontinuity scores were first calculated over all targets against which compounds with a particular scaffold were active (i.e., global scores), which provides a general measure for the propensity of a scaffold to form cliffs. Then, the scores were recalculated on a per-target basis, thus identifying target-dependent activity cliffs, if present, or target pair-dependent selectivity cliffs. For both global and target-based calculations, discontinuity scores of greater than 0.8 were considered. In our experience, this score level reliably indicates the presence of activity cliffs.

In CDB and BDB, 137 and 75 scaffolds were found, respectively, that achieved global potency-based discontinuity scores of greater than 0.8 and were represented by more than two compounds active against more than one target. The complete sets of these CDB and BDB scaffolds with potency values and intervals are provided in Tables S1 and S2 (Supporting Information), respectively. In Tables 4 and 5, 36 of these CDB and BDB scaffolds are reported, respectively. Furthermore, in Tables 6 and 7, all 34 CDB and all 23 BDB scaffolds are listed that produced selectivity-based discontinuity scores of greater than 0.8 and were represented by more than two compounds active against more than one target pair. Tables S3 and S4 (Supporting Information) list these scaffolds with associated selectivity ratios and selectivity intervals. Only two scaffolds formed both strong activity

Table 6. Scaffolds from CDB with High Selectivity-Based Discontinuity Scores[a]

rank	scaffold ID	#cpds	SScS	#TPs	#TPs ⟨MultiCpds⟩	#TPCliffs
1	9230	5	1	3	**1**	**1**
2	10683	4	1	34	**6**	**5**
3	4991	3	1	3	**3**	**3**
4	10707	16	0.99	27	**12**	**9**
5	2840	7	0.99	3	**3**	**2**
6	6582	3	0.99	10	**10**	**7**
7	12673	3	0.98	6	**3**	**3**
8	3211	3	0.97	3	**3**	**2**
9	3754	7	0.96	3	**2**	**1**
10	5304	5	0.96	6	**3**	**2**
11	572	5	0.95	3	**3**	**2**
12	7198	4	0.95	3	**3**	**2**
13	4266	3	0.95	3	**1**	**1**
14	8848	3	0.95	6	**3**	**2**
15	143	7	0.94	3	**3**	**2**
16	5834	6	0.93	6	**3**	**3**
17	7991	21	0.92	10	**10**	**3**
18	3153	3	0.92	3	**1**	**1**
19	10439	5	0.91	6	**5**	**3**
20	973	3	0.91	21	**1**	**1**
21	12298	4	0.90	3	**3**	**2**
22	1779	3	0.90	6	**6**	**3**
23	12627	3	0.90	3	**3**	**2**
24	2634	6	0.89	6	**6**	**3**
25	2712	8	0.88	6	**5**	**2**
26	1087	3	0.86	6	**3**	**1**
27	6611	13	0.84	3	**3**	**1**
28	9649	4	0.84	3	**2**	**2**
29	6332	17	0.83	35	**21**	**3**
30	12747	4	0.83	3	**3**	**1**
31	2008	3	0.82	3	**3**	**2**
32	6576	3	0.82	6	**6**	**2**
33	9011	66	0.81	3	**1**	**1**
34	347	23	0.81	3	**3**	**2**

[a] All 34 CDB scaffolds with SScS greater than 0.8 that represent more than two compounds that are active against more than one target are listed. For each scaffold, the score-based rank position (rank), discontinuity score (SScS), the number of unique compounds it represents (#cpds), the total number of target pairs (#TPs), the number of target pairs with multiple active compounds (#TPs ⟨MultiCpds⟩), and the number of selectivity cliffs it forms (#TPCliffs) are reported.

