Systematic identification of activity cliffs with dual-atom replacements and their rationalization on the basis of single-atom replacement analogs and X-ray structures

Huabin Hu | Jürgen Bajorath

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

Correspondence
Jürgen Bajorath, Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Friedrich-Hirzebruch-Allee 6, D-53115 Bonn, Germany.
Email: bajorath@bit.uni-bonn.de

Abstract
Very small chemical changes in active compounds causing large potency effects are of particular interest in medicinal chemistry and drug design. We have systematically searched active compounds with available high-confidence activity data for pairs of structural analogs with dual-atom replacements and additional analogs with corresponding single-atom replacements. From ~287,000 unique qualifying compounds with activity against nearly 1900 unique targets, ~3500 target-based analog pairs with dual-atom replacements were identified. These included 852 pairs with significant differences in compound potency, representing a set of previously unobserved activity cliffs. Comparing these pairs with corresponding single-atom replacement analogs, which were frequently identified, made it possible to systematically analyze how potency changes propagated from single- to dual-atom replacements. The analysis uncovered different potency effects and revealed that individual atom replacements were often decisive for activity cliff formation. For a limited number of activity cliffs, X-ray structures of targets in complex with cliff compounds were available, which aided in rationalizing potency alterations among analogs with single- or dual-atom replacements. The analog pairs identified herein provide a rich resource of structure-activity relationship information and attractive test cases for calibrating computational methods.

KEYWORDS
activity cliffs, analog pairs, atom modifications, potency effects, structure-activity relationships, X-ray structures

1 | INTRODUCTION

In medicinal chemistry, activity cliffs (ACs) have been introduced as pairs or groups of structurally similar compounds with large differences in potency against the same target (Cruz-Monteagudo et al., 2014; Maggiora, 2006; Stumpfe & Bajorath, 2012; Stumpfe et al., 2014, 2019a). As such, ACs uncover small chemical modifications with significant consequences for specific biological activities. Accordingly, ACs are rich in structure-activity relationship...
Huang and Bajorath

(SAR) information and may reveal SAR determinants, which explains their desirability in medicinal chemistry (Stumpfe & Bajorath, 2012; Stumpfe et al., 2014).

In the practice of medicinal chemistry, ACs may be encountered during sequential compound optimization when substituents in analog series are modified that are involved in critical ligand-target interactions. Moreover, much of our current knowledge of ACs, including their distribution across different targets and frequency of occurrence, has originated from systematic analyses of publicly available compounds and activity data (Pérez-Villanueva et al., 2015; Stumpfe & Bajorath, 2015; Stumpfe et al., 2019). To these ends, ACs were often assessed in different ways by modifying similarity (Stumpfe et al., 2014, 2019a) and potency difference criteria (Hu et al., 2019; Stumpfe et al., 2019) required for their definition. For example, ACs with single- or multi-site substitutions (Hu et al., 2012; Stumpfe et al., 2019) were systematically extracted from large compound databases on the basis of modified molecular fragmentation algorithms originally developed for different tasks (Hussain & Rea, 2010; Naveja et al., 2019).

In medicinal chemistry and drug design, minimal chemical modifications leading to large potency effects are of particular interest since they frequently indicate most critical ligand-target interactions and/or enable the generation of highly potent lead compounds by chemical fine-tuning (Leung et al., 2012; Pennington et al., 2020; Pennington & Moustakas, 2017). ACs can capture such SAR determinants very well. Specifically, ACs representing smallest possible chemical modifications include so-called chirality cliffs, which are caused by isomerism at a single stereo center (Hu & Bajorath, 2012; Schneider et al., 2018), as well as ACs with single-atom modifications such as replacements of individual non-hydrogen atoms or positional changes of a heteroatom in a compound (atom walks) (Hu & Bajorath, 2020). Such ACs are observed for different compound classes and pharmaceutical targets. For example, a systematic search in the ChEMBL database (Gaulton et al., 2012), the major public repository of compounds and activity data from medicinal chemistry, identified ~1500 ACs comprising pairs of analogs with single-atom modifications having an at least 100-fold difference in potency, which were widely distributed across different targets (Hu & Bajorath, 2020).

