Check for updates

DATA NOTE

Comprehensive knowledge base of two- and three-dimensional activity cliffs for medicinal and computational chemistry [version 1; peer review: 3 approved]

Ye Hu, Norbert Furtmann, Dagmar Stumpfe, 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

V1 First published: 25 Jun 2015, 4(Chem Inf Sci):168
<https://doi.org/10.12688/f1000research.6661.1>

Latest published: 25 Jun 2015, 4(Chem Inf Sci):168
<https://doi.org/10.12688/f1000research.6661.1>

Abstract

Activity cliffs are formed by pairs or groups of structurally similar or analogous active compounds with large differences in potency. They can be defined in two or three dimensions by comparing graph-based molecular representations or compound binding modes, respectively. Through systematic analysis of publicly available compound activity data and ligand-target X-ray structures we have in a series of studies determined all currently available two- and three-dimensional activity cliffs (2D- and 3D-cliffs, respectively). Furthermore, we have systematically searched for 2D extensions of 3D-cliffs. Herein, we specify different categories of activity cliffs we have explored and introduce an open access data deposition in ZENODO (doi: [10.5281/zenodo.18490](https://doi.org/10.5281/zenodo.18490)) that makes the entire knowledge base of current activity cliffs freely available in an organized form.

Keywords

Active compounds , X-ray structures , activity cliffs , data mining , structure-activity relationships , computational methods

Open Peer Review

Approval Status

	1	2	3
version 1
25 Jun 2015	view	view	view
<hr/>			
1. Alexandre Varnek , University of Strasbourg, Strasbourg, France			
2. Veer Shanmugasundaram , Pfizer PharmaTherapeutics Research & Development, Groton, USA			
3. Gerhard Müller , Mercachem, Nijmegen, The Netherlands			

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

This article is included in the **Chemical Information Science** gateway.

This article is included in the **Data: Use and Reuse** collection.

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

Competing interests: No competing interests declared.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2015 Hu Y *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Hu Y, Furtmann N, Stumpfe D and Bajorath J. **Comprehensive knowledge base of two- and three-dimensional activity cliffs for medicinal and computational chemistry [version 1; peer review: 3 approved]** F1000Research 2015, 4 (Chem Inf Sci):168 <https://doi.org/10.12688/f1000research.6661.1>

First published: 25 Jun 2015, 4(Chem Inf Sci):168 <https://doi.org/10.12688/f1000research.6661.1>

Introduction

The activity cliff concept has experienced increasing interest in chemical informatics and medicinal chemistry^{1–5}. A consensus definition of activity cliffs^{1–4} refers to pairs or groups of structurally similar or analogous active compounds with large differences in potency^{4,5}. For the definition of activity cliffs, the specification of similarity and potency difference criteria is required. Two-dimensional activity cliffs (2D-cliffs) have mostly been defined on the basis of Tanimoto similarity⁶ comparing molecular fingerprint representations². More recently, 2D-cliffs have also been defined on the basis of substructure relationships, preferably employing the matched molecular pair (MMP) formalism^{7,8}, leading to the introduction of MMP-cliffs⁹. An MMP is defined as a pair of compounds that are only distinguished by a structural change at a single site⁷, i.e., the exchange of a substructure, termed a chemical transformation⁸. For the definition of MMP-cliffs, transformation size restrictions have been introduced to limit transformations to small chemical changes typically observed in analog series⁹. Applying well-defined similarity and potency difference criteria, 2D-cliffs can be systematically extracted from compound databases¹⁰.

The vast majority of 2D-cliffs (i.e., close to or more than 95%, depending on the molecular representations and similarity measure used) are not formed in isolation (i.e., in the absence of structural neighbors with significant potency variations), but rather in a coordinated manner involving series of compounds with varying potency forming multiple and overlapping cliffs^{4,5,11}. In activity cliff network representations where nodes represent compounds and edges activity cliffs, coordinated cliffs emerge as individual clusters of varying composition and size¹¹, which can be isolated for further analysis.

In addition to 2D-cliffs, three-dimensional activity cliffs (3D-cliffs) can also be defined by comparing compound binding modes in X-ray structures¹². This requires the superposition of structures of a given target available in different crystallographic ligand-target complexes and the assessment of the 3D similarity of bound ligands¹². Three-dimensional activity cliffs can be further extended by taking 2D ligand information into account. This can be accomplished by systematically searching compound activity classes for analogs of 3D-cliff partners¹³. For example, for each cliff partner, MMPs with database compounds sharing the same activity can be determined and qualifying analogs can be assigned to the 3D-cliff¹³, leading to what we term herein a 3D-cliff-MMP extension. **Figure 1** shows an exemplary 2D-cliff (MMP-cliff), activity cliff cluster, 3D-cliff, and 3D-cliff-MMP extension.

