JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Systematic Identification and Classification of Three-Dimensional Activity Cliffs

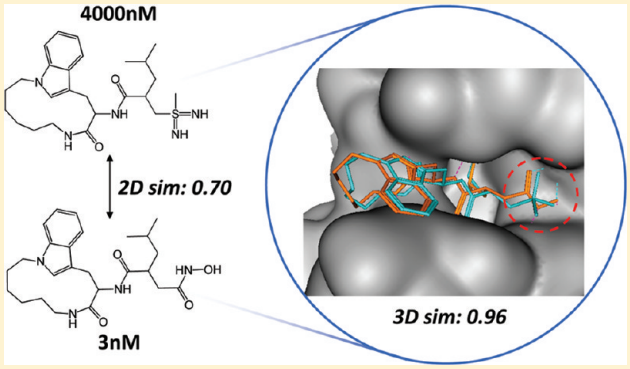
Ye Hu,[†,§] Norbert Furtmann,[†,‡,§] Michael Gütschow,[‡] and Jürgen Bajorath*[,†]

[†]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

[‡]Pharmaceutical Institute, University of Bonn, An der Immenburg 4, D-53121 Bonn, Germany

Supporting Information

ABSTRACT: Activity cliffs were systematically extracted from public domain X-ray structures of targets for which complexes with multiple ligands were available, following the concept of three-dimensional (3D) cliffs. Binding modes of ligands with well-defined potency measurements were compared in a pairwise manner, and their 3D similarity was calculated using a previously reported property density function-based method taking conformational, positional, and chemical differences into account. Requiring the presence of at least 80% 3D similarity and a potency difference of at least 2 orders of magnitude as cliff criteria, a total of 216 well-defined 3D activity cliffs were detected in the Protein Data Bank (PDB). These 3D-cliffs involved a total of 269 ligands active against 38 different targets belonging to 17 protein families. For 255 of these compounds, binding modes were available at high crystallographic resolution. All 3D-cliffs were analyzed in detail and assigned to different categories on the basis of crystallographic interaction patterns. In many instances, differences in ligand−target interactions suggested plausible causes for origins of 3D-cliffs. In other cases, short-range interactions seen in X-ray structures were insufficient to deduce possible reasons for cliff formation. The 3D-cliffs described herein further advance the rationalization of activity cliffs at the level of ligand−target interactions and should also be useful for other applications such as the calibration of energy functions for structure-based design. The pool of identified activity cliffs is provided to enable subsequent structure-based analyses of cliffs.



INTRODUCTION

Activity cliffs, pairs or groups of structurally similar compounds with large potency differences,[1−3] have mostly been studied on the basis of molecular graphs and two-dimensional (2D) similarity measures.[3−5] In medicinal chemistry, activity cliffs have become popular as focal points of systematic structure−activity relationship (SAR) analysis to identify structural modifications that determine SAR characteristics.[3,6] The assessment of activity cliffs requires the consideration of experimental potency differences and chemical/structural similarities of test compounds. In addition, it is essential to clearly define potency and similarity criteria for cliff formation.[3] For similarity analysis, a variety of molecular representations and similarity measures might be considered. Only a few studies have thus far departed from 2D similarity assessment and extended the study of activity cliffs into 3D space. For example, hypothetical compound conformations were utilized to calculate 3D molecular representations, define activity cliffs on the basis of these representations, and generate consensus cliffs for 2D and 3D descriptors by data fusion of calculated similarity values.[7] Furthermore, activity cliffs detected on the basis of 2D similarity measures were interpreted taking ligand-target interactions in X-ray structures of corresponding complexes into account.[8] These studies focused on exemplary activity cliffs extracted from different compound classes for which complex structures were available. In addition, in order to define activity cliffs on the basis of structural information, ligand−target force field interaction energies were computed for active compounds in complex structures and recorded in an interaction fingerprint format.[9] These 3D interaction fingerprints were then used as ligand representations to calculate similarity values and structure−activity landscape index scores[4] from which activity cliffs were derived.[9] Hence, such structure-based activity cliffs take information from ligand−target complexes into account, albeit indirectly via interaction energy calculations. Recently, 3D-cliffs have been introduced by directly comparing experimentally determined compound binding modes[10] using a 3D similarity function.[11] For two compound activity classes, 3D-cliffs were systematically derived and compared to standard 2D-cliffs. It was found that only ∼40% of 3D-cliffs (characterized by high similarity of

Received: March 23, 2012
Published: May 21, 2012

compound binding modes) could be reconciled on the basis of standard 2D similarity calculations.[10] These findings reflected the presence of substantial differences between pharmacophore resemblance of active compounds and global similarity of molecular graphs, at least for the data sets under study.

In this study, we have systematically extracted 3D-cliffs from ligand−target complex structures available in the PDB.[12−14] Compounds for which well-defined potency data and experimental structures were available were assembled. For ligands active against the same target, their binding modes were compared to identify 3D-cliffs. All detected cliffs were analyzed in detail and classified on the basis of ligand−target interaction information. The analysis revealed more than 200 3D-cliffs falling into different categories. The major goal of this extensive structure and activity data mining efforts has been to identify all currently available 3D-cliffs applying conservative criteria for cliff formation.

