🔴 Check for updates

COMMENTARY

Advancing the activity cliff concept, part II [version 1; peer review: 2 approved]

Dagmar Stumpfe, Antonio de la Vega de León, Dilyana Dimova, Jürgen Bajorath 🔵

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

Abstract

We present a follow up contribution to further complement a previous commentary on the activity cliff concept and recent advances in activity cliff research. Activity cliffs have originally been defined as pairs of structurally similar compounds that display a large difference in potency against a given target. For medicinal chemistry, activity cliffs are of high interest because structure-activity relationship (SAR) determinants can often be deduced from them. Herein, we present up-to-date results of systematic analyses of the *ligand efficiency* and *lipophilic efficiency* relationships between activity cliff-forming compounds, which further increase their attractiveness for the practice of medicinal chemistry. In addition, we summarize the results of a new analysis of *coordinated activity cliffs* and clusters they form. Taken together, these findings considerably add to our evaluation and current understanding of the activity cliff concept. The results should be viewed in light of the previous commentary article.

Open Peer Review

Reviewer Status ✔️ ✔️

	Invited Reviewers	
	1	**2**
version 1 18 Mar 2014	✔️ report	✔️ report

1 **Gerhard Müller**, Mercachem, Nijmegen, The Netherlands

2 **J. Richard Morphy**, Lilly Research Laboratories, Surrey, UK

Any reports and responses or comments on the article can be found at the end of the article.

Associated Commentary

Hu Y, Stumpfe D and Bajorath J » Advancing the activity cliff concept, F1000Research 2013, **2**:199 (https://doi.org/10.12688/f1000research.2-199.v1)

Corresponding author: Jürgen Bajorath (bajorath@bit.uni-bonn.de)

Competing interests: No competing interests were declared.

How to cite this article: Stumpfe D, de la Vega de León A, Dimova D and Bajorath J. **Advancing the activity cliff concept, part II [version 1; peer review: 2 approved]** F1000Research 2014, **3**:75 (https://doi.org/10.12688/f1000research.3788.1)

First published: 18 Mar 2014, **3**:75 (https://doi.org/10.12688/f1000research.3788.1)

Introduction

Over the past decade, the activity cliff concept has been increasingly discussed in the chemoinformatics and medicinal chemistry literature[1–3]. In the practice of medicinal chemistry, activity cliffs, which are formed by structurally similar or analogous compounds with large potency differences for a given target[1,2], have long been considered during chemical optimization efforts, typically for individual compound series[2]. However, the increasing popularity of the activity cliff concept can at least in part be attributed to computational exploration and large-scale analysis[2,3]. In fact, much of our current knowledge about activity cliffs has resulted from compound data mining and other chemoinformatics investigations[2–4]. Hence, in addition to supporting practical applications in compound development, activity cliff research is an area where chemoinformatics and medicinal chemistry meet.

In a previous commentary[4], we have summarized key aspects of the activity cliff concept and discussed further extensions and refinements. Among others, discussed topics have included the current frequency of occurrence of activity cliffs, their dependence on chosen molecular representations, their target distributions, and associated structure-activity relationship (SAR) information[4]. Herein, we present a follow up to this commentary, which has been catalyzed by the availability of new results concerning the ligand efficiency and lipophilic efficiency of activity cliff partners as well as the topology of coordinated activity cliffs formed across currently available bioactive compounds. These findings should also be considered as further advancements of the activity cliff concept and viewed on the basis of the previous commentary article.

Activity cliff definition

The definition of activity cliffs requires the specification of a *similarity criterion* (when are two compounds "similar"?) and a *potency difference criterion* (when is a potency difference "large" and "significant"?)[1,2]. Molecular similarity can be assessed in a variety of ways, which can roughly be divided into chemical descriptor-based approaches, which require the calculation of similarity values based on the comparison of chosen molecular representations[2], and substructure-based approaches, which directly establish structural relationships (on the basis of molecular graphs)[2]. Among substructure-based approaches, the matched molecular pair (MMP) formalism[5] has become very popular in recent years. An MMP is defined as a pair of compounds that are distinguished by the exchange of a substructure at a single site[5] termed a chemical transformation[6]. The formation of an MMP can thus be considered as a possible similarity criterion for activity cliff formation. To define activity cliffs, transformation size-restricted MMPs have been introduced in which transformations are limited to small and chemically meaningful replacements[7]. Accordingly, transformation size-restricted MMPs mostly account for structural analogs.