In medicinal chemistry, 2D-cliffs are often considered in the context of structure-activity relationship (SAR) analysis and compound design^{2,3}. For SAR exploration, activity cliff clusters are of particular interest because they provide more SAR information than 2D-cliffs studied individually. Furthermore, 3D-cliffs are of prime interest for structure-based design and also for computational chemistry applications including, for example, the calibration of scoring functions or free energy (perturbation) calculations. Last but not least, 2D extensions of 3D-cliffs bridge different applications in medicinal and computational chemistry and help to identify candidate compounds for further analysis.

Methods and materials

The activity cliff information provided herein is the result of recent surveys and systematic analyses of 2D-cliffs including clusters¹⁴, 3D-cliffs¹⁵, and extensions of 3D-cliffs¹⁶. **Table 1** summarizes the different activity cliff categories. All 2D-cliffs reported herein originated from the most recent release of ChEMBL (version 20)^{17,18} and all 3D-cliffs from the Protein Data Bank (PDB; accessed December, 2014)¹⁹. MMP extensions of 3D-cliffs were identified in ChEMBL (version 19).

For all activity cliffs, an at least 100-fold difference in potency between cliff partners was consistently required. For 2D-cliffs, only (assay-independent) K_i values were considered as potency measurements. For 3D-cliffs, K_i and IC_{50} measurements were separately considered (using a K_i and IC_{50} value-based data set, respectively). In addition, 3D-cliffs were also determined in a combined K_i/IC_{50} data set (taking into consideration that 3D-cliffs provide a much smaller knowledge base than 2D-cliffs; *vide infra*). For 3D-cliff analysis, a crystallographic resolution limit of 3.0 Å was applied.

Two-dimensional activity cliffs were determined using three different molecular representations including the extended connectivity fingerprint with bond diameter 4 (ECFP4)²⁰, molecular access system (MACCS) structural keys²¹, and transformation size-restricted MMPs (MMP-cliffs)⁹. As similarity criteria for ECFP4- and MACCS-based activity cliffs, Tanimoto coefficient threshold values of 0.55 and 0.85 were applied, respectively^{2,14}. For an MMP-cliff, our preferred 2D-cliff definition^{3,4}, the formation of a transformation size-restricted MMP served as a similarity criterion. By definition 2D-cliffs do not contain stereochemical information.

For the identification of 3D-cliffs, the normalized overlap of atomic property density functions calculated for a pair of bound ligands was used as a measure of 3D similarity, taking conformational, positional, and atomic property differences into account¹². An at least 80% calculated 3D similarity was required as a threshold for 3D-cliff formation^{12,15}.

Data description

Activity cliff statistics are reported in **Table 1**. A total of 17,111 MMP-cliffs, 31,975 ECFP4-, and 34,813 MACCS-based 2D-cliffs were identified formed by compounds active against more than 300 targets in each case. The corresponding number of activity clusters (comprising at least 3 compounds) was 1267, 1462, and 1402, respectively. Therefore, a very large knowledge base of well-defined 2D-cliffs is currently available. In addition, on the basis of K_i and IC_{50} measurements, 236 and 292 3D-cliffs were detected and were formed by crystallographic ligands of 26 and 43 targets, respectively. The combined K_i/IC_{50} data set yielded 595 3D-cliffs for 58 targets. Although many more 2D- than 3D-cliffs are currently available, as one would expect, the number of 3D-cliffs is larger than we anticipated, hence providing substantial opportunities for structural and computational studies. **Table 2** provides details for the 61 different targets for which 3D-cliffs were detected. Furthermore, more than 1000 3D-cliff-MMPs were identified for each of the K_i and IC_{50} data sets and 2608 for the combined set (**Table 1**). Hence, for many 3D-cliffs, active analogs are available whose SAR characteristics and possible interaction patterns can be explored

