JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

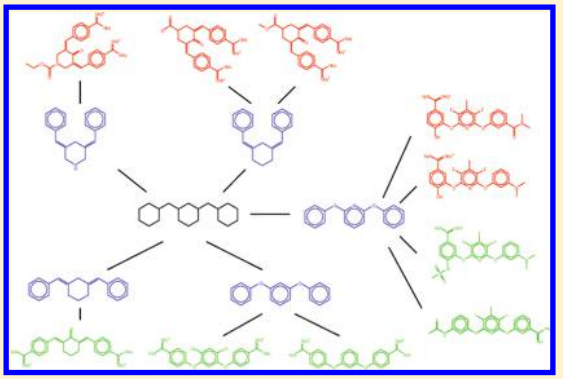
From Activity Cliffs to Activity Ridges: Informative Data Structures for SAR Analysis

Martin Vogt,[†] Yun Huang,[†] and Jürgen Bajorath*

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

Supporting Information

ABSTRACT: The extraction of SAR information from structurally diverse compound data sets is a challenging task. One of the focal points of systematic SAR analysis is the search for activity cliffs, that is, structurally similar compounds having large potency differences, from which SAR determinants can be deduced. The assessment of SAR information is usually based on pairwise similarity and potency comparisons of data set compounds. As a consequence, activity cliffs are mostly evaluated at a compound pair level. Here, we present an extension of the activity cliff concept by introducing "activity ridges" that are formed by overlapping "combinatorial" activity cliffs between participating compounds, giving rise to ridge-like structures in activity landscapes. Activity ridges are rich in SAR information. In a systematic analysis of 242 compound data sets, we have identified well-defined activity ridges in 71 different sets. In addition, an information-theoretic approach has been devised to characterize the structural composition of activity ridges. Taken together, our results show that activity ridges frequently occur in sets of active compounds and that different categories of ridges can be distinguished on the basis of their structural content. The computational identification of activity ridges provides access to compound subsets having high priority for SAR analysis.



INTRODUCTION

Computational approaches for SAR analysis of large and heterogeneous compound data sets have received increasing attention in recent years.[1−3] These data sets originate from biological screening, hit-to-lead, and/or lead optimization efforts and have often accumulated over time in the course of drug discovery projects. When studying such compound sets in medicinal chemistry, a primary goal is SAR information extraction. Within this process, the identification of activity cliffs[3,4] typically is one of the initial focal points. Activity cliffs are formed by pairs of structurally similar (analogous) compounds where one compound is weakly and the other highly potent.[4] Activity cliffs are particularly interesting for SAR analysis because they encode small structural changes leading to large differences in potency.

It should be noted that activity cliffs result from different interactions between structurally similar compounds and their target sites. Thus, interactions leading to activity cliff formation can often (but not always) be rationalized in three dimensions (provided the structural information is available). However, critical interactions are implicitly encoded in structure−activity relationship information provided by molecular graphs and associated compound potency information (which also captures other influences such as (de)solvation effects that cannot be directly inferred from 3D interactions). Thus, activity cliffs are usually studied on the basis of ligand information and often using molecular graphs as structural input.

Regardless of their representation, large-magnitude activity cliffs form the highest level of SAR discontinuity in compound data sets[1,2] and are associated with high SAR information content.[3−5] Different computational approaches have been introduced to systematically screen compound data sets for discontinuous local SARs[6,7] and activity cliffs.[7,8] Irrespective of their specific features, these methods generally have in common that they rely on pairwise similarity and potency comparisons of test compounds. Consequently, activity cliffs are usually considered on a compound pair basis.[2]

Herein, we introduce a data structure termed "activity ridge". According to our definition, activity ridges consist of multiple overlapping activity cliffs that are formed by series of compounds where each compound participates in cliff formation. Accordingly, an activity ridge contains more SAR information than isolated cliffs, and compound subsets forming activity ridges would thus be of high relevance for SAR analysis.

Although an activity ridge can be easily conceptualized, it represents a previously unconsidered and as of yet hypothetical data structure. Therefore, we have systematically investigated whether such activity ridges are found in actual compound data sets. To these ends, we have searched for activity ridges in 242 compound activity classes and detected more than 100 well-defined ridges in

Received: June 3, 2011
Published: July 17, 2011

a total of 71 compound sets. Thus, activity ridges are indeed rather frequently observed and provide another attractive starting point for SAR analysis. Furthermore, we have analyzed the composition of activity ridges in detail and distinguished between different types of ridges on the basis of their structural content.

MATERIALS AND METHODS

Similarity Assessment. For our analysis of activity ridges, we have concentrated on a core structure-based assessment of compound similarity, rather than calculated similarity values, which have thus far been mostly utilized in the context of activity cliff and activity landscape analysis.[3] Importantly, calculation of compound similarity requires the use of molecular representations such as fingerprints or numerical property descriptors and the application of similarity metrics. It is well appreciated that activity cliff formation and activity landscape topologies much depend on the choice of molecular representations and similarity measures,[9,10] which introduces a bias in SAR analysis and often renders the results representation-dependent. Therefore, we have applied a more robust and chemically intuitive structure-based measure of similarity relationships that utilizes shared core structures (vide infra) as a similarity criterion. The application of this criterion alleviates the need for pairwise structural comparison and is particularly suitable for the organization of analogue series.