Table 7. Scaffolds from BDB with High Selectivity-Based Discontinuity Scores[a]

rank	scaffold ID	SScS	#cpds	#TPs	#TPs ⟨MultiCpds⟩	#TPCliffs
1	2115	1	3	3	**3**	**3**
2	312	0.99	9	3	**3**	**3**
3	381	0.99	6	3	**3**	**3**
4	424	0.99	5	3	**3**	**3**
5	196	0.97	4	2	**1**	**1**
6	857	0.97	4	7	**1**	**1**
7	1425	0.96	6	3	**1**	**1**
8	733	0.96	3	3	**1**	**1**
9	1615	0.94	6	4	**1**	**1**
10	1008	0.94	3	4	**3**	**2**
11	1169	0.93	8	13	**8**	**4**
12	1001	0.92	4	4	**2**	**2**
13	1100	0.90	6	3	**3**	**2**
14	2005	0.88	8	3	**3**	**2**
15	511	0.87	10	3	**1**	**1**
16	466	0.87	9	3	**3**	**3**
17	1238	0.87	5	3	**2**	**1**
18	1152	0.86	21	9	**4**	**2**
19	362	0.85	6	3	**3**	**1**
20	1319	0.85	3	3	**3**	**3**
21	1106	0.84	5	10	**6**	**4**
22	2208	0.81	45	6	**6**	**2**
23	446	0.81	19	26	**9**	**4**

[a] All 23 BDB scaffolds with SScS greater than 0.8 that represent more than two compounds that are active against more than one target are listed. For each scaffold, the score-based rank position (rank), discontinuity score (SScS), the number of unique compounds it represents (#cpds), the total number of target pairs (#TPs), the number of target pairs with multiple active compounds (#TPs ⟨MultiCpds⟩), and the number of selectivity cliffs it forms (#TPCliffs) are reported.

and selectivity cliffs, scaffold 10 707 (Table 4 and 6) and scaffold 1152 (Table 5 and 7).

Many of the scaffolds in Tables 4 and 5 form activity cliffs against multiple targets. For example, the top-scoring scaffold in Table 4 (rank 1) corresponds to 30 compounds that are active against 24 targets and form significant activity cliffs against four of these targets. The scaffold at score rank 17 corresponds to nine compounds that are active against three targets and form activity cliffs in each case. Furthermore, the scaffold at score rank 39 is represented by 46 compounds active against 22 targets, forming activity cliffs for seven of these targets. In Table 5, the scaffold at rank 5 corresponds to 32 compounds active against six targets and forming activity cliffs against two of them. Moreover, the scaffold at rank 15 is represented by 13 compounds that form activity cliffs for all five targets they are active against. Similar observations were made for selectivity cliffs. For example, in Table 6, the scaffold at rank 4 is represented by 16 compounds forming nine selectivity cliffs, and the scaffold at rank 2 in Table 7 corresponds to nine compounds forming three selectivity cliffs. A total of 25 CDB (Table 6) and 15 BDB (Table 7) scaffolds were found to yield compounds forming multiple selectivity cliffs. In many instances, fewer than 10 compounds representing a particular scaffold produced activity or selectivity cliffs for multiple targets. Thus, thorough chemical exploration of these scaffolds was not required for cliff formation.

Multi-target Activity Cliffs. Which types of scaffolds form multi-target activity cliffs? Figure 4a and b show representative CDB and BDB scaffolds that are represented by more than two compounds and form at least three activity cliffs for distinct targets. The target annotations of the scaffolds are shown in Figures S1 and S2 (Supporting Information). These scaffolds are of different sizes and chemical natures ranging from small generic structures, e.g., simple aliphatic rings such as cyclohexane, tetrahydrofuran, or pyrrolidine, to complex multiring scaffolds. Hence, multi-target activity cliff scaffolds were diverse, and there were no apparent preferences for specific chemotypes. Importantly, many of these scaffolds formed activity cliffs for targets belonging to different communities. Similar observations were made for multi-target selectivity cliff scaffolds. These BDB and CDB scaffolds and their selectivity annotations are shown in Figures S3 and S4 (Supporting Information), respectively. Thus, the formation of multiple-target activity or selectivity cliffs was not limited to closely related targets but also involved different classes of targets. Figure 5 shows representative multi-target activity cliffs. Different pairs of compounds representing a scaffold introduce activity cliffs of varying magnitudes against different targets. Such