In the following, we consistently adhere to our previously rationalized preferred activity cliff definition[4]:

(a) Similarity criterion: Formation of a transformation size-restricted MMP.

(b) Potency difference criterion: At least two orders of magnitude (≥ 100-fold).

(c) Activity measurements: Equilibrium constants (K_i values).

So defined activity cliffs have also been termed MMP-cliffs[7].

Ligand efficiency and lipophilic efficiency

For the assessment of activity cliffs, compound potency has thus far been a focal point, consistent with the original activity cliff definition. However, during compound optimization, other criteria are often applied to monitor progress that relate potency to changes in molecular weight (MW) or hydrophobicity[8]. These criteria are formalized as optimization indices and include, among others, *ligand efficiency* (LE)[8,9] or *ligand lipophilic efficiency* (LLE)[8,10], which is also termed *lipophilic efficiency* (LipE)[8,11,12]. For an active compound, LE yields the fraction of potency per non-hydrogen atom or MW unit[9]. In the presence of strong and specific ligand-target interactions, LE should increase during compound optimization; in other words, a gain in potency should not primarily be attributed to molecular size effects. Herein, LE was calculated using the binding efficiency index (BEI)[13] defined as:

$$\text{BEI (LE)} = pK_i/\text{MW [log unit/kDa]}$$

Furthermore, LipE was calculated as[10]:

$$\text{LipE (LLE)} = pK_i - cLogP \text{ [log unit].}$$

LipE is obtained by subtracting the logarithm of the calculated octanol/water partition coefficient, a measure of hydrophobicity, from the logarithm of the equilibrium constant. Hence, LipE indirectly accounts for the influence of hydrophobicity on potency. LipE should also increase during compound optimization because a gain in potency should not primarily be attributed to increasing hydrophobicity of a compound (which often gives rise to nonspecific binding effects).

Because LE and LipE are important and widely applied measures for compound optimization in drug discovery, it makes sense to also consider them in the context of activity cliff formation, given their immediate relevance for SAR analysis. In a recent analysis[14], 18,208 activity cliffs were extracted from more than 41,000 unique ChEMBL[15] (release 15) compounds with activity against the current spectrum of human targets. Then, differences in LE between lowly and highly potent cliff partners were systematically assessed. For activity cliffs based upon our preferred definition, one would hope that favorable changes in LE might be observed in many instances. However, whether or not systematic trends might be detectable was an open question. The analysis revealed that the formation of 99.1% of all activity cliffs across different targets was accompanied by consistent increases in LE values between lowly and highly potent cliff partners, with, on average, ΔLE = 6.25[14]. For activity cliffs defined on the basis of calculated (molecular fingerprint-based) similarity values, comparable observations were made[14].

Here, we report results of LE and, in addition, LipE analysis for the most recent release of the ChEMBL database (version 17). From a total of ~45,000 unique ChEMBL compounds active against 661 different human targets (~77,000 K_i values), 20,080 activity cliffs were isolated. For the highly and lowly potent partners of each cliff, LE and LipE were calculated. The resulting value distributions are

displayed in Figure 1. An increase in LE and LipE values for the highly potent cliff partner was detected for 99.1% and 96.7% of all activity cliffs, respectively, with, on average, ΔLE = 6.27 and ΔLipE = 2.42. Hence, similarly positive LE and LipE trends were observed for activity cliff formation (for LipE, this was difficult to predict). These findings further emphasize the relevance of activity cliff information for medicinal chemistry applications because chemical modifications encoded by activity cliffs consistently increase potency, LE, and LipE.