Data Sets. From BindingDB,[11] 242 different activity classes covering a wide spectrum of target proteins were extracted. As selection criteria, we required that each class consisted of least 50 compounds with ring-containing core structures (i.e., linear aliphatic molecules were omitted) and that pK_i values were available for all compounds. The average number of compounds per activity class was ~344.

Core Structures. From compounds of each activity class, scaffolds were extracted following the definition of Bemis and Murcko[12] (BM scaffolds) that consist of all rings and linker fragments between rings (after removal of R-groups). BM scaffolds were further reduced to carbon skeletons (CSKs), also referred to as cyclic skeletons,[13] by setting all bond orders to one and converting all heteroatoms to carbon. CSKs were utilized as core structures for compound classification (similarity criterion).

Compounds from each activity class were divided into CSK-based subsets, and for each subset the presence of activity ridges was investigated.

Activity Ridge Definition. We define an activity ridge on the basis of the following five criteria:

(1) Activity cliff threshold: For a cliff to be considered a part of a ridge, the difference in potency between a weakly and a highly potent compound must be at least 2 orders of magnitude (i.e., two pK_i units).

(2) Upper boundary potency threshold: The potency (K_i) of a compound considered to be highly potent must at least be 100 nM.

(3) Lower and upper boundary potency variance: The potencies of all weakly and all highly potent compounds must be within 1 order of magnitude.
This means that if the most potent compound in a ridge has a potency of 1 nM, high potency values can range from 1 to 10 nM. In addition, if the least potent compound has a potency of 10 μM, low potency values can range from 10 to 1 μM.

(4) Compound number threshold: A minimum of five weakly and five highly potent compounds are required for ridge formation.

(5) Subset optimization: Subsets of weakly and highly potent compounds are selected such that the largest possible ridges are formed.
Among sets of similar compounds sharing the same CSK, overlapping activity ridges might be formed, and ridge formation might in some instances thus be ambiguous. Therefore, a two-step procedure is applied. In the first step, the weakly and highly potent compound subsets (with at least 2 orders of magnitude difference in potency) are determined such that the product of the number of weakly and highly potent compounds is maximal. In the second step, weakly and highly potent compounds falling within 1 order of magnitude in potency, respectively, are selected such that the largest possible selection is obtained. If several selection sets of equal size are identified, the one with the largest potency margin between the weakly and highly potent compounds is chosen. On the basis of this procedure, larger ridges are prioritized, and potential ambiguities in ridge formation are avoided.

The search procedure for the identification of activity ridges was implemented in Java using the OpenEye chemistry toolkit.[14]

Mutual Information-Based Scaffold Analysis. We quantitatively analyzed whether certain BM scaffolds captured by a CSK would predominantly yield highly or weakly potent compounds. From an information-theoretic point of view, this question can be reformulated as follows: does a given BM scaffold contain any information concerning the potency of the analogues it represents? This type of information is termed Mutual Information (MI)[15] between an object (scaffold) and two property states (i.e., a highly potent and a weakly potent ridge compound) that can be quantitatively accounted for.

The definition of MI is based on the concept of the entropy of a random variable. For instance, the property P of a ridge compound (weakly potent or highly potent) can be represented as a binary random variable with the two possible values ("weak" or "high"). The probability of a randomly chosen ridge compound belonging to the weak or high category is given by the relative frequency of weakly and highly potent compounds.

If n compounds form a ridge including w weakly and h highly potent ones, this probability becomes: Pr("weak") = w/n and Pr("high") = h/n.

The entropy $H(P)$ then is:

$$H(P) = -\frac{w}{n}\log_2\frac{w}{n} - \frac{h}{n}\log_2\frac{h}{n}$$

This equation accounts for the information content of P and is measured in bits. If the numbers of weakly and highly potent compounds are the same, $H(P) = 1$ bit. If these numbers are not the same, $H(P) < 1$ bit.

Analogously, one can consider the entropy of the BM scaffold distribution. Here, the random variable is B (BM scaffold of a ridge compound), which can adopt any of m values with probability b_i, $i = 1,...,m$, if m different scaffolds occur in ridge compounds with relative frequencies b_i. The entropy of B then is:[15]