Figure 1. Exemplary 2D- and 3D-cliffs. Different categories of activity cliffs are shown formed by inhibitors of tyrosine kinase ABL. MMP-cliffs are used to represent 2D-cliffs. For each compound, the ChEMBL or Protein Data Bank (PDB) ID and its negative logarithmic potency value are reported. **(a)** An exemplary MMP-cliff (structural modification highlighted in red) taken from an activity cliff cluster (dashed blue box) is shown. In an activity cliff network, nodes represent compounds and edges cliffs. Nodes are colored according to potency values using a continuous color spectrum from red (lowest potency) via yellow (intermediate) to green (highest potency). In network representations, coordinated activity cliffs emerge as clusters. **(b)** An exemplary 3D-cliff and its 2D extension are shown (3D-cliff-MMP). The extension results from MMP-based mapping of analogs from ChEMBL to 3D-cliff compounds. Structural differences between 3D-cliff compounds and their 2D (MMP) partners are highlighted in red.

Table 1. Activity cliff categories and statistics. Reported are the number of 2D-cliffs belonging to three different categories (FP, fingerprint; Tc, Tanimoto coefficient), corresponding activity cliff clusters (comprising at least three compounds), 3D-cliffs for different potency measurement-dependent data sets, and corresponding 3D-cliff-MMPs (giving the total number of MMPs detected for 3D-cliffs from each data set).

				#Cliffs/clusters	#Targets
2D (K _i)	FP Tc-based	MACCS	Cliffs	34,813	320
			Cliff clusters	1402	
		ECFP4	Cliffs	31,975	339
			Cliff clusters	1462	
	MMP		Cliffs	17,111	301
			Cliff clusters	1267	
3D	3D-cliffs		K _i	236	26
			IC ₅₀	292	43
			K _i /IC ₅₀	595	58
	3D-cliff-MMPs		K _i	1017	18
			IC ₅₀	1113	35
			K _i /IC ₅₀	2608	48

Table 2. Targets with available 3D-cliffs. A total of 61 targets are listed for which 3D-cliffs were detected. For each target, the ChEMBL ID, UniProt accession ID (UniProtID)²², and the number of available 3D-cliffs are reported. 3D-cliffs were separately determined on the basis of only K_i or IC₅₀ measurements (available for active compounds) as well as for the combined K_i and IC₅₀ data set (K_i/IC₅₀). In addition, the number of MMP-cliffs (if available; defined on the basis of K_i values) is also reported for each target.

Target ID (ChEMBL)	UniProtID	Target name	#3D-cliffs			#MMP-cliffs
			K _i	IC ₅₀	K _i /IC ₅₀	K _i
202	P00374	Dihydrofolate reductase	0	3	5	0
204	P00734	Thrombin	158	2	166	296
205	P00918	Carbonic anhydrase II	8	6	24	397
206	P03372	Estrogen receptor alpha	3	2	10	3
235	P37231	Peroxisome proliferator-activated receptor gamma	0	1	2	24
239	Q07869	Peroxisome proliferator-activated receptor alpha	0	0	1	0
242	Q92731	Estrogen receptor beta	1	0	2	10
244	P00742	Coagulation factor X	21	1	28	1130
260	Q16539	MAP kinase p38 alpha	0	11	13	68
262	P49841	Glycogen synthase kinase-3 beta	1	0	2	11
267	P12931	Tyrosine-protein kinase SRC	0	2	2	20
275	Q07343	Phosphodiesterase 4B	0	2	2	0
279	P35968	Vascular endothelial growth factor receptor 2	1	2	4	12
280	P45452	Matrix metalloproteinase 13	0	2	1	31
283	P08254	Matrix metalloproteinase 3	0	0	2	18
284	P27487	Dipeptidyl peptidase IV	0	4	5	40
286	P00797	Renin	0	8	7	7
288	Q08499	Phosphodiesterase 4D	0	2	2	0
301	P24941	Cyclin-dependent kinase 2	3	26	41	7