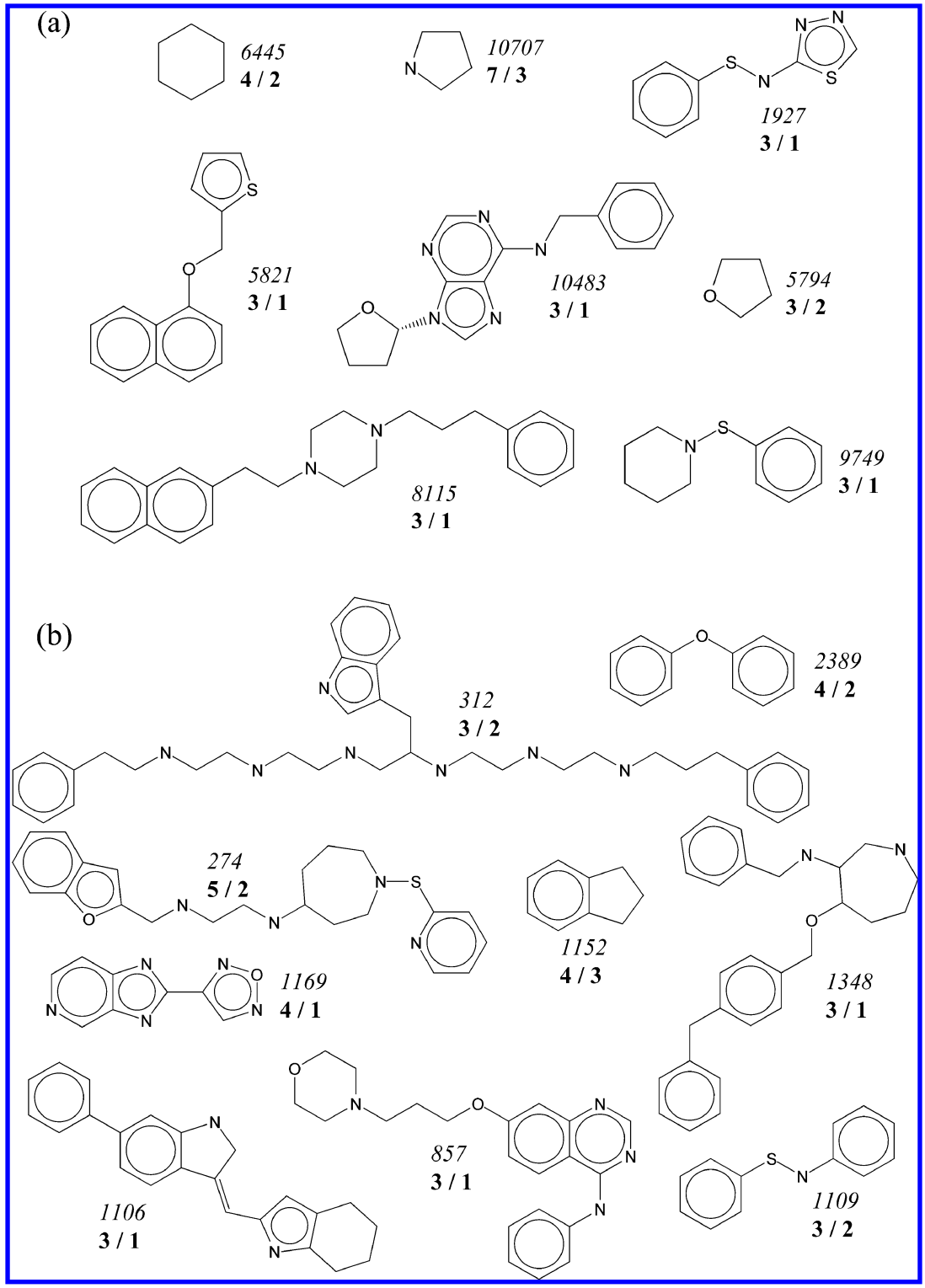

Figure 4. Scaffolds with high propensity to form activity cliffs. Scaffolds are shown that produce activity cliffs for at least three targets (a, CDB; b, BDB). Scaffolds IDs (in italics) and target/community numbers (bold) are provided. For example, "3/2" means that a scaffold forms activity cliffs for three targets in two communities.

patterns are representative for many multi-target activity cliff scaffolds. In Table S5 (Supporting Information), we provide SMILES representations of the complete set of multi-target activity cliff scaffolds represented by multiple compounds. Table S6 (Supporting Information) provides a corresponding list of multi-target selectivity cliff scaffolds.

