From activity cliffs to promiscuity cliffs

First draft submitted: 26 May 2017; Accepted for publication: 2 June 2017; Published online: 27 July 2017

Keywords: activity cliffs • bioactive compounds • compound promiscuity • multitarget activities • polypharmacology • promiscuity cliffs • structure–promiscuity relationships • target hypotheses • target proteins

In medicinal and computational chemistry, the ‘activity cliff’ (AC) concept is applied to aid in the analysis of structure–activity relationships (SARs) of small molecules and the identification of SAR determinants. In addition to target-specific activities, multitarget activities of compounds, rationalized as molecular promiscuity, play an increasingly important role in drug discovery. Therefore, as an extension of the AC concept, ‘promiscuity cliffs’ (PCs) have been introduced to relate structural modifications to differences in promiscuity. Recently, PCs have been systematically identified in compound datasets from different sources, revealing surprising structure–selectivity relationships and suggesting many experimentally testable target hypotheses.

Structure–activity relationships

The study of SARs is one of the central topics in medicinal chemistry [1]. In SAR analysis, structural modifications of compounds are made to identify regions that are critical for biological activity and improve compound potency. If subsequent R-group replacements or other structural modifications lead to small-magnitude changes in potency, SARs are continuous in nature [1]. By contrast, small chemical changes causing large potency effects indicate SAR discontinuity [1], the presence of which may or may not be desirable, depending on the stages of

chemical optimization efforts. However, discontinuous SARs often arise from modulating important ligand–target interactions. Especially during early stages of compound optimization, inactive compounds might then be generated or, on the other hand, significant potency improvements might be achieved.

Activity cliffs

ACs were primarily conceptualized to capture SAR discontinuity [2]. An AC is defined as a pair of structurally similar (analogous) active compounds having a large difference in potency [2]. As such, ACs represent the apex of SAR discontinuity and often reveal SAR determinants. In the practice of medicinal chemistry, ACs are encountered in analog series with varying frequency, depending on the underlying SAR characteristics. In addition, they have also been systematically identified and studied through large-scale mining of compound activity data [2]. On the basis of high-confidence activity data, on average, every fifth compound participates in the formation of ACs with an at least 100-fold difference in potency between cliff partners. Importantly, groups of structural analogs often form multiple and overlapping ACs [2]. These ‘coordinated’ cliffs represent the vast majority of available ACs [2,3]. Currently, ACs are found in compound activity classes covering more than 300 pharmaceutical targets [3].

Jürgen Bajorath^{*1}

¹Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology & Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Dahlmannstr. 2, D-53113 Bonn, Germany

*Author for correspondence:

Tel.: +49 228 2699 306

Fax: +49 228 2699 341

bajorath@bit.uni-bonn.de

FUTURE
SCIENCE

part of

fsg

For a consistent assessment of ACs, potency difference and similarity criteria must be specified. While a variety of similarity measures are applicable, our preferred similarity criterion for ACs is the formation of transformation size-restricted matched molecular pairs (tsr-MMPs) [4]. An MMP is defined as a pair of compounds that are only distinguished by a structural change at a single site [5], corresponding to the exchange of two substructures, termed a (chemical) transformation [6]. Transformation size restrictions are applied to limit MMP compounds to typical analogs from medicinal chemistry [4]. Hence, in addition to AC assessment, a systematic search for tsr-MMPs is also our primary approach for identifying analog series in large compound datasets.

Rationalizing molecular promiscuity

By definition, ACs capture large potency differences between analogs sharing the same activity. In addition to single-target activity, multitarget activities of small molecules can be investigated. Compound promiscuity is often associated with nonspecific binding and assay artifacts resulting from colloidal aggregation or other interference effects [7,8]. However, an alternative – and scientifically more appropriate – definition of molecular promiscuity is the ability of compounds to specifically interact with multiple targets [9], as opposed to nonspecific binding or assay artifacts. This view of promiscuity is applied in the following: while single-target activity is consistent with the specificity paradigm in medicinal chemistry and drug discovery, multitarget activities are relevant for chemogenomics [10] and provide the molecular basis of polypharmacology [11,12], another important paradigm in drug discovery. Drugs and other bioactive compounds frequently engage multiple targets and the resulting polypharmacology is responsible for efficacy and also undesired side effects [12].

Computational promiscuity analysis

Unprecedented growth in volumes of compounds and activity data from medicinal chemistry [13] and biological screening [14] has provided an excellent basis for the computational identification and analysis of promiscuous compounds [15]. For database compounds, all unique targets they are reported to be active against are assembled and recorded. This target profile yields the ‘promiscuity degree’ (PD) of a molecule (i.e., the number of its unique targets). To limit false-positive target assignments, careful data curation is essential and it is strongly advisable to focus promiscuity analysis on high-confidence activity data [15]. Computational promiscuity analysis has yielded some unexpected findings. For example, the majority of recently identified promiscuous

compounds were active in the sub- μM range against two to five proteins from the same family and comparably potent against these targets [16]. Thus, the situation that a promiscuous compound might be strongly potent against a primary target and weakly potent against others, as one might intuitively expect, was only rarely observed [16].