MATERIAL AND METHODS

Data Preparation. In the first step, PDB ligand-target pairings and ligands for which detailed potency measurements were available were determined. For this purpose, UniProt[15] accession IDs (UniProtIDs) were collected for all human target proteins available in ChEMBL[16,17] (release 10) and BindingDB[18,19] (December 2011 release). Then, the list of unique UniProtIDs was queried in the PDB utilizing its UniProtID search function. Targets were selected for which complex structures with two or more small molecular ligands were available. These complexes were organized into target sets on the basis of conserved UniProtIDs. Ligands (and corresponding complex structures) were retained if they were annotated with defined potency measurements (i.e., K_i or IC_{50}). Potency annotations were manually curated by eliminating all approximate measurements reporting ">", "<", or "∼" relations. Moreover, if multiple potency values of a ligand from different data sources or experiments differed by more than one order of magnitude, the ligand was no longer considered. If all potency values fell within an order of magnitude, their geometric mean was calculated to generate the final potency annotation of the ligand. In addition, ligands that were natural substrates or inorganic compounds were excluded. It should be noted that the so-derived target sets were found to contain only X-ray but no NMR structures. Target sets were assigned to different protein families following the UniProt target family organization.[15]

Similarity Evaluation. For the assessment of 3D similarity, a previously developed 3D similarity function was applied that compares compound binding modes using property density functions taking conformational, positional, and atomic property differences into account.[11] For comparison, standard 2D Tanimoto similarity[20] calculations were carried out using MACCS structural keys[21] as descriptors.

For the calculation of pairwise ligand binding mode similarity, α-carbon atoms of target structures in complex with different ligands (i.e., structures of a target in different PDB entries) were optimally superposed using the sequence/structure alignment function of the Molecular Operating Environment (MOE).[22] The superposition defined the positions of ligands for 3D similarity calculation. The normalized overlap of property density functions calculated for a ligand pair was used as a measure of 3D similarity. Therefore, each ligand atom was represented by a spherically symmetric Gaussian density function centered at the position of the nucleus with a width determined by the van der Waals radius of the atom. A global density function was defined for different atomic properties by the weighted sum of the density functions of all atoms of a ligand. The atomic Gaussians were weighted relative to selected atom properties. The atomic property density weight was set to 1 if the respective atom contained this property and to 0 if the property was absent. Four atomic properties were considered: aromatic character, hydrogen bond acceptor potential, hydrogen bond donor potential, and hydrophobic character. The overlap of the property density functions of two binding modes was determined as the sum (Gaussian) of the individual property density functions:

$$F(X, Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{w_i^p w_j^p + w_i^q w_j^q + \ldots}{mn} \left(\frac{a^2}{2\pi(r_i^2 + r_j^2)}\right)^{3/2} \exp\left\{-\frac{a^2}{2(r_i^2 + r_j^2)} |x_i - y_j|^2\right\}$$

$F(X,Y)$: overlap of property density functions of conformations X and Y. X,Y: matrices of spatial atom coordinates for the two molecules with dimension $3 \times m$ and $3 \times n$, respectively. m, n: numbers of atoms in molecules X and Y, respectively. x_i: vector of coordinates of atom i in conformation X. w_i^p: weight of atom i with respect to property p. $w_i^p = 1$ if atom i has property p, otherwise $w_i^p = 0$. a: scaling factor; set to 2 in our calculations. r_i: van der Waals radius of atom i.

Atomic properties were calculated using MOE according to the following definitions: $w_i^{\text{aromatic}} = 1$ if atom i is in a ring satisfying the Hückel rule and is sp^2 hybridized; $w_i^{\text{donor}} = 1$ if atom i is in pharmacophore class "donor" or in class "base";$w_i^{\text{acceptor}} = 1$ if atom i is in pharmacophore class "acceptor" or in class "acid"; $w_i^{\text{hydrophobic}} = 1$ if atom i is in pharmacophore class "hydrophobe".

A normalization procedure was applied to obtain 3D similarity values between 0 and 1. Thus the overlap of the molecular property density functions was divided by the mean self-overlap of the ligand binding modes:

$$F^{\text{norm}}(X, Y) = \frac{F(X, Y)}{\frac{1}{2}F(X, X)F(Y, Y)}$$

This 3D similarity approach has been previously reported and shown to be a sensitive measure of positional, conformational, and chemical property differences between compound binding modes.[11]

Activity Cliff Criteria. For the formation of activity cliffs, the following criteria were applied. For statistical analysis, ligand pairs with 2D or 3D similarity above 80%, 85%, 90%, or 95% and potency differences of 1 or 2 orders of magnitude were considered to form cliffs. For each combination of a similarity threshold value and potency difference, cliffs were separately calculated. For the final selection of 3D-cliffs and structural analysis, a 3D similarity threshold of at least 80% and a potency difference of at least 2 orders of magnitude were required as cliff criteria.

3D-Cliff Categorization. Superposed ligand−target complexes representing 3D-cliffs were analyzed in detail using MOE. Ligand−target interactions were determined and compared for 3D-cliff partners. On the basis of differences in

dx.doi.org/10.1021/ci300158v | J. Chem. Inf. Model. 2012, 52, 1490−1498

crystallographically defined interactions between cliff forming compounds, 3D-cliffs were assigned to six categories:

(1) *H-bond and/or ionic interactions:* A 3D-cliff was assigned to this category if differences in intermolecular interactions involving the cliff forming compounds only included hydrogen bonds or salt bridges.
(2) *Lipophilic/aromatic groups:* Interaction differences between cliff partners only included lipophilic or aromatic groups. The presence or absence of lipophilic/aromatic substituents often led to notable differences in the shape complementarity between ligands and the target binding site (e.g., by only partly or completely filling a hydrophobic pocket).
(3) *Water molecules:* Interaction differences were limited to the presence or absence of water-mediated hydrogen bonds.
(4) *Stereochemistry:* Cliff forming ligands were stereoisomers, leading to differences in shape complementarity. It should be noted that stereochemistry-dependent differences in hydrogen bond patterns were not detected in our set of 3D-cliffs.
(5) *Multiple effects:* Differences included several types of interactions, as defined above.
(6) *No apparent interaction differences:* Ligands forming a 3D-cliff displayed no apparent interaction differences or only very subtle differences. Hence, in these instances, there was no possible structural rationale for cliff formation.