CONCLUSIONS

In this study, we have primarily explored the question of whether molecular scaffolds exist that display a general tendency to form activity or selectivity cliffs against different targets. From the ChemblDB and BindingDB databases, scaffolds and associated compound activity data were systematically extracted, target communities were estab-

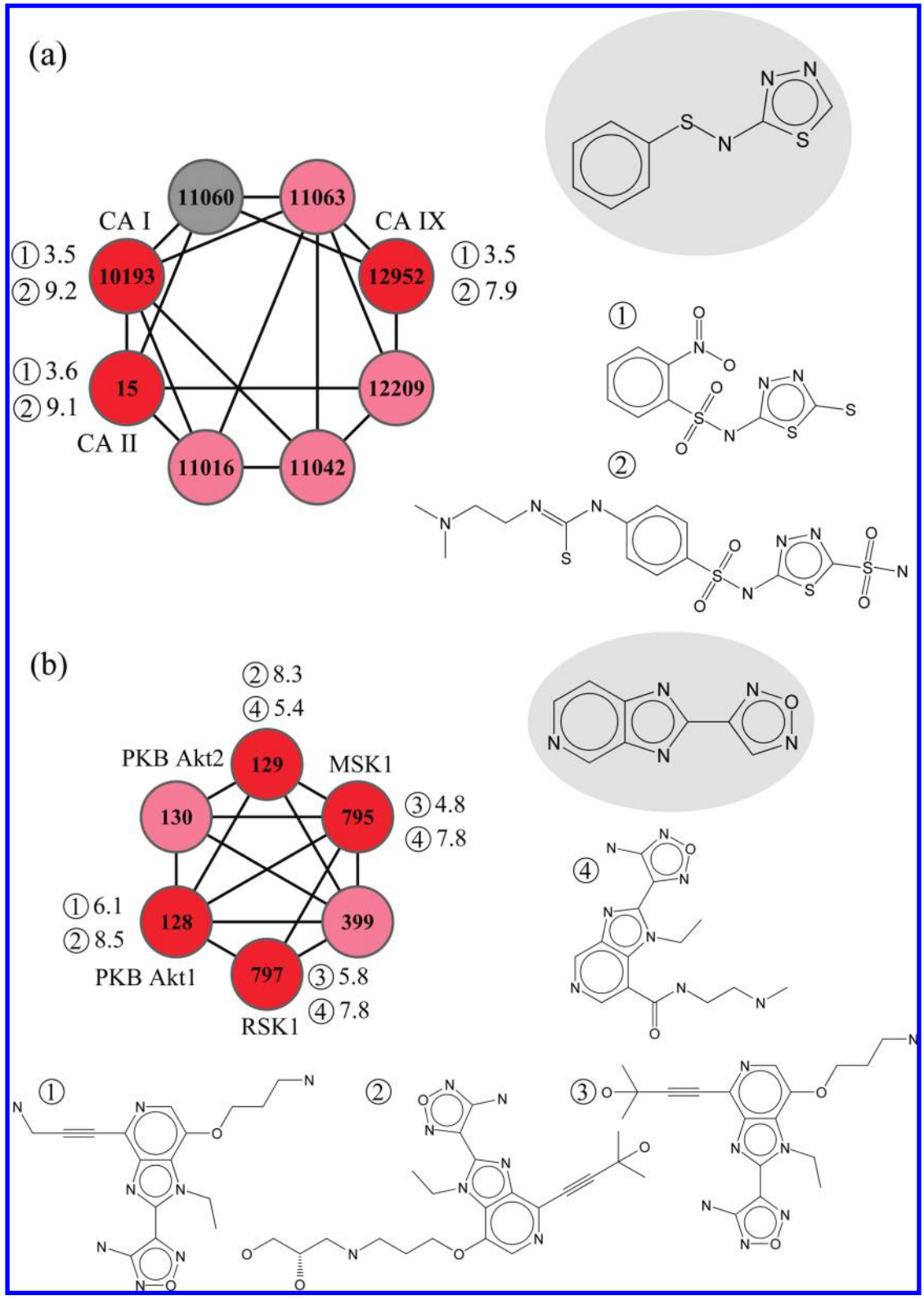

Figure 5. Representative multi-target activity cliffs. Two representative scaffolds are shown that form multi-target activity cliffs (a, CDB; b, BDB). The scaffolds are shown on a gray background. Nodes represent targets. Two nodes are connected if they share compounds containing the scaffold. A target is colored gray if only a single compound is reported to be active against it (hence, for such targets, activity cliffs cannot be detected). A node is colored red if compounds active against the target yield a PScS value greater than 0.8 (indicating strong discontinuity). Representative compounds containing the scaffold are shown and labeled. For these compounds, negative logarithmic potency values for individual targets are reported. The differences between these values indicate the magnitude of the target-dependent activity cliffs the compounds form. Target abbreviations: CA, carbonic anhydrase; MSK, mitogen- and stress-activated protein kinase; PKB, protein kinase B; RSK, ribosomal S6 kinase.

lished, and target-selective and cliff-forming scaffolds were identified. Consistent with our early findings, compound selectivity data is only sparse in public domain compound databases, which currently limits a reliable assignment of target-selective scaffolds. By contrast, we have identified a significant number of scaffolds that are represented by

compounds forming activity or selectivity cliffs against multiple targets. These targets are often unrelated and occur in different target communities. Many multi-target cliff scaffolds have not yet been extensively explored; i.e., they are currently represented by only fewer than 10 compounds, yet they already display strong tendencies of cliff formation. Multi-target activity cliff scaffolds are in part promiscuous in nature and able to bind to different types of targets. However, these scaffolds yield compounds that form substantial activity cliffs for different targets and are thus phenotypically distinct from "non-specific" molecules. Thus, multi-target activity cliff scaffolds might be interesting candidates for compound optimization when considered