$$H(B) = -\sum_{i=1}^{m} b_i \log_2 b_i$$

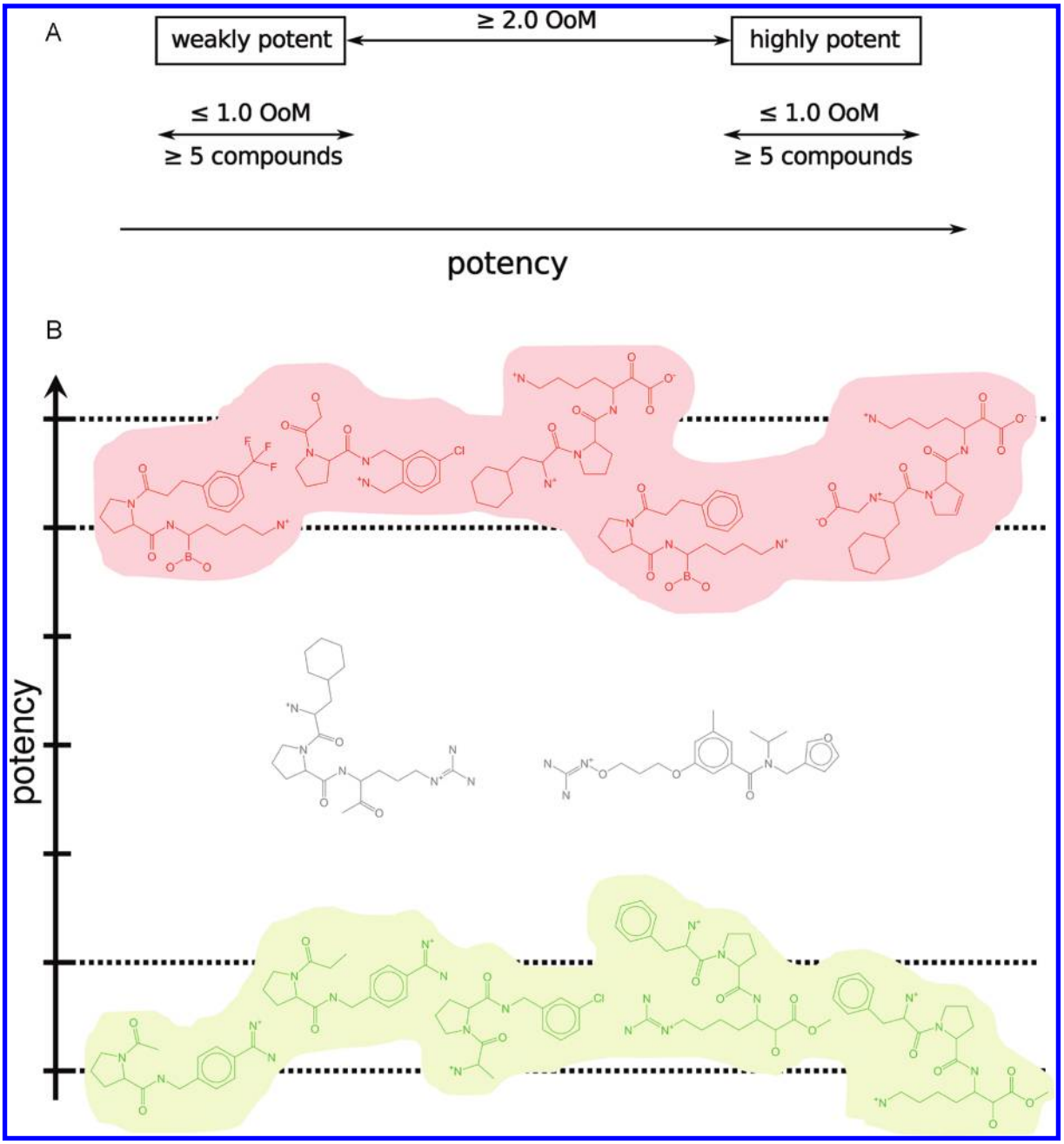


Figure 1. Activity ridge model structure. In (A), a schematic illustration is shown that describes the concept of an activity ridge. OoM means orders of magnitude. Part (B) shows a set of thrombin inhibitors (green, weakly potent; red, highly potent) forming an activity ridge. Gray inhibitors are not part of the ridge.

The MI of the random variables P and B then tells us how much information about P (the weakly potent and highly potent states) is accounted for by the scaffold B. This value depends on the distribution of scaffolds among the weakly and the highly potent ridge compounds. If there is no difference in the distribution of scaffolds between weakly and highly potent compounds, then there also is no scaffold-specific information (i.e., MI is zero). If on the other hand the two distributions are disjoint, that is, the weakly and highly potent compounds do not share any scaffold, knowledge of the BM scaffold completely determines the property distribution P and MI is maximal. These relationships are formalized as follows. The distribution of scaffolds among active compounds is given by the conditional distributions $\Pr(B|P = \text{“weak”})$ and $\Pr(B|P = \text{“high”})$.

MI is then quantified as the difference between the entropy of B and the entropy of the conditional distributions:

$$\mathrm{MI}(B, P) = H(B) - H(B|P)$$

where

$$H(B|P) = \Pr(P = \text{“weak”})H(B|P = \text{“weak”}) + \Pr(P = \text{“high”})H(B|P = \text{“high”})$$

dx.doi.org/10.1021/ci2002473 |J. Chem. Inf. Model. 2011, 51, 1848–1856

Table 1. Summary of the Search Results

number of compound data sets	242
total number of active compounds	83 234
total number of nonsingleton CSK-based compound subsets	9895
number of CSK-based subsets containing activity cliffs	2315
compound data sets containing activity cliffs	233
number of isolated activity cliffs	1514
compound data sets containing activity ridges	71
number of activtiy ridges	125

If the relative frequencies of the scaffolds are w_i for weakly potent and h_i for highly potent compounds, the conditional entropies are calculated as:

$$H(B|P = \text{“weak”}) = -\sum_{i=1}^{m} w_i \log_2 w_i$$

$$H(B|P = \text{“high”}) = -\sum_{i=1}^{m} h_i \log_2 h_i$$

It follows that $MI(B,P)$ can maximally be $H(P)$. Thus, MI can be normalized to $MI'(B,P) = MI(B,P)/H(P)$, which yields values within the range 0−1.

The MI analysis procedure was implemented in Java.