Coordinated activity cliffs

The assessment of activity cliffs has conventionally focused on compound pairs. Hence, cliffs are typically considered on an individual basis (including statistical analysis). However, activity cliffs are only formed in isolation if participating compounds have no structural neighbors with which they also form cliffs. This is unlikely for many compound series and data sets, especially those resulting from compound optimization efforts. In earlier studies, series of highly and lowly potent analogs have been identified in different data sets that formed multiple overlapping activity cliffs[16]. These cliff arrangements have been termed *coordinated activity cliffs*[17]. Indeed, based upon global statistical analysis, we have determined that ~97% of all activity cliffs are formed in a coordinated manner[4]. In principle, coordinated activity cliffs have higher SAR information content than and are thus of particular interest for medicinal chemistry. However, only very little information has thus far been available about how coordinated activity cliffs are formed and what the size of coordinated cliff arrangements might be.

Therefore, in a recent study, all activity cliffs extracted from active compounds in ChEMBL (version 17) were subjected to network analysis[18]. Activity cliff forming compounds were represented as
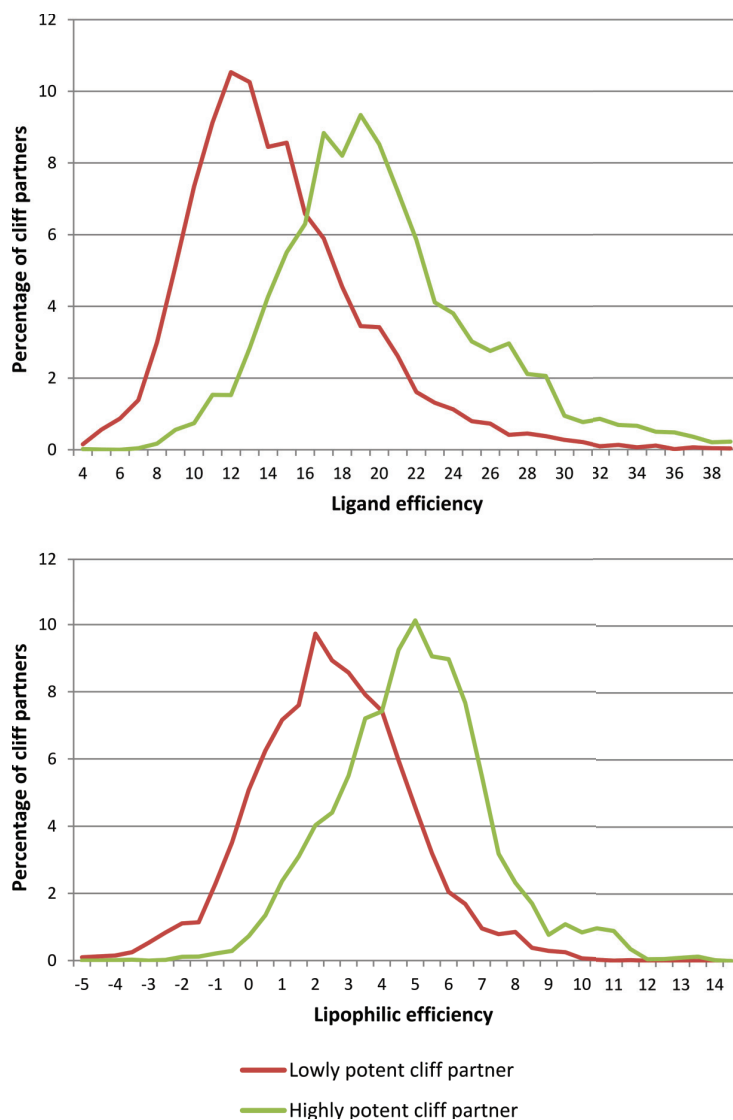

Figure 1. Ligand and lipophilic efficiency value distributions. LE (top) and LipE (bottom) value distributions are shown for lowly (red line) and highly potent (green) cliff partners.

nodes that were connected by edges accounting for individual activity cliffs. The global network is depicted in Figure 2A. It consisted of activity cliffs formed by compounds with activity against a total of 293 targets, and more than 93% of all activity cliffs were found to be single-target cliffs[18]. Only 769 (3.8%) of a total of 20,080 activity cliffs were formed in isolation. Coordinated activity cliffs appeared as different-sized disjoint clusters of varying topologies. In total, 19,311 coordinated activity cliffs formed 1303 separate clusters. Among these were 26 clusters consisting of more than 50 compounds and 420 clusters containing six to 15 compounds, hence reflecting a high degree of activity cliff coordination.