Structure and Cliff Representations. Structural representations and 2D interaction diagrams[23] were generated with MOE. For target sets containing at least 15 3D-cliffs, a network representation was designed to visualize relationships between cliff forming compounds. In this network, nodes represented ligands and edges 3D-cliffs. In addition, nodes were color-coded according to the potency values of corresponding ligands using a continuous spectrum from black (lowest potency in the set) to red (highest potency). Networks were drawn with Cytoscape.[24]

RESULTS AND DISCUSSION

Data Sets. On the basis of human targets for which ligands were available in BindingDB or ChEMBL, a list of 2426 unique UniProtIDs was assembled and used for searching the PDB. From the PDB, a total of 12 145 ligand-target complex structures were retrieved. However, for ligands in only 4718 of these complexes defined potency measurements were available. Furthermore, for ligands in 2112 complexes, potency measurements were consistently and explicitly defined, and these complexes were grouped into 265 unique target sets. For only 93 ligands (i.e., < 5%), qualifying multiple potency measurements were available. Only 189 of our target sets contained at least two complex structures. For two of these sets, i.e., inhibitors of farnesyl pyrophosphate synthase and S-adenosylmethionine decarboxylase proenzyme, target structures could not be accurately superposed. Hence, complex structures belonging to a total of 187 target sets were ultimately subjected to the search for activity cliffs. On average, these target sets contained ~11 complex structures with different ligands. Importantly, for ligands in all selected complexes, high-confidence potency data were available, a prerequisite for meaningful activity cliff assessment. The summary of the analysis is reported as a flowchart in **Figure S1** of the Supporting Information.

Activity Cliffs. For the 187 target sets, 3D- and 2D-cliffs falling into eight categories were systematically identified on the basis of different similarity threshold values (i.e., 0.80, 0.85, 0.90, 0.95) and potency differences (i.e., at least 10- or 100-fold). For each category, 3D- and 2D-cliffs were compared. The results are reported in Table 1. In general, the numbers of 3D- and 2D-cliffs differed significantly in each category. In addition, with increasing stringency of similarity and potency criteria, the numbers of 3D- and 2D-cliffs decreased, as to be expected. For instance, when the most stringent criteria were applied, i.e., at least 95% structural similarity and a 100-fold difference in potency, only 16 3D-cliffs and 15 2D-cliffs were detected in 11 and 10 target sets, respectively. By contrast, when the least stringent criteria were applied, i.e., at least 80% structural similarity and a 10-fold potency difference, 740 3D- and 452 2D-cliffs were identified in 74 and 64 sets, respectively. The number of 3D-cliffs was always larger than the number of 2D-cliffs. Thus, conservation of 3D- and 2D-cliffs was overall limited, ranging from 5 to 222 conserved cliffs from the most to the least stringent cliff categories. Taken together, these findings indicated that many compounds with similar binding modes and chemical feature distributions qualifying for the formation of 3D-cliffs were not considered 2D-cliffs (because their two 2D structures were overall less similar). It follows that the assignment of activity cliffs on the basis of calculated 2D similarity values might often be prone to false-negatives. For subsequent analysis, 216 3D-cliffs were selected falling into the 80% similarity and 100-fold potency difference category. These cliffs were formed by a total of 269 ligands. For only 17 (~6%) of these ligands, multiple potency measurements were available, all of which fell into the same order of magnitude (as required by our selection criteria to ensure the exclusive use of high confidence activity data). For all cliffs, PDB entry information and sources of activity data are reported in **Table S1** of the Supporting Information.

Table 1. Activity Cliffs[a]

thresholds		number of				
potency difference	similarity	3D-cliffs	sets with 3D-cliffs	2D-cliffs	sets with 2D-cliffs	3D/2D-cliffs
10-fold	0.80	740	74	452	64	222
	0.85	404	62	268	50	107
	0.90	216	50	145	34	62
	0.95	78	28	62	20	21
100-fold	0.80	216	38	126	27	48
	0.85	104	29	73	22	25
	0.90	47	20	37	17	14
	0.95	16	11	15	10	5

[a]For each combination of similarity threshold values and potency differences, the numbers of 3D-, 2D-, and conserved 3D/2D-cliffs are reported. In addition, the numbers of target sets in which cliffs were detected are provided.

3D-Cliffs. We then studied the 216 selected 3D-cliffs in detail, first from a statistical and then from a structural point of view.

Distribution of 3D-Cliffs. These cliffs were distributed over 38 target sets belonging to 17 different protein families, as reported in Table 2. These families contained rather diverse targets. The serine protease and metalloprotease families contained the most target sets (five and six, respectively).