RESULTS AND DISCUSSION

Concept of an Activity Ridge. The intuitive notion of an activity ridge is that of a set of structurally similar compounds that includes both weakly and highly potent ones, leading to the formation of multiple overlapping activity cliffs (i.e., cliffs sharing a weakly or highly potent compound). Figure 1 provides a schematic representation of an activity ridge and a compound example. If we consider an activity ridge from an activity landscape perspective, the ridge formally consists of a base surface (lower boundary) formed by weakly potent compounds and an elevated surface (upper boundary) formed by highly potent compounds. On the basis of criteria 1−5 applied to define an activity ridge (detailed in the Materials and Methods), qualifying ridges share a number of architectural features but can significantly differ in the magnitude and distribution of activity cliffs they contain. The distance between the base surface and the elevated surface defines the minimal magnitude of individual activity cliffs required for ridge formation. As indicated in Figure 1, other similar compounds might have potency values falling between the lower and upper boundaries of the ridge but are not considered because they cannot consistently form activity cliffs of the minimally required magnitude to ridge compounds. Importantly, each compound belonging to a ridge participates in the formation of activity cliffs. As a consequence of criterion 3 (lower and upper boundary potency variance), a ridge compound participates in the formation of activity cliffs with all of its (either weakly or highly potent) counterparts. Therefore, an activity ridge is characterized by the presence of combinatorial activity cliffs formed between participating compounds (i.e., all possible activity cliffs).

Considering the difficulties in numerically quantifying molecular similarity in a consistent manner, a structural criterion is applied as a similarity measure: compounds are regarded as similar if they belong to the same topological structure class (i.e., if they share the same CSK). This criterion captures all analogue series including those with heteroatom substitutions in rings or linker fragments between rings. For each CSK-based compound subset, the BM scaffold composition of weakly and highly potent ridge compounds is then quantitatively assessed and compared utilizing an information-theoretic approach, as further discussed below.

Table 2. Activity Classes Containing at Least Three Activity Ridges[a]

activity class/target	number of ridges
factor Xa	9
adenosine receptor A3	8
thrombin	6
cannabinoid receptor 2	5
melanocortin receptor 4	4
sigma opioid receptor	4
dopamine D4 receptor	3
kappa opioid receptor	3
melanin-concentrating hormone receptor 1 (MCH1R)	3
carbonic anhydrase I	3
cannabinoid receptor 1	3

[a] The activity classes consist of enzyme inhibitors or receptor antagonists.

Identification of Activity Ridges. We searched 242 activity classes for activity ridges meeting criteria 1−5. The results are summarized in Table 1. The activity classes contained a total of 83 234 compounds that yielded 23 030 different CSK-based compound sets including 13 135 singletons, defined here as single compounds yielding a unique CSK (and thus also a unique BM scaffold). Only sets containing at least 10 compounds could be considered for our analysis. Activity cliffs with a potency difference of at least 2 orders of magnitude involving at least one nanomolar compound (criterion 2) were found in 2315 CSK-based sets originating from 233 activity classes (hence, nine of the 242 classes did not contain activity cliffs of the required magnitude). In these 2315 sets, we detected a total of 1514 isolated activity cliffs (formed between individual compound pairs). Moreover, we detected 125 well-defined activity ridges that occurred in a total of 71 activity classes. The majority of these classes contained a single ridge, but 11 classes were found to contain three or more activity ridges, as reported in Table 2. These classes predominantly consisted of G protein coupled receptor antagonists. However, the three activity classes containing most activity ridges included two classes of serine protease inhibitors: factor Xa inhibitors (nine ridges) followed by adenosine receptor A3 antagonists (eight) and thrombin inhibitors (five).

Figure 2 shows two exemplary activity ridges from the factor Xa data set having different size and composition. In Figure S1 of the Supporting Information, additional examples are provided from the factor Xa and other compound sets. In Figure 2A, the subset-defining CSK covers a single BM scaffold that represents six weakly and five highly potent analogues. By contrast, the CSK in Figure 2B covers five different BM scaffolds: two representing either one or two weakly potent compounds, two others representing one or two highly potent compounds, and a fifth scaffold representing 27 weakly potent and four highly potent analogues. These examples already illustrate differences in the nature and composition of the identified activity ridges, as described below in detail.

Size and Potency Distribution. Next, we determined the size of the activity ridges and the ratio of weakly and highly potent