The activity cliff clusters displayed 449 distinct topologies with different frequency of occurrence. Examples are provided in Figure 2A and 2B. The majority of activity cliff clusters, i.e., 861, were assigned to only three recurrent main topologies, termed the *star*, *chain*, and *rectangle* topology, and a limited number of extensions and combinations of these topologies, as illustrated in Figure 2B.
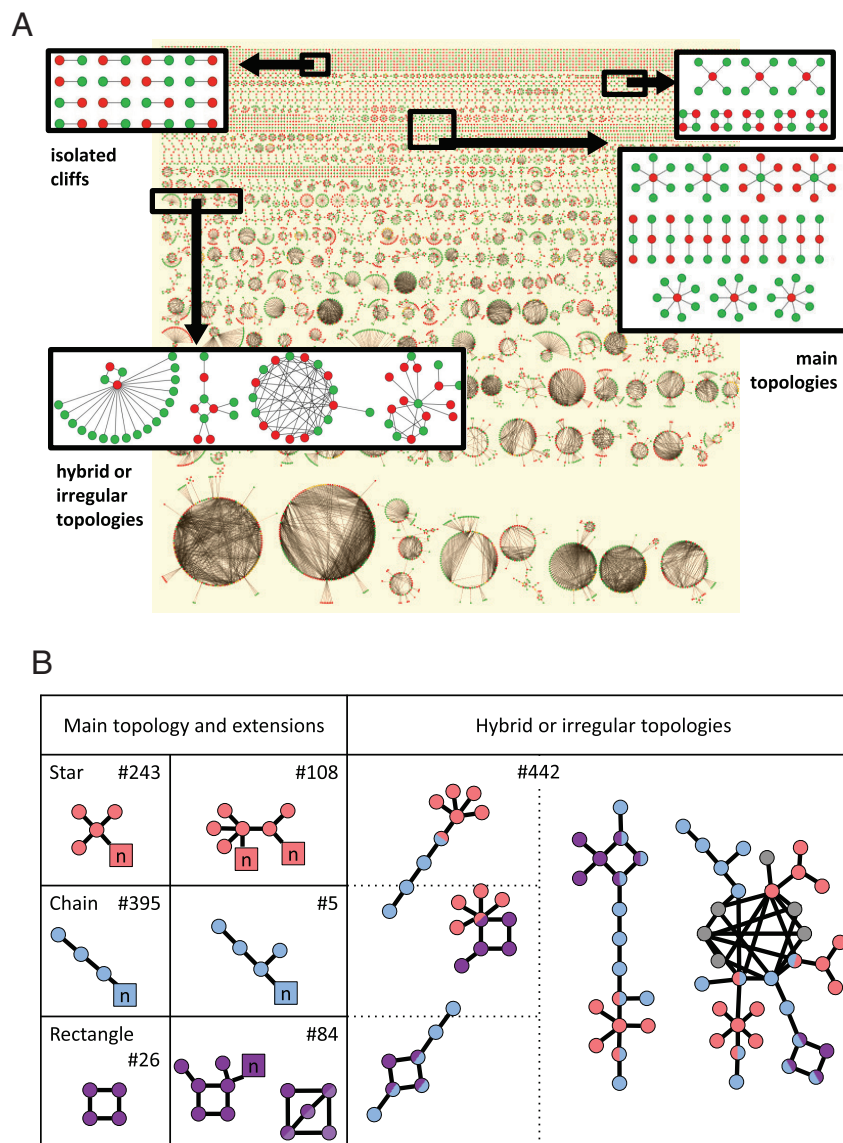