Table 2. Target Family Classification[a]

target family	number of targets	ligands	3D-cliffs	3D/2D-cliffs
apoptosis protein	1	6	1	0
cytosolic other-chaperone	1	44	17	2
cytosolic other-motor protein	1	21	4	2
hydrolase-glycosidase	1	5	1	1
hydrolase-phosphodiesterase	3	39	6	5
hydrolase-protease-aminopeptidase	1	10	3	1
hydrolase-protease-aspartic protease	1	106	18	7
hydrolase-protease-metalloprotease	6	86	31	5
hydrolase-protease-serine protease	5	256	43	6
hydrolase-protein phosphatase	1	70	3	0
hydrolase-other	2	88	13	2
oxidoreductase	3	39	11	3
transcription factor-nuclear receptor	4	53	7	1
transferase/isomerase	1	6	1	0
transferase-Ser/Thr protein kinase	4	269	39	8
transferase-Tyr protein kinase	2	40	17	7
transferase-other	1	3	1	0

[a]For each target family, the numbers of targets, crystallographic ligands, and 3D- and conserved 3D/2D-cliffs are reported.

The numbers of ligands, 3D-cliffs, and conserved 3D/2D-cliffs differed significantly for the 17 families. Most ligands and 3D-cliffs were found for the serine protease (256 ligands forming 43 3D-cliffs) and the Ser/Thr protein kinase family (269 ligands forming 39 cliffs). By contrast, the glycosidase family contained only one target set with five ligands forming a single 3D-cliff. The distribution of 3D-cliffs mirrored at least to some extent the popularity of crystallographic targets. For example, many proteases and kinases are therapeutically relevant and have already been intensely investigated. For these targets, large numbers of inhibitors are available and crystallization conditions are often well-established, making it relatively easy to obtain complex X-ray structures.

Binding Mode Quality. In addition to high-confidence compound potency data, the structural integrity of binding modes also influences the determination of 3D-cliffs. In Table 3, we report the crystallographic resolution of structures that

Table 3. Crystal Structure Resolutions[a]

resolution range	number of ligands
≤1.5 Å	14
>1.5 Å, ≤2 Å	142
>2 Å, ≤2.5 Å	99
>2.5 Å, ≤3 Å	14

[a]For each resolution range, the number of crystallographic ligands forming 3D-cliffs is reported.

were the source of the 216 selected 3D-cliffs. For 156 of the 269 ligands involved in the formation of these cliffs, binding modes were resolved at 2 Å or higher resolution and for only 14 ligands, and binding modes were available at medium resolution between 2.5 and 3.0 Å. Hence, in addition to the potency data used for the identification of these 3D-cliffs, ligand binding modes generally were of high confidence.

Coordinated Formation of 3D-Cliffs. Six target sets belonging to six different protein families were found to contain at least 15 3D-cliffs. For these sets, activity cliff networks were generated (Figure 1). With 26 cliffs, inhibitors of cyclin-dependent kinase 2 formed the largest number of 3D-cliffs per target set. The network topologies of the six target sets differed. For example, the coagulation factor II (thrombin) and leukotriene A4 hydrolase inhibitor sets contained comparable numbers of 24 and 23 3D-cliffs, respectively, but displayed rather different network topologies. Similar observations were made, for example, for inhibitors of heat shock protein 90 and tyrosine kinase Src. However, despite these differences, in most cases 3D-cliffs were formed in a coordinated manner, i.e., ligands were involved in the formation of multiple and overlapping cliffs, similar to observations made for 2D-cliffs.[2] Overall, 111 of the 122 3D-cliffs displayed in Figure 1 were coordinated and only 11 cliffs were formed by isolated ligand pairs.

Structural Analysis of 3D-Cliffs. All 216 3D-cliffs were also systematically analyzed by studying the structural alignments from which they were derived and the specifics of their ligand−target interactions. Several different aspects were taken into consideration.

Calculated versus Observed 3D Similarity. The analysis initially focused on the comparison of the 216 pairs of binding modes that qualified as 3D-cliffs. We determined whether there were instances where high calculated 3D similarity was apparently not consistent with visual inspection of binding modes and the spatial relations of functional groups involved in ligand−target interactions. In previous studies,[10,11] the property density function overlap-based approach to quantifying 3D similarity of bound ligands was found to be a robust and stable measure of binding mode resemblance, as mentioned above. Our current analysis provided further support for this conjecture. Among the 216 3D-cliffs, we did not detect any instances where 3D similarity of at least 80% did not correspond to clear binding mode correspondence (that was interpretable in chemical terms). This was consistently the case, although the 2D similarity of 3D-cliffs fell often clearly below the 80% level. The 2D similarity distribution for all 3D-cliffs is reported in **Figure S2** of the Supporting Information. The peak of the distribution is at similarity value of 0.7.

Covalently Bound Ligands. We also searched for covalently bound inhibitors among 3D-cliffs. One 3D-cliff involving covalently bound compounds was found in the cathepsin K inhibitor set and three covalent cliffs were detected in the monoamine oxidase B (MAO-B) inhibitor set. In case of MAO-B, the cliff forming inhibitors were covalently connected to the cofactor FAD. In all four instances, covalent inhibitors in a pair were distinguished by small structural modifications. These four covalent activity cliffs were not included in the classification of 3D-cliffs presented below.