on a per-target basis. Yet it should be taken into account that compounds derived from such scaffolds might often be highly potent against multiple targets. However, depending on the therapeutic application, the use of multi-target activity cliff scaffolds might also be desirable, for example, when optimizing compounds for series of closely related targets having similar or overlapping functions. The collection of multi-target cliff scaffolds we have identified and provide as part of this study should be helpful in order to evaluate and prioritize scaffolds for compound optimization efforts.

Supporting Information Available: Tables S1−S4 list scaffolds with high potency- or selectivity-based discontinuity scores, and Tables S5 and S6 provide SMILES representations for multi-target activity and selectivity cliff scaffolds that are represented by at least three compounds. Figures S1−S4 show target and target pair annotations of multi-target activity and selectivity cliff scaffolds. This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

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

(2) Merlot, C.; Domine, D.; Cleva, C.; Church, D. J. Chemical substructures in drug discovery. *Drug Discovery Today* **2003**, *8*, 594–602.

(3) Brown, N.; Jacoby, E. On Scaffolds and Hopping in Medicinal Chemistry. *Mini Rev. Med. Chem.* **2006**, *6*, 1217–1229.

(4) Zhao, H. Scaffold Selection and Scaffold Hopping in Lead Generation: A Medicinal Chemistry Perspective. *Drug Discovery Today* **2007**, *12*, 149–155.

(5) Wang, J.; Hou, T. Drug and Drug Candidate Building Block Analysis. *J. Chem. Inf. Model.* **2010**, *50*, 55–67.

(6) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. RECAP--retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511–522.

(7) Ertl, P.; Jelfs, S.; Mühlbacher, J.; Schuffenhauer, A.; Selzer, P. Quest for the Rings. In Silico Exploration of Ring Universe To Identify Novel Bioactive Heteroaromatic Scaffolds. *J. Med. Chem.* **2006**, *49*, 4568–4573.

(8) Pitt, W. R.; Parry, D. M.; Perry, B. G.; Groom, C. R. Heteroaromatic Rings of the Future. *J. Med. Chem.* **2009**, *52*, 2952–2963.

(9) Krier, M.; Bret, G.; Rognan, D. Assessing the Scaffold Diversity of Screening Libraries. *J. Chem. Inf. Model.* **2006**, *46*, 512–524.

(10) Lipkus, A. H.; Yuan, Q.; Lucas, K. A.; Funk, S. A.; Bartelt, W. F., III; Schenck, R. J.; Trippe, A. J. Structural Diversity of Organic Chemistry. A Scaffold Analysis of the CAS Registry. *J. Org. Chem.* **2008**, *73*, 4443–4451.