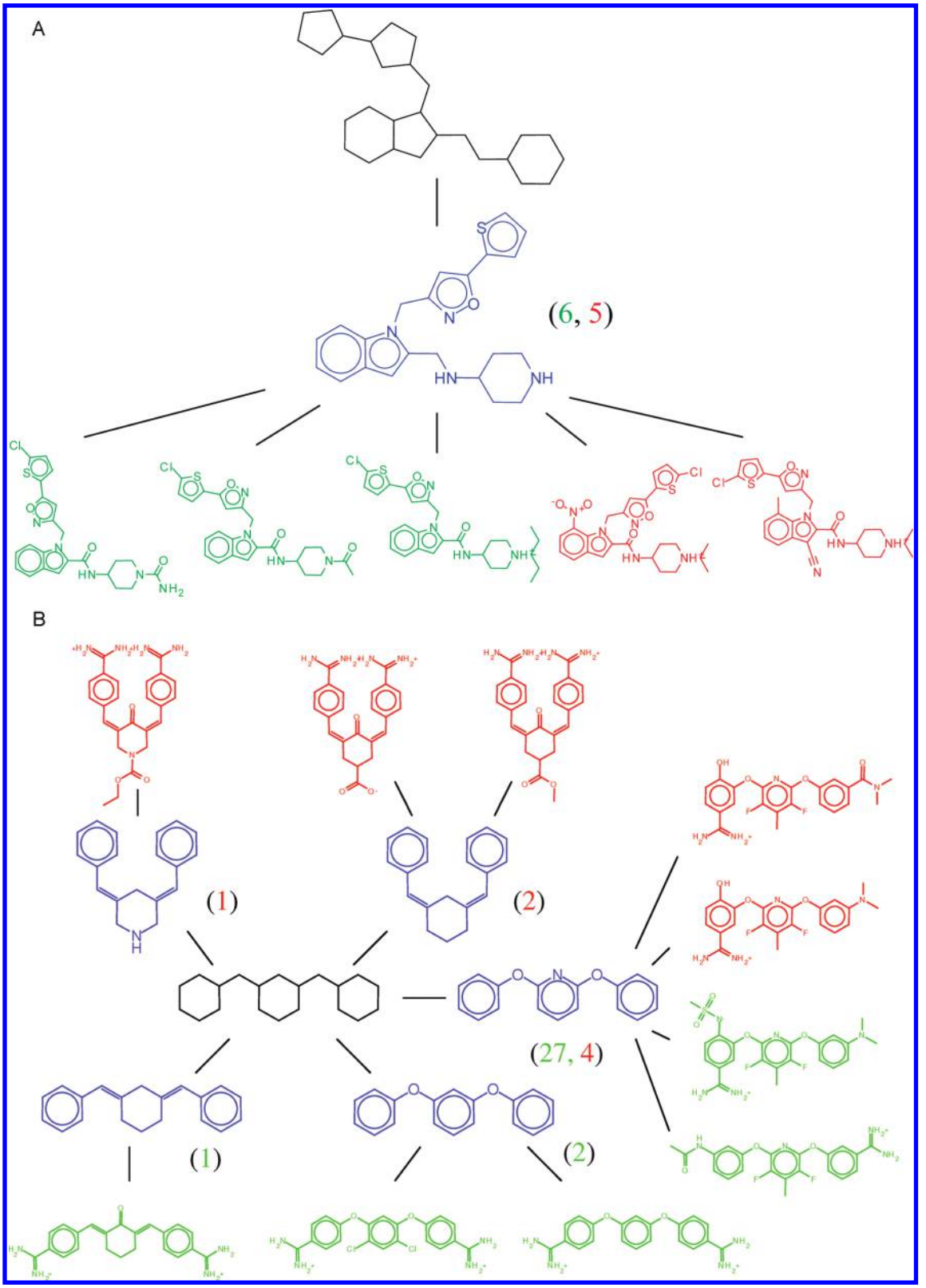


Figure 2. Activity ridges formed by factor Xa inhibitors. In (A) and (B), inhibitor subsets forming two activity ridges are displayed that involve a single or multiple BM scaffolds, respectively. Carbon skeletons, BM scaffolds, weakly potent compounds, and highly potent compounds are shown in black, blue, green, and red, respectively. Scaffolds are labeled with colored numbers or colored number pairs that report the number of weakly potent (green) and/or highly potent (red) compounds represented by each scaffold. Black lines between structures indicate CSK−scaffold−compound relationships. For clarity, not all weakly and highly potent compounds are displayed.

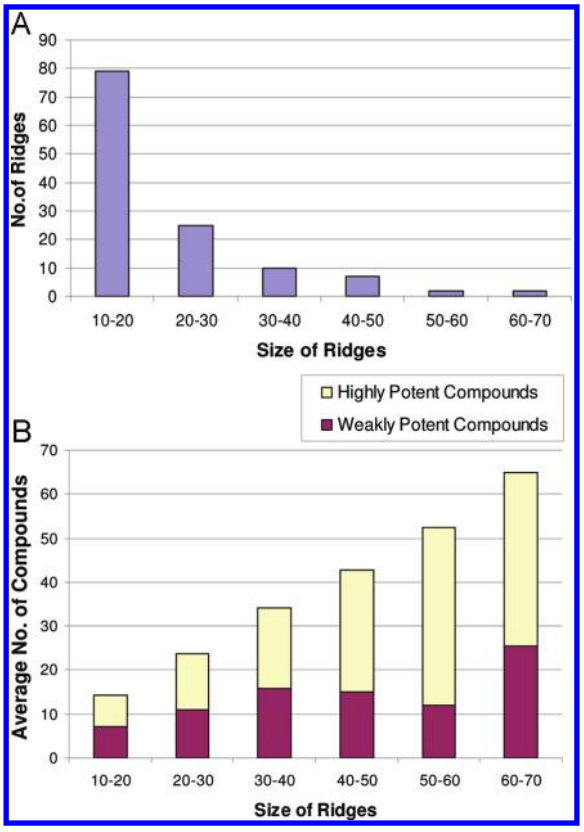


Figure 3. Activity ridge size and compound distribution. In (A), the size distribution of activity ridges is shown. Size refers to the number of ridge compounds. In (B), average numbers of weakly and highly potent compounds are reported for ridges of increasing size.