Stereoisomers. It was also determined whether 3D similarity calculations distinguished between structures of stereoisomers that might or might not form 3D-cliffs. In all 187 target sets, a total of 26 ligand pairs from 13 sets were found to consist of stereoisomers. However, only six of these pairs formed 3D-cliffs at the 80% similarity threshold level with at least 2 orders of magnitude potency difference. In addition, 15 pairs yielded high 3D similarity but had only small potency differences. By contrast, three pairs had only limited 3D similarity, visibly different binding modes, and large potency differences. Thus, in these cases, potency differences between stereoisomers could be directly attributed to the presence of different binding modes. The remaining two ligand pairs had low 3D similarity

dx.doi.org/10.1021/ci300158v | J. Chem. Inf. Model. 2012, 52, 1490−1498

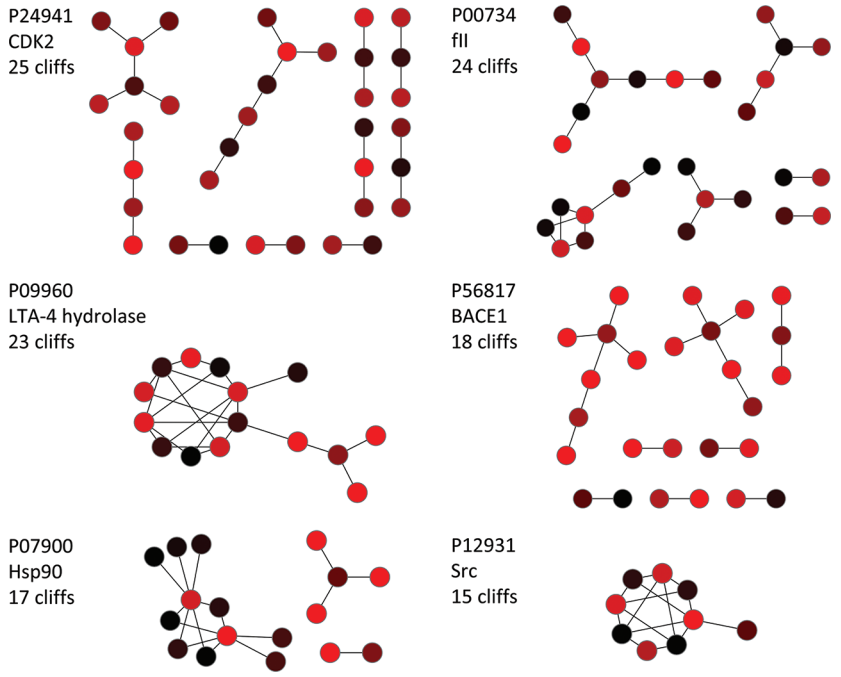


Figure 1. Activity cliff networks. Shown are activity cliff networks for six target sets that contained at least 15 3D-cliffs. In each network, nodes represent ligands that are connected by an edge if they form a 3D-cliff. In addition, nodes are color-coded according to potency values of the ligands applying a continuous color spectrum from black (lowest potency in the set) to red (highest potency). For each set, the UniProtID, target name, and the number of 3D-cliffs are provided. Target abbreviations: CDK2, cyclin-dependent kinase 2; fII, coagulation factor II; LTA-4, leukotriene A4 hydrolase; BACE1, beta-secretase 1; Hsp90, heat shock protein 90; Src, proto-oncogene tyrosine protein kinase Src.

and only small potency differences. Thus, calculated 3D similarity differentiated between the magnitudes of structural effects introduced by changes in stereochemistry. Among the six 3D-cliffs formed by pairs of stereoisomers, two cliffs were identified where stereoisomerization modified ligand–target interactions in a defined manner, despite the presence of high 3D similarity, as further discussed below.

Exchange of Hydrogen to Fluorine Atoms. We also inspected all 216 3D-cliff pairs for the potential addition of fluorine atoms that are similar in size to the hydrogens they replace. Significant potency differences that might potentially result from hydrogen-to-fluorine replacements would be of particular interest for medicinal chemistry. However, no such replacements were detected among the 216 3D-cliffs.

Classification of 3D-Cliffs. Then ligand–target interactions for all cliff forming compound pairs were analyzed in detail and compared. In many instances, we detected differences in individual interactions between ligands forming a 3D-cliff. Importantly, the detection of such interaction differences does not necessarily mean that these interactions are indeed responsible, or largely responsible, for the thermodynamics underlying cliff formation (although this was likely in a number of instances). Clearly, the ligand binding process is energetically much more complex than is reflected by the presence of short-range interactions in ligand–target complexes. For example, the analysis of these types of interactions does not take entropic or desolvation effects into account that might also make large contributions to binding events. However, what can be stated on the basis of interaction analysis is that many of the 3D-cliffs we identified were indeed characterized by well-defined differences in ligand–target interactions that were readily interpretable. This should make these cliffs also interesting for further studies of binding energies or for the calibration of energy functions for drug design.

Distribution of Cliff Categories. We defined six categories of interaction differences to characterize 3D-cliffs. This classification is qualitative in nature. Often, multiple interaction differences were observed and decisions had to be made how to best assign these 3D-cliffs. In such instances, preference was given to most obvious interaction differences. The assignment of cliffs to these categories and their distribution over target sets and families are reported in Table 4. We identified a total of 74

Table 4. Categorization of 3D-Cliffs[a]

category ID	cliff category	number of families	number of target sets	number of 3D-cliffs	number of 3D/2D-cliffs
1	H-bond and/or ionic interactions	9	19	74	11
2	lipophilic/aromatic groups	15	27	89	26
3	water molecules	1	1	1	1
4	stereochemistry	2	2	2	2
5	multiple effects	5	10	32	5
6	no apparent interaction differences	8	10	18	3

[a]For each 3D-cliff category, the numbers of target families, target sets, and 3D- and conserved 3D/2D-cliffs are reported.