Figure 2. Activity cliff network and cluster topologies. In (**A**), the complete activity cliff network is shown. Nodes represent compounds and edges activity cliffs. Nodes of highly and lowly potent cliff partners are colored green and red, respectively, and nodes representing a compound that is a highly and lowly potent partner in different activity cliffs are colored yellow. Small sections of the network containing exemplary activity cliff cluster topologies are magnified. On the right, examples of the three most frequently occurring (main) topologies are displayed, which include so-called *star*, *chain*, and *rectangle* topology. In (**B**), main activity cliff cluster topologies and observed extensions as well as hybrid and irregular topologies are schematically illustrated (pink nodes: *star*; light blue: *chain*; purple: *rectangle* topology; gray: no topology assignment). Dual-color nodes indicate compounds belonging to cluster components with hybrid topologies. Squared nodes represent variable compound numbers (*n*) for a given topology. For each topology, the number (*#*) of instances in the network is reported.

The recurrent topologies covered many clusters of small to medium size. Topologies of increasingly large size often had hybrid character or became irregular, as also illustrated in Figure 2B.

The star topology reflects the presence of a highly or lowly potent compound and multiple analogs having opposite potency, a situation frequently observed in compound optimization. In total, the star topologies and its extensions were detected 351 times. Different from clusters with star topology, chains with more than three compounds and rectangles require the presence of alternating highly and lowly potent compounds forming sequences of activity cliffs or circular arrangements, which are less likely than stars. However, these topologies were also recurrent.

Activity cliff network analysis has revealed how coordinated activity cliffs are formed across currently available compound activity classes. Taken together, the results clearly indicate that many coordinated activity cliffs occur as well-defined clusters with recurrent topologies, which can be easily isolated and subjected to SAR exploration. For a detailed characterization of the global activity cliff network and individual cluster topologies, the interested reader is referred to the original publication[18].

Conclusions

Herein, we have presented an update on the state-of-the-art in rationalizing activity cliffs. Focal points of our analysis have been the large-scale characterization of activity cliffs in terms of ligand efficiency and lipophilic efficiency as well as the visualization and systematic assessment of coordinated activity cliffs. The finding that activity cliff formation is generally accompanied by improvements in ligand and lipophilic efficiency further increases the attractiveness of activity cliff information for compound optimization. In addition, the observation that coordinated activity cliffs often form clusters of well-defined topology, irrespective of specific compound activities, is relevant for SAR analysis. Because activity clusters are rich in SAR information, an important topic for future research will be how such SAR information might be systematically extracted from clusters with different topology.

References

1. Maggiora GM: **On outliers and activity cliffs--why QSAR often disappoints.** *J Chem Inf Model.* 2006; **46**(4): 1535.
   PubMed Abstract | **Publisher Full Text**

2. Stumpfe D, Bajorath J: **Exploring activity cliffs in medicinal chemistry.** *J Med Chem.* 2012; **55**(7): 2932–2942.
   PubMed Abstract | **Publisher Full Text**

3. Stumpfe D, Hu Y, Dimova D, *et al.*: **Recent progress in understanding activity cliffs and their utility in medicinal chemistry.** *J Med Chem.* 2014; **57**(1): 18–28.
   PubMed Abstract | **Publisher Full Text**

4. Hu Y, Stumpfe D, Bajorath J: **Advancing the activity cliff concept [v1; ref status: indexed, http://f1000r.es/1wf].** *F1000Res.* 2013; **2**: 199.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Kenny PW, Sadowski J: **Structure modification in chemical databases.** *Chemoinformatics in Drug Discovery.* Oprea, T. I., Ed.; Wiley-VCH: Weinheim, Germany, 2005; pp 271–285.
   Publisher Full Text

6. Hussain J, Rea C: **Computationally efficient algorithm to identify matched molecular pairs (MMPs) in large data sets.** *J Chem Inf Model.* 2010; **50**(3): 339–348.
   PubMed Abstract | **Publisher Full Text**

7. Hu X, Hu Y, Vogt M: **MMP-Cliffs: systematic identification of activity cliffs on the basis of matched molecular pairs.** *J Chem Inf Model.* 2012; **52**(5): 1138–1145.
   PubMed Abstract | **Publisher Full Text**