Target ID (ChEMBL)	UniProtID	Target name	#3D-cliffs			#MMP-cliffs
			K _i	IC ₅₀	K _i /IC ₅₀	K _i
335	P18031	Protein-tyrosine phosphatase 1B	1	3	8	20
1782	P14324	Farnesyl diphosphate synthase	0	2	2	3
1808	P12821	Angiotensin-converting enzyme	2	0	2	19
1827	O76074	Phosphodiesterase 5A	0	1	1	2
1862	P00519	Tyrosine-protein kinase ABL	0	2	0	18
1871	P10275	Androgen Receptor	1	0	0	15
1892	Q04609	Glutamate carboxypeptidase II	2	0	2	42
1918	P39086	Glutamate receptor ionotropic kainate 1	3	0	3	43
1966	Q02127	Dihydroorotate dehydrogenase	1	1	4	2
2147	P11309	Serine/threonine-protein kinase PIM1	0	7	7	21
2179	P04062	Beta-glucocerebrosidase	1	0	1	8
2288	Q13526	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1	0	3	4	26
2308	Q9UKM7	Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase	0	1	1	7
2360	P00492	Hypoxanthine-guanine phosphoribosyltransferase	1	0	1	7
2524	P06280	Alpha-galactosidase A	3	0	3	0
2527	O96017	Serine/threonine-protein kinase Chk2	0	12	12	0
2534	O15530	3-phosphoinositide dependent protein kinase-1	0	4	5	0
2835	P23458	Tyrosine-protein kinase JAK1	1	0	1	7
3267	P48736	PI3-kinase p110-gamma subunit	0	0	2	8
3286	P00749	Urokinase-type plasminogen activator	7	2	12	6
3587	Q02750	Dual specificity mitogen-activated protein kinase kinase 1	0	11	11	0
3589	P55263	Adenosine kinase	0	1	1	1
3717	P08581	Hepatocyte growth factor receptor	0	3	5	30
3835	P51955	Serine/threonine-protein kinase NEK2	0	4	4	0
3880	P07900	Heat shock protein HSP 90-alpha	0	39	43	0
3922	P50579	Methionine aminopeptidase 2	0	4	4	33
3959	P16083	Quinone reductase 2	0	1	2	0
3975	P09467	Fructose-1,6-bisphosphatase	0	1	1	0
3991	P08709	Coagulation factor VII	4	0	4	2
4073	P09237	Matrix metalloproteinase 7	1	0	1	3
4393	P39900	Matrix metalloproteinase 12	1	0	2	4
4439	P36897	TGF-beta receptor type I	0	1	0	187
4581	P52732	Kinesin-like protein 1	0	4	3	110
4588	P22894	Matrix metalloproteinase 8	1	3	3	13
4617	P11086	Phenylethanolamine N-methyltransferase	6	0	6	67
4618	P09960	Leukotriene A4 hydrolase	0	32	37	0
4630	O14757	Serine/threonine-protein kinase Chk1	0	8	8	1
4722	O14965	Serine/threonine-protein kinase Aurora-A	0	4	4	1
4822	P56817	Beta-secretase 1	2	57	57	72
5147	P54760	Ephrin type-B receptor 4	0	4	4	0
5879	O60760	Hematopoietic prostaglandin D synthase	0	3	3	0
1795117	Q8TEK3	Histone-lysine N-methyltransferase, H3 lysine-79 specific	2	0	2	32

based upon 3D-cliff information, for example, by superposing them onto 3D-cliff compounds with which they form transformation size-restricted MMPs.

Data availability

The activity cliff information described above is made freely available in four separate data files containing 2D-cliffs, 3D-cliffs, 3D-cliff-MMP extensions, and superpositions of complex X-ray structures and 3D ligands for selected targets:

- (1) *2D-Cliffs_and_Cliff-Clusters.xlsx* (Excel format): 2D-cliffs and clusters belonging to different categories are separately recorded using ChEMBL IDs.
- (2) *3D-Cliffs.xlsx* (Excel): 3D-cliffs from the K_i , IC_{50} , and K_i/IC_{50} data sets are separately provided using PDB IDs for compounds and UniProt²² IDs for targets.
- (3) *3D-Cliff_Extension.xlsx* (Excel): Analogs of 3D-cliff compounds identified by MMP search are reported. For each of the three data sets, all 3D-cliff-MMPs are listed.
- (4) *Superpositions.zip* (all files in MOL2 format): For each target in Table 2, superpositions of complex X-ray structures and 3D ligands are provided.

These data sets are contained in an open access ZENODO deposition²³. The deposition also contains a *README* document that details the data organization and information provided.

Conclusions

In this study, we have discussed different categories of activity cliffs (including cliff extensions) and reported the distribution of cliffs belonging to these categories. Given the cliff definitions applied herein, the activity cliff information we provide as an open access deposition is up-to-date and comprehensive. We hope that this large knowledge base of activity cliffs will be helpful in the practice of medicinal chemistry and structure-based drug design as well as in further evaluating and advancing computational methods.