3D-cliffs where cliff forming ligands were mostly differentiated by hydrogen bonding and/or ionic interactions. In addition, one cliff was found where water-mediated hydrogen bonds exclusively distinguished ligand interaction patterns and another one where water-mediated hydrogen bonds and the introduction of an additional lipophilic group differentiated two ligands. Moreover, in 89 3D-cliffs, interaction differences

dx.doi.org/10.1021/ci300158v | J. Chem. Inf. Model. 2012, 52, 1490–1498

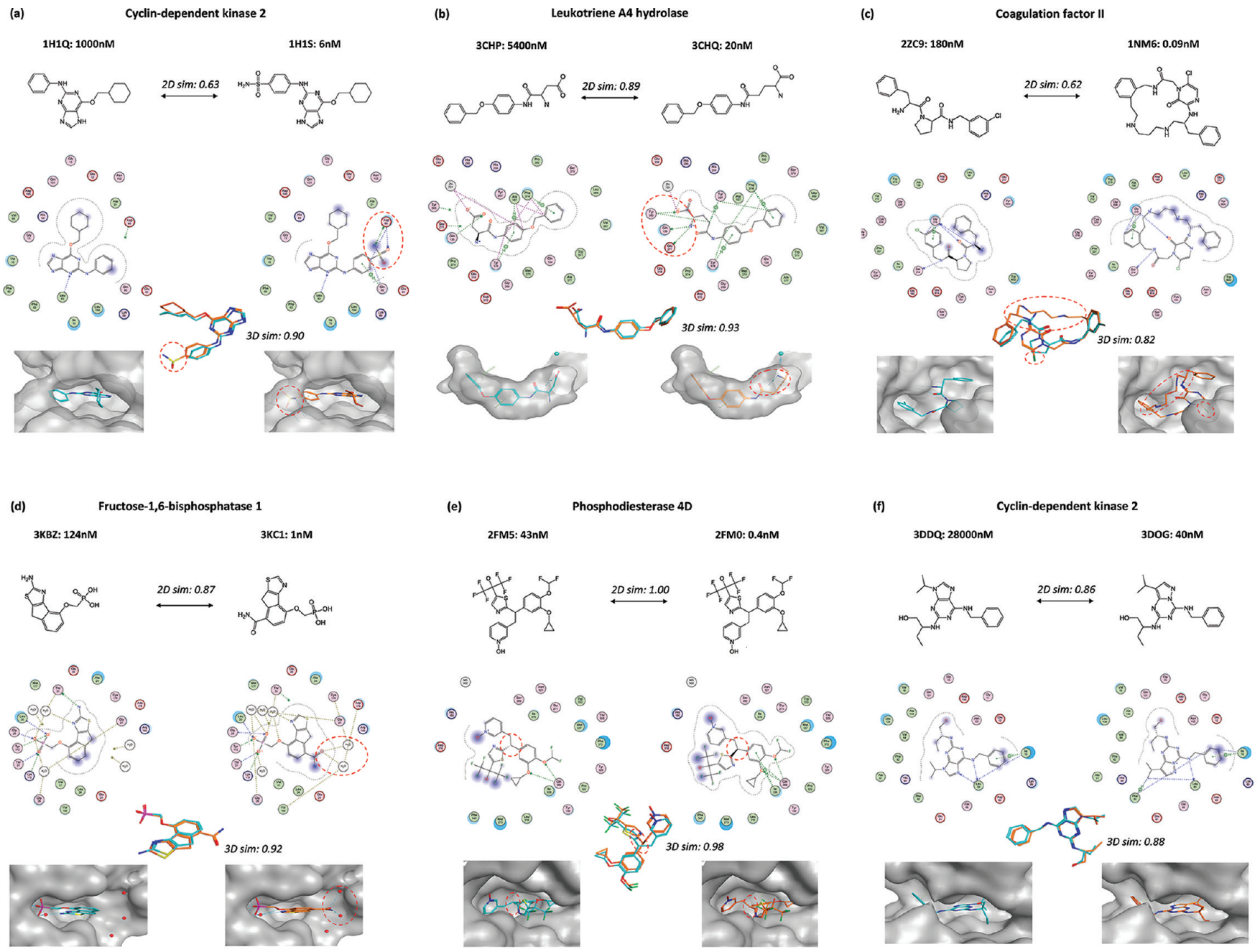


Figure 2. continued

dx.doi.org/10.1021/ci300158v | J. Chem. Inf. Model. 2012, 52, 1490−1498

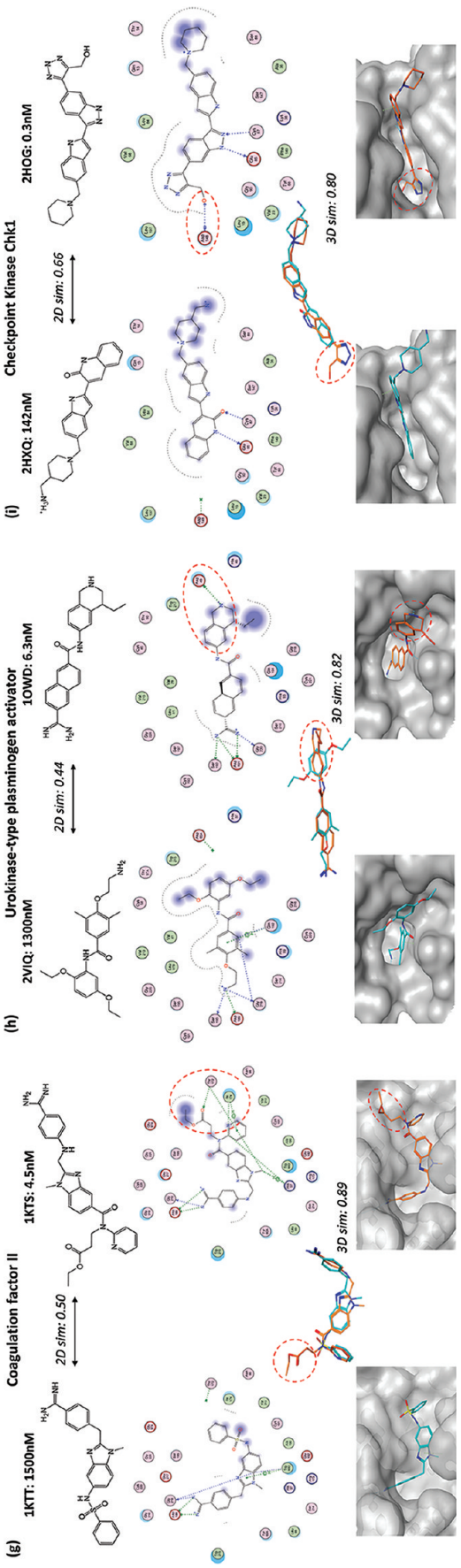


Figure 2. Exemplary 3D-cliffs. In a–i, representative 3D-cliffs are shown that belong to different cliff categories according to Table 4. For each 3D-cliff, the name of the target is provided, 2D structures of cliff-forming ligands are shown, and PDB IDs and potency values are reported. In addition, interaction diagrams, aligned binding modes, and binding site views of ligand-target complexes are shown. Furthermore, 2D and 3D similarity values are reported for comparison of molecular graphs and binding modes, respectively. In each case, the less potent ligand is shown on the left and the more potent one on the right. Regions containing interaction differences discussed in the text are highlighted using red dashed circles in the complementary structural representations.

between cliff partners predominantly involved hydrophobic or aromatic groups, which often led to differences in the shape complementarity between ligands and targets. 3D-Cliffs belonging to the hydrogen bond/ionic interaction or lipophilic/aromatic group category were found in 19 and 27 target sets, respectively, and represented the most prominent cliff categories. In addition, 32 3D-cliffs were identified in which ligands were distinguished by multiple types of clearly defined crystallographic interactions and 2 cliffs where stereoisomerization led to different degrees of shape complementarity in the presence of high binding mode similarity. Finally, we found 18 3D-cliffs in 10 target sets where ligand–target interaction patterns were virtually undistinguishable, despite the presence of large potency differences. Hence, in these cases, different energetic effects were responsible for cliff formation that could not be deduced from the analysis of structural data.

Representative Examples. In Figure 2, representative examples are provided for different cliff categories. In many cases, it was readily possible to predict the highly potent cliff partner on the basis of observed interaction differences. For example, the two cyclin-dependent kinase 2 (CDK2) inhibitors in Figure 2a form a 3D-cliff are only distinguished by charged hydrogen bond interactions between a sulfonamide group in one of the inhibitors and residue Asp 86 of the enzyme. A similar example is shown in Figure 2b for a cliff formed by inhibitors of