compounds participating in their formation. According to criterion 4, the minimal size of a ridge was 10 compounds including at least five weakly and five highly potent ones. Figure 3A shows that the majority of ridges that were found contained 10−20 compounds. However, much larger ridges were also detected. In total, 25 ridges consisted of 20−30 compounds, 10 of 30−40, seven of 40−50, two of 50−60, and two of 60−70. The largest ridge contained 69 inhibitors of protein kinase Cα. Figure 3B reports the weakly versus highly potent compound composition of ridges falling into different size intervals. The small ridges containing 10−20 compounds displayed a balanced composition of weakly and highly potent compounds (by definition, ridges with 10 compounds must consist of equal numbers of weakly and highly potent compounds). However, with increasing ridge size, the ratio of highly over weakly potent compounds increased, indicating that larger ridges predominantly originated from advanced lead optimization efforts.

Activity Ridge Composition. The exemplary activity ridges in Figure 2 indicated that the composition of the detected ridges was variable and characterized by different relationships between a subset-defining CSK, the corresponding BM scaffolds, and active compounds represented by the scaffolds. To quantitatively assess the relative distributions of BM scaffolds and the weakly or highly potent ridge compounds they represented, we adapted the MI concept. MI analysis provided the basis to determine how much potency information was encoded in individual BM scaffolds present in activity ridges of different composition.

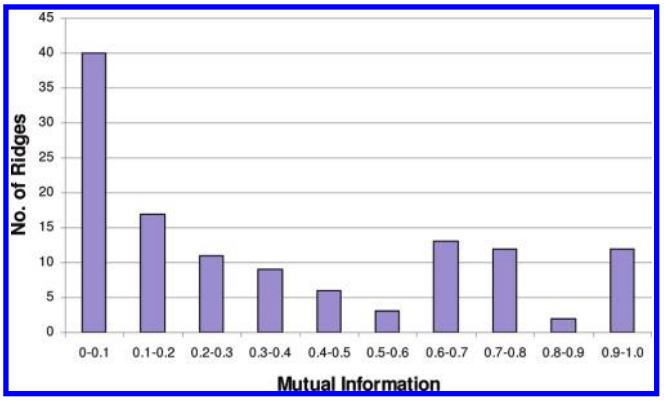


Figure 4. Mutual information analysis of activity ridges. The numbers of activity ridges falling into 10 MI value intervals are reported.

Mutual Information Analysis. Figure 4 summarizes the results of MI analysis for all 125 activity ridges. Lowest MI values close to 0 are indicative of ridges where participating scaffolds represent both weakly and highly potent compounds and where no firm statistical trends for scaffolds with respect to compound potency states are detectable. By contrast, high MI values close to 1 are characteristic of ridges where scaffolds exclusively, or almost exclusively, represent compounds with either low or high potency. Figure 4 shows that activity ridges were found to cover the entire MI value range, which further emphasized the variability of ridge composition. However, the majority of ridges displayed values in the [0, 0.1) MI interval. Hence, in these cases, single or multiple scaffolds yielded both weakly and highly potent analogues. Moreover, it is also observed in Figure 4 that the number of ridges gradually decreased toward midrange MI values and then increased again for higher MI values (with the exception of ridges falling into the [0.8, 0.9) interval). In particular, 12 ridges were found to yield MI values of 1, which indicated strong potency state preferences of individual scaffolds, as discussed above. Figure 5 shows a representative example of low-MI (Figure 5A) and two representative examples of high-MI (Figure 5B and C) activity ridges that illustrate substantial differences in activity ridge composition.

Scaffold and Compound Distribution. On the basis of MI analysis, we further characterized the BM scaffold and compound distribution over activity ridges. For each ridge, we determined the number of weakly and highly potent compounds that originated from scaffolds yielding only weakly potent, only highly potent, or both weakly and highly potent compounds, respectively. Consistent with the results of MI analysis, we found that activity ridges could be divided into three different categories depending on specific scaffold−compound potency relationships. Representative results are shown in Figure 6 that reports the compound composition reported for ridges arranged in the order of increasing MI values. At low MI values, activity ridges consisted of compounds represented by scaffolds that yielded both weakly and highly potent compounds. At midrange MI values, the composition of ridges was heterogeneous. Here, ridges contained compounds that were produced by scaffolds with exclusive low or high compound potency as well as scaffolds associated with low and high compound potency. At high MI values, ridges consisted of compounds represented by scaffolds that exclusively yielded weakly or highly potent compounds. Thus, the quantitative estimate of activity ridge composition based on information theory strongly correlated with observed