Data availability

ZENODO: Knowledge base of two- and three-dimensional activity cliffs, doi: [10.5281/zenodo.1849023](https://doi.org/10.5281/zenodo.1849023)

Author contributions

JB designed the study, NF, YH, and DS collected, organized, and deposited the data, JB wrote the manuscript, all authors examined the manuscript.

Competing interests

No competing interests declared.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgements

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

References

- Maggiore GM: **On outliers and activity cliffs--why QSAR often disappoints.** *J Chem Inf Model.* 2006; **46**(4): 1535.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Stumpfe D, Bajorath J: **Exploring activity cliffs in medicinal chemistry.** *J Med Chem.* 2012; **55**(7): 2932–2942.
[PubMed Abstract](#) | [Publisher Full Text](#)
- 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](#)
- Hu Y, Stumpfe D, Bajorath J: **Advancing the activity cliff concept [v1; ref status: indexed, <http://f1000r.es/1wrf>].** *F1000Research.* 2013; **2**: 199.
[Publisher Full Text](#)
- Hu Y, Stumpfe D, de la Vega de León A, Dimova D, *et al.*: **Follow up: Advancing the activity cliff concept, part II [v1; ref status: indexed, <http://f1000r.es/34p>].** *F1000Research.* 2014; **3**: 75.
[Publisher Full Text](#)
- Willett P, Barnard JM, Downs GM: **Chemical similarity searching.** *J Chem Inf Comput Sci.* 1998; **38**(6): 983–996.
[Publisher Full Text](#)
- Kenny PW, Sadowski J: **Structure modification in chemical databases.** In *Chemoinformatics in Drug Discovery*; Oprea, T. I., Ed.; Wiley-VCH: Weinheim, Germany, 2005; 271–285.
[Publisher Full Text](#)
- 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](#)
- Hu X, Hu Y, Vogt M, *et al.*: **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](#)
- Stumpfe D, Bajorath J: **Frequency of occurrence and potency range distribution of activity cliffs in bioactive compounds.** *J Chem Inf Model.* 2012; **52**(9): 2348–2353.
[PubMed Abstract](#) | [Publisher Full Text](#)
- 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](#)
- Hu Y, Furtmann N, Gütschow M, *et al.*: **Systematic identification and classification of three-dimensional activity cliffs.** *J Chem Inf Model.* 2012; **52**(6): 1490–1498.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Furtmann N, Hu Y, Bajorath J: **Comprehensive analysis of three-dimensional activity cliffs formed by kinase inhibitors with different binding modes and cliff mapping of structural analogues.** *J Med Chem.* 2015; **58**(1): 252–264.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Stumpfe D, Bajorath J: **Monitoring global growth of activity cliff information over time and assessing activity cliff frequencies and distributions.** *Future Med Chem.* 2015; **7**: in press.
- Furtmann N, Hu Y, Gütschow M, *et al.*: **Identification and analysis of the**

- currently available high-confidence three-dimensional activity cliffs. *RSC Adv.* 2015; **5**(54): 43660–43668.
[Publisher Full Text](#)
16. Hu Y, Furtmann N, Bajorath J: **Extension of three-dimensional activity cliff information through systematic mapping of active analogs.** *RSC Adv.* 2015; **5**(54): 43006–43015.
[Publisher Full Text](#)
 17. Gaulton A, Bellis LJ, Bento AP, *et al.*: **ChEMBL: a large-scale bioactivity database for drug discovery.** *Nucleic Acids Res.* 2011; **40**(Database issue): D1100–D1107.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 18. Bento AP, Gaulton A, Hersey A, *et al.*: **The ChEMBL bioactivity database: an update.** *Nucleic Acids Res.* 2014; **42**(Database issue): D1083–D1090.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 19. Berman HM, Westbrook J, Feng Z, *et al.*: **The protein data bank.** *Nucleic Acids Res.* 2000; **28**(1): 235–242.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 20. Rogers D, Hahn M: **Extended-connectivity fingerprints.** *J Chem Inf Model.* 2010; **50**(5): 742–754.
[PubMed Abstract](#) | [Publisher Full Text](#)
 21. **MACCS Structural keys.** Accelrys, San Diego, CA, USA.
[Reference Source](#)
 22. UniProt Consortium: **The Universal Protein Resource (UniProt) in 2010.** *Nucleic Acids Res.* 2010; **38**(Database issue): D142–D148.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 23. Hu Y, Furtmann N, Stumpfe D, *et al.*: **Knowledge base of two- and three-dimensional activity cliffs.** *Zenodo.* 2015.
[Data Source](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 16 July 2015