leukotriene A4 hydrolase where differences in hydrogen bond and ionic interactions are also observed. By contrast, the two cliff forming coagulation factor II inhibitors in Figure 2c display differences in the interactions involving an aliphatic moiety. Here, the larger aliphatic substituent in the highly potent macrocylic compound leads to a further improved fit into a hydrophobic region of the binding site and to a higher degree of shape complementarity. Furthermore, compared to its weakly potent cliff partner, the macrocyclic compound is also conformationally restricted, which provides an entropy advantage. In addition, the two fructose-1,6-bisphosphatase inhibitors in Figure 2d provide an example of a 3D-cliff where the ligands are only distinguished by hydrogen bonds involving crystallographic water molecule. In the X-ray structure, there is no hydrogen bond formed between the more potent ligand and residue Glu 29. In Figure 2e, a 3D-cliff is shown involving two phosphodiesterase 4D inhibitors that are stereoisomers. In this case, different stereochemistry at a single stereocenter leads to a slightly more constrained fit of the less potent ligand, i.e. the distomer, into the binding site. In Figure 2f, another 3D-cliff involving two CDK2 inhibitors is shown that differ by positional change of one nitrogen ring atom. Here, only very subtle interaction differences are observed, although these compounds have a large difference in potency. In Figure 2g, the replacement of a substituent in two coagulation factor II inhibitors leads to an improved fit of the more potent compound into the binding site that also forms an additional hydrogen bond compared to the less potent inhibitor. Hence, this is an example of a 3D-cliff characterized by multiple interaction differences. In Figure 2h, an example of a urokinase-type plasminogen activator inhibitor cliff is shown where chemically different ligands with low 2D similarity form very similar binding modes, with multiple conserved as well as differentiating interactions. For example, the positively charged nitrogen in the condensed ring of the highly potent inhibitor forms a strong salt bridge interaction with an aspartic acid in the binding site that is absent in the weakly potent cliff partner. In addition, the presence of the condensed ring also leads to a

better fit into the binding pocket than the two ethyl phenyl ether groups in the weakly potent inhibitor. In addition, there are other less prominent interaction differences. This clearly defined 3D-cliff would have not been identified as a 2D-cliff, given the limited 2D similarity of these inhibitors. Finally, in Figure 2i, another example of a 3D-cliff for kinase inhibitors is presented where multiple interaction differences are observed. The differences include the formation of a well-defined hydrogen bond and the fit of an additional ring into a pocket of the enzyme. Both of these interactions are formed by the highly potent cliff partner.