(11) Hu, Y.; Bajorath, J. Scaffold Distributions in Bioactive Molecules, Clinical Trials Compounds, and Drugs. *ChemMedChem* **2010**, *5*, 187–190.

(12) Lameijer, E.; Kok, J. N.; Bäck, T.; Ijzerman, A. P. Mining a Chemical Database for Fragment Co-Occurrence: Discovery of "Chemical Clichés". *J. Chem. Inf. Model.* **2007**, *46*, 553–562.

(13) Sutherland, J. J.; Higgs, R. E.; Watson, I.; Vieth, M. Chemical Fragments as Foundations for Understanding Target Space and Activity Prediction. *J. Med. Chem.* **2008**, *51*, 2689–2700.

(14) Schuffenhauer, A.; Ertl, P.; Roggo, S.; Wetzel, S.; Koch, M. A.; Waldmann, H. The scaffold tree--visualization of the scaffold universe by hierarchical scaffold classification. *J. Chem. Inf. Model.* **2007**, *47*, 47–58.

(15) Wetzel, S.; Klein, K.; Renner, S.; Rauh, D.; Oprea, T. I.; Mutzel, P.; Waldmann, H. Interactive Exploration of Chemical Space with Scaffold Hunter. *Nat. Chem. Biol.* **2009**, *5*, 581–583.

(16) Siegel, M. G.; Vieth, M. Drugs in Other Drugs: A New Look at Drugs as Fragments. *Drug Discovery Today* **2007**, *12*, 71–79.

(17) Villar, H. O.; Hansen, M. R. Computational Techniques in Fragment-Based Drug Discovery. *Curr. Top. Med. Chem.* **2007**, *7*, 1509–1513.

(18) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. Recent Developments in Fragment-based Drug Discovery. *J. Med. Chem.* **2008**, *51*, 3661–3680.

(19) Evans, B. E.; Rittle, K. E.; Bock, M. G.; Dipardo, R. M.; Freidinger, R. M.; Whitter, W. L.; Lundell, G. F.; Veber, D. F.; Anderson, P. S. Methods for Drug Discovery: Development of Potent, Selective, Orally Effective Cholecystokinin Antagonists. *J. Med. Chem.* **1988**, *31*, 2235–2246.

(20) Horton, D. A.; Bourne, G. T.; Smythe, M. L. The Combinatorial Synthesis of Bicyclic Privileged Structures or Privileged Substructures. *Chem. Rev.* **2003**, *103*, 893–930.

(21) Constantino, L.; Barlocco, D. Privileged Substructures as Leads in Medicinal Chemistry. *Curr. Med. Chem.* **2006**, *13*, 65–85.

(22) Schnur, D. M.; Hermsmeier, M. A.; Tebben, A. J. Are Target-Family-Privileged Substructures Truly Privileged. *J. Med. Chem.* **2006**, *49*, 2000–2009.

(23) Hu, Y.; Wassermann, A. M.; Lounkine, E.; Bajorath, J. Systematic Analysis of Public Domain Compound Potency Data Identifies Selective Molecular Scaffolds across Druggable Target Families. *J. Med. Chem.* **2010**, *53*, 752–758.

(24) Hu, Y.; Bajorath, J. Exploring Target-Selectivity Patterns of Molecular Scaffolds. *ACS Med. Chem. Lett.* [Online] DOI: 10.1021/ml900024v.

(25) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: a Web-Accessible Database of Experimentally Determined Protein-Ligand Binding Affinities. *Nucleic Acids Res.* **2007**, *35*, D198–D201.

(26) PubChem. http://pubchem.ncbi.nlm.nih.gov/ (accessed October 1, 2009).

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

(28) 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.

(29) 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.

(30) ChemblDB. http://www.ebi.ac.uk/chembl/ (accessed January 2, 2010).

(31) 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.

(32) 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.

(33) 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.

(34) *MACCS Structural Keys*; Symyx Software: San Ramon, CA, 2005.

(35) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983–996.

(36) *Scitegic Pipeline Pilot*, Student ed., version 6.1; Accelrys, Inc.: San Diego, CA, 2007.

CI100059Q