Promiscuity cliffs

A key question arising from promiscuity analysis, which goes beyond statistics and is only beginning to be addressed, is which structural/chemical features might be responsible for compound promiscuity. Hence, just as much as one aims to rationalize SARs, one would like to explore and better understand structure–promiscuity relationships; a new research topic. For this purpose, PCs have been introduced [17], as an extension of the AC concept. PCs are defined as pairs or groups of structurally analogous compounds that display large PD differences. Thus, departing from – and further extending – the AC concept, our current PD definition considers single- and multi-target activities of compounds and also inactivity, if this information is available, but it does not include potency differences, for reasons described above [16]. PCs were first generated to aid in the analysis of compound array experiments [17] under the condition of which very large differences in apparent promiscuity of analogs were observed. Such examples were highlighted using PCs [17].

As a similarity criterion for PCs, we also require the formation of tsr-MMPs, in analogy to ACs. Thus, PCs encode small structural modifications that can be directly associated with large changes in promiscuity, in cases where causal relationships exist. Exemplary PCs are shown in Figure 1.

Recently, two studies have been carried out to systematically identify PCs in compounds from medicinal chemistry [18] and biological screening [19]. For these analyses, it has been important to eliminate compounds that are prone to activity/assay artifacts [7,8] to the extent possible and concentrate on reliable experimental data [15].

For the analysis of compounds from the medicinal chemistry literature available in ChEMBL [13], a PD difference of at least 10 (targets) was applied as a PC criterion, in addition to the requirement that compounds for which high-confidence measurements were available formed a tsr-MMP [18]. Since ChEMBL exclusively contains active compounds, a highly potent PC compound was required to be active against at least 11 different targets (yielding a PC with compound PDs of 1 and 11, respectively). Under these conditions, 784 PCs were identified in ChEMBL (release 22; excluding all screening compounds incorporated

Figure 1. Promiscuity cliffs. On the right, a cluster from the global promiscuity cliff (PC) network of compounds extensively tested in primary PubChem assays is shown. In the network, compounds are represented as nodes that are connected by edges if they form a PC. As a cliff criterion, a PD difference of at least 20 was required. Nodes are colored by PD values ranging from 0 (consistently inactive compounds) to ≥ 20 (compounds with activity against 20 or more targets). In the cluster, a pathway comprising seven compounds (1–7) and six PCs is traced. On the left, the pathway is displayed in detail. Color-coded substitution sites in compounds indicate chemical modifications that distinguish pairs of analogs forming PCs.

PD: Promiscuity degree.

from PubChem [14]). These PCs involved 77 highly promiscuous compounds (with a maximal PD of 38) and 763 qualifying analogs. The targets of 42 of the 75 highly promiscuous compounds belonged to different protein families.

For ChEMBL compounds, no assay frequency and inactivity records are available. Therefore, PD differences encoded by PCs might be influenced by differences in test frequency of cliff compounds and ensuing data sparseness. Accordingly, these PCs primarily provide suggestions for additional targets of structural analogs of highly promiscuous compounds. Indeed, PCs extracted from ChEMBL revealed many new target hypotheses [18].

Assay frequency and inactivity records can be obtained for screening compounds originating from PubChem [14], provided assay data are thoroughly analyzed outside the database environment. In a subsequent investigation [19], compounds tested in both primary (single compound concentration) and confirmatory (dose–response) screening assays were retrieved from PubChem and their activity data were curated. Then, test frequency and activity statistics were determined and PCs formed by extensively tested compounds identified. In this case, a PD difference of at least 20 was required as a more stringent PC criterion and consistently inactive compounds (PD = 0) were taken into account.

PCs were extracted from 437,257 extensively tested PubChem compounds with a mean and median value of 411 and 437 assays per compound, respectively. For primary assays, a total of 2070 PCs (involving 2158 compounds) were obtained and for confirmatory assays, 282 PCs (318 compounds). These PCs included 1024 (primary assays) and 82 (confirmatory) promiscuous compounds, 850 (primary) and 70 (confirmatory) of which were active against multiple targets from different families.

PCs can be viewed in networks in which nodes represent compounds and edges PCs (akin to AC networks). Such representations make it possible to view PCs in context and examine possible coordination and cluster formation. In Figure 1, an exemplary cluster from the PC network for primary assays is shown. In this cluster, a PC pathway is delineated and shown in detail. This pathway includes a sequence of coordinated PCs formed by structural analogs that were either consistently inactive in all assays they were tested in or active in assays against at least 20 different targets. Such PCs provide attractive starting points for experimental follow-up studies to further explore promiscuity patterns.

Importantly, 75% of the compounds from primary assays forming PCs were each tested in more than 300 assays and 83% of the PC compounds from confirmatory assays in more than 50 assays. Moreover, when only compounds were considered that were tested in at least 400 primary or 100 confirmatory assays, 713 (primary) and 142 (confirmatory) PCs were obtained. In addition, using newly introduced numerical indices, PCs were prioritized for which both partner compounds were tested in comparably large numbers of assays with at least 85% assay overlap [19].