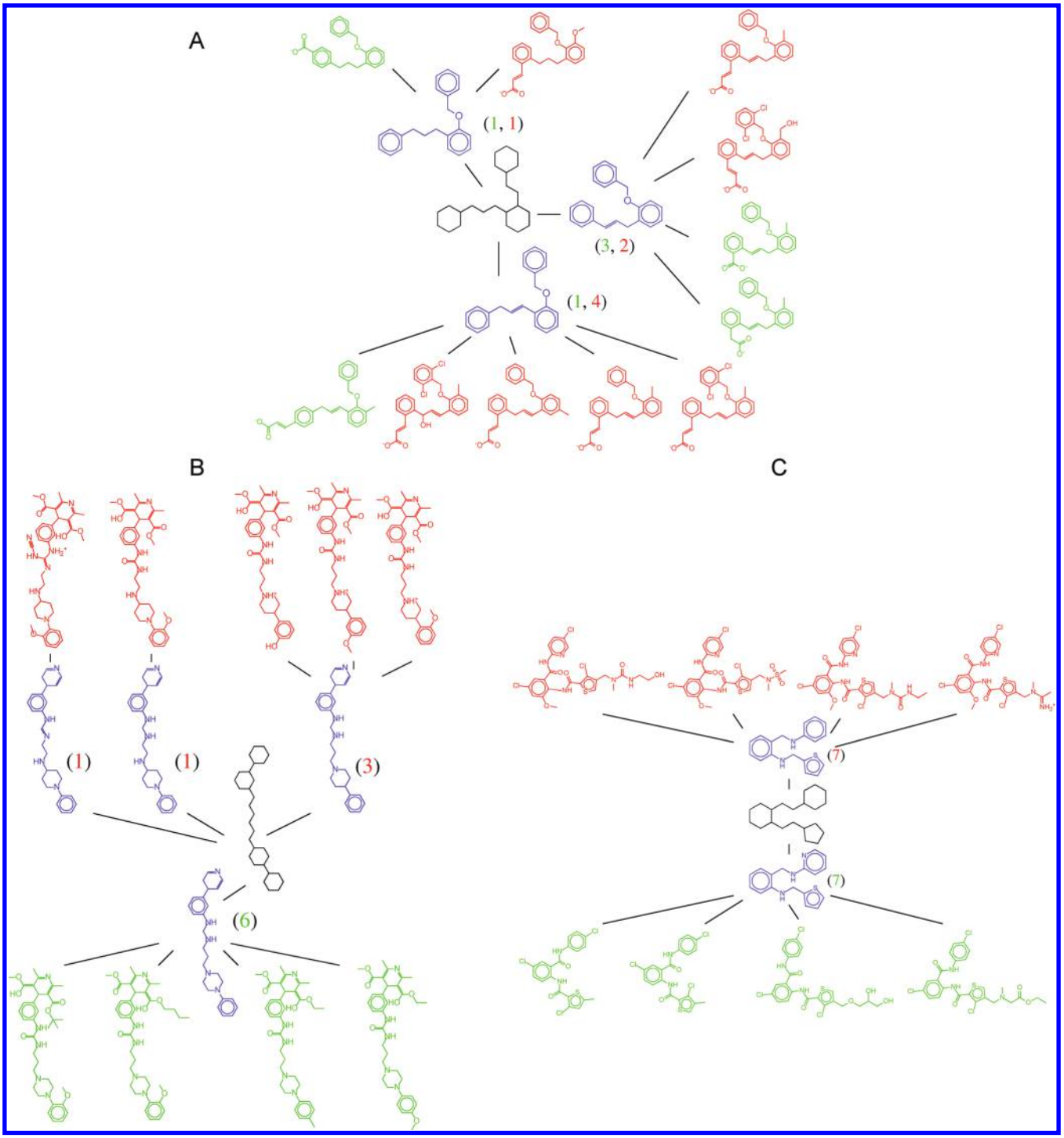


Figure 5. Activity ridges representing different levels of mutual information. Examples of activity ridges with different MI values are shown: (A) prostanoid EP3 receptor antagonists, low MI; (B) neuropeptide Y receptor type-1 antagonists, high MI; (C) factor Xa inhibitors, high MI. The representation is according to Figure 2.

scaffold−compound potency relationships that characterized different types of activity ridges. In Table 3, examples of ridges from different activity classes are reported that belong to the different categories observed at low, intermediate, and high MI values.

Mutual Information and SAR Analysis. MI values of activity ridges provide an immediate measure of their composition and also indicate what type of SAR information can be obtained from them, which is an important prerequisite for the selection of activity ridges for SAR analysis. Low-MI activity ridges are characterized by the presence of single or multiple compound series (each of which is centered on an individual BM scaffold) that contain both weakly and highly potent analogues. Hence, these types of ridges carry much information relevant for conventional SAR analysis focusing on individual compound series with large potency variations. However, high-MI ridges are equally relevant for SAR analysis because they reveal preferences

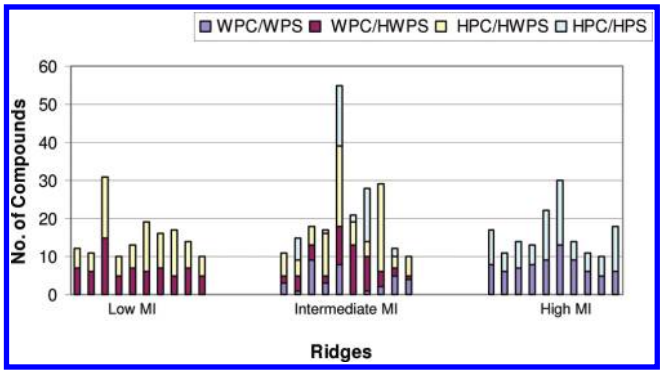


Figure 6. Composition of activity ridges with increasing mutual information. On the horizontal axis, activity ridges are arranged in the order of increasing MI. On the vertical axis, numbers of compounds represented by three categories of BM scaffolds are reported for each ridge. Four different compound/scaffold combinations are distinguished: weakly potent compounds (WPC) represented by scaffolds exclusively yielding weakly potent compounds (WPS, scaffold category 1), WPC represented by scaffolds yielding both highly and weakly potent compounds (HWPS, category 2), highly potent compounds (HPC) represented by HWPS, and HPC represented by scaffolds exclusively yielding highly potent compounds (HPS, category 3). For clarity, representative activity ridge subsets of different composition are shown that define different types of ridges at low, midrange, and high MI values.