CONCLUDING REMARKS

In this study, we have systematically identified and characterized three-dimensional activity cliffs available in the Protein Data Bank. Considerations were limited to ligand-target complexes for which high-confidence structural and activity data were available and that could be assigned to defined target sets. Given these requirements, only approximately 20% of available ligand-target complex structures were ultimately subjected to activity cliff analysis. We detected 3D-cliffs at different similarity and potency threshold levels that were often not conserved when 2D similarity measures were applied. These systematically identified activity cliffs represent all 3D-cliffs that are currently available in the public domain (when rigorous criteria for cliff formation are applied). Hence, this pool of 3D cliffs should provide a relevant knowledge base for further rationalization of activity cliff formation beyond 2D similarity considerations. Perhaps it might also be possible to utilize this information for compound design directed against selected targets, although this has not been a primary purpose of our analysis. We finally selected a set of more than 200 3D-cliffs that were characterized by the presence of at least 80% 3D similarity and an at least 100-fold difference in potency between the participating ligands. In many but not all instances, the cliff forming compounds displayed well-defined differences in crystallographic ligand−target interactions. As discussed, determining to what extent such structural features give rise to large potency differences responsible for cliff formation is the subject of subsequent analyses. Thus, these 3D-cliffs might also be of interest beyond activity cliff analysis, for example, to study binding energetics or calibrate energy functions in the context of structure-based drug design.

ASSOCIATED CONTENT

Supporting Information

Supplementary Table S1 reports PDB and activity source information for 216 3D-cliffs and Supplementary Figures S1 and S2 provide PDB entry and measurement statistics and 2D similarity values for 3D-cliffs, 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 two authors should be considered equal.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

N.F. is supported by a fellowship from the Jürgen Manchot Foundation, Düsseldorf, Germany.

REFERENCES

(1) Maggiora, G. M. On Outliers and Activity Cliffs − Why QSAR often Disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535−1535.
(2) Vogt, M.; Huang, Y.; Bajorath, J. From Activity Cliffs to Activity Ridges: Informative Data Structures for SAR Analysis. *J. Chem. Inf. Model.* **2011**, *51*, 1848−1856.
(3) Stumpfe, D.; Bajorath, J. Exploring Activity Cliffs in Medicinal Chemistry. *J. Med. Chem.* **2012**, *55*, 2932−2942.
(4) 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.
(5) Wassermann, A. M.; Dimova, D.; Bajorath, J. Comprehensive Analysis of Single- and Multi-Target Activity Cliffs Formed by Currently Available Bioactive Compounds. *Chem. Biol. Drug Des.* **2011**, *78*, 224−228.
(6) Wassermann, A. M.; Wawer, M.; Bajorath, J. Activity Landscape Representations for Structure-Activity Relationship Analysis. *J. Med. Chem.* **2010**, *53*, 8209−8223.
(7) Medina-Franco, J. L.; Martínez-Mayorga, K.; Bender, A.; Marín, R. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. Characterization of Activity Landscapes using 2D and 3D Similarity Methods: Consensus Activity Cliffs. *J. Chem. Inf. Model.* **2009**, *49*, 477−491.
(8) Sisay, M. T.; Peltason, L.; Bajorath, J. Structural Interpretation of Activity Cliffs Revealed by Systematic Analysis of Structure-Activity Relationships in Analog Series. *J. Chem. Inf. Model.* **2009**, *49*, 2179−2189.
(9) Seebeck, B.; Wagener, M.; Rarey, M. From Activity Cliffs to Target-specific Scoring Models and Pharmacophore Hypotheses. *ChemMedChem* **2011**, *6*, 1630−1639.
(10) Hu, Y.; Bajorath, J. Exploration of 3D Activity Cliffs on the Basis of Compound Binding Modes and Comparison of 2D and 3D-cliffs. *J. Chem. Inf. Model.* **2012**, *52*, 670−677.
(11) Peltason, L.; Bajorath, J. Molecular Similarity Analysis Uncovers the Presence of Heterogeneous Structure-Activity Relationships and Variable Activity Landscapes within Active Sites. *Chem. Biol.* **2007**, *14*, 489−497.
(12) Berman, H.; Henrick, K.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(13) Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L. The Worldwide Protein Data Bank (wwPDB). *Nucleic Acids Res.* **2007**, *35*, D301−D303.
(14) *RCSB Protein Data Bank*. http://www.rcsb.org (accessed March 1, 2012).
(15) UniProtConsortium. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.* **2010**, *38*, D142−D148.
(16) 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.
(17) *ChEMBL*. http://www.ebi.ac.uk/chembldb/ (accessed March 1, 2012).
(18) 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.
(19) *BindingDB*. http://www.bindingdb.org/ (accessed March 1, 2012).
(20) Willett, P. Searching Techniques for Databases of Two- and Three-Dimensional Structures. *J. Med. Chem.* **2005**, *48*, 4183−4199.
(21) *MACCS Structural Keys*; Symyx Software: San Ramon, CA, 2005.
(22) *MOE (Molecular Operating Environment)*; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2011.

(23) Clark, A. M.; Labute, P. 2D Depiction of Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2007**, *47*, 1933−1944.
(24) 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.