<https://doi.org/10.5256/f1000research.7155.r9508>

© 2015 Müller G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gerhard Müller

Medicinal Chemistry, Mercachem, Nijmegen, The Netherlands

Y. Hu, N. Furtmann, D. Stumpfe, and J. Bajorath report on a conceptual extension of their well-exemplified activity cliff concept from the previously established graph-based two-dimensional representation to a three-dimensional version by factoring crystallographically solved high-resolution complex structures of respective ligands into the comparative analyses of sub-structural molecular changes linked to changes in activity. With that extension into three-dimensional space they account for ligand-target interactions, thus providing activity cliff-forming compound sets a design-relevant context that is of immediate assistance when embedded in a molecular design campaign. This clearly is a useful enrichment of the toolbox of computational, as well as medicinal chemists.

The authors provide all the relevant molecular similarity and potency difference criteria underlying this analyses that are essential for the definition of activity cliff-forming compound clusters, together with a sound description of all methodological details. While the traditional 2D cliffs were based on e.g. Tanimoto similarity and the size-restricted matched molecular pair concept, introduced by the Bajorath group earlier, 3D cliffs are based on mutual molecular similarity for ligands for which high-resolution complex structures are available. Obtained 3D activity cliffs were further enriched to 3D-cliff-MMP extended sets by including molecular similarity considerations to active compounds for which no x-ray structure is available.

In total, app. 17.000 MMP cliffs have been identified at more than 300 distinct biological targets revealing close to 1.300 activity clusters. Cumulatively, 600 3D cliffs have been detected involving app. 60 distinct protein targets, respectively. In Table 2, the authors provide a comprehensive overview of the target landscape and underlying statistics for the detected 3D activity cliffs. In addition, all relevant data are accessible and an open access data deposition in ZENODO has been established.

It is especially the extended 3D-cliff-MMP datasets in which ligands with experimentally determined binding modes and interaction patterns serve as probe compound for closely related active analogues with available SAR information that bears a huge potential to extrapolate

medicinal chemists' understanding of structure-activity relationships from a pure comparative framework into a 3D direct design concept. An immediate interrogation of the binding site's functionalities becomes amenable to a previously restricted indirect design approach. It will be interesting to see as to whether the 3D activity concept introduced in this contribution can be extended to a better understanding of structure-selectivity relationships of compound sets acting e.g. at different isoforms of densely populated target families. Summarizing, this contribution laid the basis for migrating a formerly indirect design-restricted tool for comprehensive SAR analysis into 3D space, actually into the binding pocket of investigated ligand sets, thus increasing the interpretability and the feasibility of identified cliff information for the community of practicing medicinal chemists

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 09 July 2015

<https://doi.org/10.5256/f1000research.7155.r9206>

© 2015 Shanmugasundaram V. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Veer Shanmugasundaram

Center of Chemistry Innovation & Excellence, Pfizer PharmaTherapeutics Research & Development, Groton, CT, USA

Recently there has been a lot of interest and effort in curating databases and providing MMP information for analysis of various properties by both academic and industrial groups as noted by some references below

<http://dx.doi.org/10.1021/ci5005256>

<http://dx.doi.org/10.1021/jm400223y>

<http://dx.doi.org/10.1021/jm500317a>

Further, a few companies have started providing content and are in the data-provisioning business accessing corporate information through consortia like the one below from Medchemica:

<http://www.medchemica.com/index.html>

<http://www.astrazeneca.com/Research/news/Article/260613-roche-and-astrazeneca-launch-medicinal-chemistry-datasha>

Or developing capabilities to use such information in drug discovery applications such as NextMove:

<http://www.nextmovesoftware.com>

The data note and open access data submitted here from the Bajorath Group based on their comprehensive work using 2D and 3D activity cliff information provides a wealth of information that directly will address many questions that medicinal chemistry project teams typically tend to ask. Very relevant, timely and valuable effort that should be communicated broadly to the research community.

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 06 July 2015

<https://doi.org/10.5256/f1000research.7155.r9335>

© 2015 Varnek A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Alexandre Varnek

Department of Chemistry, University of Strasbourg, Strasbourg, France

This is short but nice paper describing the data related to 2D and 3D activity cliffs for large variety of biological targets. The data can freely be downloaded from ZENODO which opens a way for numerous computational experiments. They also may become an important support of particular medchem projects. I recommend this paper for indexation

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

F1000Research