8. Hopkins AL, Keserü GM, Leeson PD, *et al.*: **The role of ligand efficiency metrics in drug discovery.** *Nat Rev Drug Discov.* 2014; **13**(2): 105–121.
   PubMed Abstract | **Publisher Full Text**

9. Hopkins AL, Groom CR, Alex A: **Ligand efficiency: a useful metric for lead selection.** *Drug Discov Today.* 2004; **9**(10): 430–431.
   PubMed Abstract | **Publisher Full Text**

10. Leeson PD, Springthorpe B: **The influence of drug-like concepts on decision-making in medicinal chemistry.** *Nat Rev Drug Discov.* 2007; **6**(11): 881–890.
    PubMed Abstract | **Publisher Full Text**

11. Freeman-Cook K, Hoffman RL, Johnson TW: **Lipophilic efficiency: the most important efficiency metric in medicinal chemistry.** *Future Med Chem.* 2013; **5**(2): 113–115.
    PubMed Abstract | **Publisher Full Text**

12. Shultz MD: **The thermodynamic basis for the use of lipophilic efficiency (LipE) in enthalpic optimizations.** *Bioorg Med Chem Lett.* 2013; **23**(21): 5992–6000.
    PubMed Abstract | **Publisher Full Text**

13. Abad-Zapatero C, Metz JT: **Ligand efficiency indices as guideposts for drug discovery.** *Drug Discov Today.* 2005; **10**(7): 464–469.
    PubMed Abstract | **Publisher Full Text**

14. de la Vega de León A, Bajorath J: **Formation of activity cliffs is accompanied by systematic increases in ligand efficiency from lowly to highly potent compounds.** *AAPS J.* 2014; **16**(2): 335–341.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Gaulton A, Bellis LJ, Bento AP, *et al.*: **ChEMBL: a large-scale bioactivity database for drug discovery.** *Nucleic Acids Res.* 2012; **40**(Database issue): D1100–D1107.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Vogt M, Huang Y, Bajorath J: **From activity cliffs to activity ridges: informative data structures for SAR analysis.** *J Chem Inf Model.* 2011; **51**(8): 1848–1856.
    PubMed Abstract | **Publisher Full Text**

17. Namasivayam V, Bajorath J: **Searching for coordinated activity cliffs using particle swarm optimization.** *J Chem Inf Model.* 2012; **52**(4): 927–934.
    PubMed Abstract | **Publisher Full Text**

18. Stumpfe D, Dimova D, Bajorath J: **Composition and topology of activity cliff clusters formed by bioactive compounds.** *J Chem Inf Model.* 2014; **54**(2): 451–461.
    PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 02 April 2014

https://doi.org/10.5256/f1000research.4057.r4182

✔ **J. Richard Morphy**
Department of Medicinal Chemistry, Lilly Research Laboratories, Surrey, UK

The twin concepts of molecular matched pair (MMP) analysis and activity cliffs have been widely embraced by medicinal chemists in recent years as a means to explore and understand complex SAR landscapes. The contributions of the Bajorath group have been pivotal to the acceptance of these approaches. Activity cliffs are large changes in a biological activity arising from modest structural differences. In this study size restrictions are imposed on the activity cliff/MMP partners to ensure that structural changes are relatively small. Therefore it might be expected that the activity cliff trends based upon ligand efficiency would mirror those from activity alone. Still it is reassuring that this paper confirms that activity cliffs are indeed associated with large benefits in terms of ligand efficiency. The more interesting observation is that activity cliffs are also associated with significant improvements in lipophilic ligand efficiency (LLE or LipE) since this is less intuitive based on the selection criteria used for the MMP analysis. Therefore this latest study adds real weight to the argument that MMPs and activity cliffs can be very useful concepts within hit and lead optimisation projects.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 21 March 2014

https://doi.org/10.5256/f1000research.4057.r4181

✔ **Gerhard Müller**
Medicinal Chemistry, Mercachem, Nijmegen, The Netherlands

Dagmar Stumpfe, Antonio de la Vega de Leon, Dilyana Dimova, and Jürgen Bajorath refer in the submitted commentary entitled "**Follow up: Advancing the activity cliff concept, part II**" to a previously published commentary on the activity cliff concept, now reporting on (i) how this concept correlates with ligand efficiency measures of "cliff-forming" compound pairs, and (ii) how "coordinated activity cliffs" give new insight into a putative course of optimization by tracing the activity-critical compounds introducing activity cliff network and cluster topologies.