Table 3. Selected Activity Ridges with Varying Mutual Information[a]

target	WPS	HPS	WHPS	% WHPS	MI
		Low MI			
factor Xa	0	0	1	100	0.00
carbonic anhydrase I	0	0	2	100	0.00
sigma opioid receptor	0	0	2	100	0.00
carbonic anhydrase XII	0	1	1	88	0.10
thrombin	0	4	3	81	0.20
		Intermediate MI			
sigma opioid receptor	1	3	1	73	0.30
adenosine receptor A3	5	0	2	87	0.40
cannabinoid receptor 2	0	1	2	90	0.55
dopamine D4 receptor	4	2	2	42	0.60
carbonic anhydrase I	5	0	1	77	0.70
		High MI			
carbonic anhydrase I	3	4	1	22	0.80
factor Xa	1	1	0	0	1.00
thrombin	2	3	0	0	1.00
sigma opioid receptor	1	3	0	0	1.00

[a] WPS reports the number of BM scaffolds occurring exclusively in the weakly potent ridge compounds, HPS the number of BM scaffolds occurring exclusively in highly potent compounds, and WHPS the number of BM scaffolds occurring in both weakly and highly potent compounds. In addition, % WHPS gives the percentage of compounds represented by BM scaffolds occurring in weakly and highly potent compounds. MI is the normalized mutual information between the BM scaffold distribution and the compound potency property state, as discussed in the text.

among one or more similar scaffolds to yield highly potent compounds. Therefore, such ridges provide a basis for scaffold selection and further chemical exploration. Moreover, intermediate-MI ridges combine both types of SAR information at varying degrees. They usually contain informative analogue series and might also reveal scaffold preferences for highly potent compounds. Thus, from this point of view, the intermediate-MI category of activity ridges provides the most comprehensive SAR information.

CONCLUDING REMARKS

Herein, we have introduced the concept of an activity ridge as an information-rich data structure for SAR analysis and have demonstrated that such activity ridges are indeed frequently found in different compound activity classes. Five criteria have been formulated for the formation of an activity ridge defining its global structure. Using our implementation, the search for activity ridges has been automated, and search parameters can be easily modified. Regardless of chosen potency thresholds or numbers of ridge compounds, a characteristic feature of activity ridges is that they consist of combinatorial activity cliffs formed between the participating compounds. For the definition and assessment of activity ridges, we have applied a compound similarity criterion that is not based on calculated pairwise similarity values but rather on the assignment of active compounds to topological scaffold classes. The application of this similarity criterion led to a chemically intuitive description of activity ridges. Moreover, the choice of this similarity criterion has also enabled us to study the composition of activity ridges in detail by analyzing scaffold−compound potency relationships. For a systematic exploration of these relationships, we have introduced an analysis scheme based on the concept of mutual information. On the basis of MI analysis, we have been able to organize activity ridges into three different categories. These activity ridge categories are characterized by different composition and capture SAR information in different ways. Low, intermediate, and high MI values serve as an indicator of different activity ridge categories and can be utilized to select activity ridges that are most relevant for a given SAR analysis scenario. In summary, our study provides a previously unconsidered framework for the identification of compound subsets that are rich in SAR information, and our findings indicate that activity ridges consisting of combinatorial activity cliffs represent an interesting data structure for SAR analysis.

ASSOCIATED CONTENT

Supporting Information. Figure S1 provides examples of activity ridges from additional data sets. 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

†These authors contributed equally.

ACKNOWLEDGMENT

We thank Anne Mai Wassermann for providing help with the assembly of compound data sets. Furthermore, we are grateful to Dagmar Stumpfe for help with illustrations and tables.

REFERENCES

(1) Peltason, L.; Bajorath, J. Systematic Computational Analysis of Structure−Activity Relationships: Concepts, Challenges and Recent Advances. *Future Med. Chem.* **2009**, *1*, 451–466.

(2) 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.

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

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

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

(6) Peltason, L.; Bajorath, J. SAR Index: Quantifying the Nature of Structure−Activity Relationships. *J. Med. Chem.* **2007**, *50*, 5571–5578.

(7) Guha, R.; Van Drie, J. H. Structure−Activity Landscape Index: Identifying and Quantifying Activity Cliffs. *J. Chem. Inf. Model.* **2008**, *48*, 646–658.

(8) 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.

(9) 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.

(10) Peltason, L.; Iyer, P.; Bajorath, J. Rationalizing Three-Dimensional Activity Landscapes and the Influence of Molecular Representations on Landscape Topology and the Formation of Activity Cliffs. *J. Chem. Inf. Model.* **2010**, *50*, 1021–1033.

(11) 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.

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

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

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

(15) Cover, T. M.; Thomas, J. A. *Elements of Information Theory*; John Wiley & Sons, Inc.: New York, 1991.