Thus, a large knowledge base of PCs was obtained on the basis of extensively assayed compounds. In this case, a statistically significant influence of differences in assay frequency or lack of assay overlap on PC formation was ruled out. Accordingly, these PCs can be investigated with a high level of confidence, not only for deriving new target hypotheses for compounds, but also for exploring structure–promiscuity relationships, which may also lead to the design of new analogs to probe structural features that might render compounds promiscuous. To these ends, prioritized PCs and their target annotations have been made freely available [19].

Taken together, the findings discussed herein indicate that the PC concept should be rather useful to further study multitarget activities of compounds and possible structural origins of promiscuity. In addition, PCs provide excellent test cases for comparing polypharmacological features of selected analogs with differences in promiscuity involving targets of therapeutic interest.

Second-generation PCs might also include potency difference information, phylogenetic distances between targets and/or functional annotations, which are subject to further investigations.

Acknowledgements

The author thanks Dilyana Dimova for help with illustrations.

Financial & competing interests disclosure

The author has no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Creative Commons Attribution 4.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

References

Papers of special note have been highlighted as:

• of interest

- 1 Wassermann AM, Wawer M, Bajorath J. Activity landscape representations for structure-activity relationship analysis. *J. Med. Chem.* 53(23), 8209–8223 (2010).
- 2 Stumpfe D, Hu Y, Dimova D, Bajorath J. Recent progress in understanding activity cliffs and their utility in medicinal chemistry. *J. Med. Chem.* 57(1), 18–28 (2014).
- **Recent review of activity cliff research.**
- 3 Stumpfe D, Bajorath J. Monitoring global growth of activity cliff information over time and assessing activity cliff frequencies and distributions. *Future Med. Chem.* 7(12), 1565–1579 (2015).
- 4 Hu X, Hu Y, Vogt M, Stumpfe D, Bajorath J. MMP-cliffs: systematic identification of activity cliffs on the basis of matched molecular pairs. *J. Chem. Inf. Model.* 52(5), 1138–1145 (2012).
- 5 Griffen E, Leach AG, Robb GR, Warner DJ. Matched molecular pairs as a medicinal chemistry tool. *J. Med. Chem.* 54(22), 7739–7750 (2014).
- **Thorough review of the matched molecular pair concept and applications.**
- 6 Hussain J, Rea C. Computationally efficient algorithm to identify matched molecular pairs (MMPs) in large data sets. *J. Chem. Inf. Model.* 50(3), 339–348 (2010).
- 7 Shoichet BK. Screening in a spirit haunted world. *Drug Discovery Today* 11(13–14), 607–615 (2006).
- 8 Baell J, Walters MA. Chemical con artists foil drug discovery. *Nature* 513(7519), 481–483 (2014).
- 9 Hu Y, Bajorath J. Compound promiscuity: what can we learn from current data? *Drug Discov. Today* 18(13), 644–650 (2013).
- 10 Bredel M, Jacoby E. Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat. Rev. Genet.* 5(4), 262–275 (2004).
- 11 Paolini GV, Shapland RH, van Hoorn WP, Mason JS, Hopkins AL. Global mapping of pharmacological space. *Nat. Biotechnol.* 24(7), 805–815 (2006).
- **Key study providing a foundation of polypharmacology.**
- 12 Boran AD, Iyengar R. Systems approaches to polypharmacology and drug discovery. *Curr. Opin. Drug Discov. Devel.* 13(3), 297–309 (2010).
- 13 Gaulton A, Hersey A, Nowotka M *et al.* The ChEMBL database in 2017. *Nucleic Acids Res.* 45(D1), D945–D954 (2017).
- **Major public source of compounds and activity data from medicinal chemistry.**
- 14 Wang Y, Suzek T, Zhang J *et al.* PubChem BioAssay: 2014 update. *Nucleic Acids Res.* 42(D1), D1075–D1082 (2014).
- **Major public source of biological screening data.**
- 15 Hu Y, Bajorath J. Entering the ‘big data’ era in medicinal chemistry: molecular promiscuity analysis revisited. *Future Sci. OA* 3(2), FSO179 (2017).
- 16 Hu Y, Bajorath J. Promiscuity profiles of bioactive compounds: potency range and difference distributions and the relation to target numbers and families. *Med. Chem. Commun.* 4(8), 1196–1201 (2013).
- 17 Dimova D, Hu Y, Bajorath J. Matched molecular pair analysis of small molecule microarray data identified promiscuity cliffs and identifies molecular origins of extreme compound promiscuity. *J. Med. Chem.* 55(22), 10220–10228 (2012).
- **Introduction of promiscuity cliffs.**
- 18 Dimova D, Gilberg E, Bajorath J. Identification and analysis of promiscuity cliffs formed by bioactive compounds and experimental implications. *RSC Adv.* 7, 58–66 (2017).
- **Promiscuity cliffs formed by compounds from medicinal chemistry.**
- 19 Hu Y, Jasial S, Gilberg E, Bajorath J. Structure-promiscuity relationship puzzles – extensively assayed analogs with large differences in target annotations. *AAPS J.* 19(3), 856–864 (2017).
- **Promiscuity cliffs from screening compounds taking assay frequency and inactivity records into account, revealing many surprising examples.**