The group around Jürgen Bajorath has made major contributions over the last few years to the development and establishment of the activity cliff concept within Medicinal Chemistry, documented by numerous seminal publications in the field, that provide a constantly maturing cheminformatics tool useful for the medicinal chemist to assess the steepness of an unfolding structure-activity relationship landscape. Activity cliff-forming compounds give rise to sharp curvatures in the structure-activity relationship surface and thus might suggest that affected groups establish significant pharmacophoric features in an underlying compound class. It is especially the introduction and consequent application of so-called "transformation size-restricted matched molecular pairs" that renders this concept useful for the end-user, since an abstract cheminformatics tool becomes more tangible for the practitioner in the field.

In this context the authors set out to consider as to whether the occurrence of activity cliffs is associated with a general and systematic increase in molecular weight, as this is often found in consecutive medicinal chemistry optimization rounds (e.g. transforming a primary hit into a structurally more elaborate lead-like compound).This optimization process is often accompanied by a significant increase in molecular weight, as well as in lipophilicity. For that reason, normalization of detected binding affinities or inhibition data onto e.g. the number of heavy atoms in an underlying molecule, allows to better assess and cross-compare the true efficiency of a given compound acting at a specific target. Compounds of high ligand efficiency but low molecular weight and low lipophilicity are preferred in that they show fewer liabilities in e.g. unspecific binding to other targets, or in the number of putative metabolic soft spots.

By applying stringent quality criteria for data mining, the authors have carried out a systematic analysis within the ChEMBL version 17 database identifying MMP cliffs. Approximately 20,000 activity cliffs were generated and systematically analysed for ligand efficiency changes. In more than 99% of all activity cliffs, the more potent cliff-forming compound exhibits consistently higher ligand efficiency, and in more than 96% of all identified MMP cliffs, the more active compound showed an increased lipophilic efficiency. And this finding is derived from more than 45,000 distinct compounds with biological activities reported for more than 661 targets, rendering the underlying dataset truly diverse in its chemical and biological nature.

Activity cliff analyses by their algorithmic nature focus on compound pairs, in this contribution so-called matched molecular pairs with a size-restricted transformation accounting for a single structural difference between the two cliff-forming compounds. It seems obvious that the identified and isolated 20,000 MMP-cliffs might cluster into common underlying optimization programs, since the majority of medicinal chemistry optimization campaigns cover a number of subsequent iterative feedback cycles with many structurally related compounds belonging to consecutive design generations. To account for this interdependence of isolated cliff-forming compound pairs, the group around Jürgen Bajorath has embarked into the concept of "coordinated activity cliffs".

Again, this is a very helpful attempt to back-translate an abstract and simplified view on a compound-activity space into the operational world of medicinal chemistry in which those formerly isolated pairs appear in a broader context of more comprehensive chemistry campaigns. By generating and analysing activity cliff network and cluster topologies, a hypothetical optimization pathway can be unfolded that provides additional guidelines to the practitioner on the optimization philosophy.

Admittedly, the naive end-user being confronted for the first time with the activity cliff concept will require some time to fully appreciate the intrinsic value of this cheminformatics-based approach towards the analysis of structure-activity relationships. However, the concept becomes more and more intuitive and as such is a true asset that should be applied in small-molecule drug discovery projects.

As in the previous commentary of the Bajorath group, I see the attempt to reach high user-friendliness (e.g. by explaining the concept of coordinated activity cliffs) which renders this commentary helpful in alerting the medicinal chemistry community to this very useful, but still under-appreciated concept. The group around Jürgen Bajorath continue to qualify as advocates in that sense, and the community of practicing medicinal chemists should start to move in their direction accordingly.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000**Research**