Given the SAR relevance of single-atom modifications, we have been interested in exploring how frequently ACs may be formed by compounds with dual-atom replacements and how compound potency changes might propagate from analogs with corresponding single- to dual-atom replacements. Therefore, we have systematically searched for analogs pairs (APs) distinguished by dual-atom replacements and corresponding single-atom replacement analogs, analyzed their potency differences, identified ACs formed by such APs, and attempted to rationalize their formation whenever possible. The analysis led to the identification of a set of previously undetected ACs providing additional opportunities for SAR exploration.

2 | MATERIALS AND METHODS

2.1 | Compounds and activity data

Compounds with available high-confidence activity data were extracted from ChEMBL (version 27) (Gaulton et al., 2012) and organized in different target-based activity classes. Only compounds forming direct interactions (assay relationship type: “D”) with human targets at the highest level of assay confidence (assay confidence score: 9) were considered. Two different types of potency measurements, that is, numerically defined equilibrium constants ($K_i$ values) and IC$_{50}$ measurements were collected and separately analyzed. Multiple potency values available for a compound active against a given target were only used if they fell into the same order of magnitude. Otherwise, the compound was discarded, leading to the elimination of a total of 12,557 compounds (~4% of all compounds). For qualifying compounds with multiple potency values, they were averaged to yield a final $K_i$ or IC$_{50}$ potency annotation.

On the basis of these selection criteria, a total of 217,007 unique compounds with IC$_{50}$-based activity against 1795 targets were obtained as well as 82,769 unique compounds with $K_i$-based activity against 954 targets. There was limited compound overlap between these sets. Their combination yielded ~287,000 unique compounds. The resulting target- and measurement-based activity classes were used as a compound pool to search for pairs of structural analogs (analog pairs, APs) distinguished by dual-atom replacements.

2.2 | Analog pairs with dual-atom replacements and single-atom replacement analogs

APs with dual-atom replacements (dual-atom APs) in rings and aliphatic moieties and corresponding single-atom replacement analogs (single-atom analogs) were systematically identified with the aid of RDKit (http://www.rdkit.org/). Several combinations of atom replacements were considered including two nitrogen-to-carbon replacements, (N-C & N-C), a nitrogen-to-carbon plus oxygen-to-carbon replacement (N-C & O-C), two oxygen-to-carbon (O-C & O-C), and two nitrogen-to-oxygen (N-O & N-O) replacements. Typical small R-group replacements were not considered, which precluded the inclusion of halogen atoms. For each dual-atom AP, a search was carried out
for two single-atom analogs capturing the individual atom replacements. Figure 1a shows an exemplary dual-atom AP and corresponding single-atom analogs.

2.3 | Classification of analog pairs, activity cliffs, and potency effects

Dual-atom APs were divided into two different subsets on the basis of potency differences between analogs from a pair including (i) APs with comparable potency and (ii) APs capturing a potency difference of at least one order of magnitude (i.e., 10-fold). APs forming the latter subset were designated dual-atom ACs.

For dual-atom ACs for which two single-atom analogs were available, potency values and differences were compared. If the potency difference between the two analogs forming a dual-atom AC was attributable to a single-atom replacement (represented by a single-atom analog), the AC was classified as redundant. For the determination of redundant ACs, potency comparisons are not compound order-dependent (only absolute potency differences are considered).

For the subset of non-redundant dual-atom ACs with two available single-site analogs, the observed potency difference between the two AC compounds was compared to the sum of potency differences between the weakly potent AC compound and the two single-atom analogs. The classification criteria are based on the following conditions:

1. Additive effect: \( \Delta p_{POT} - \Delta p_{POT} \times 10\% \leq \Delta p_{POT1} + \Delta p_{POT2} \leq \Delta p_{POT} + \Delta p_{POT} \times 10\% \)
2. Synergistic effect: \( \Delta p_{POT1} + \Delta p_{POT2} < \Delta p_{POT} - \Delta p_{POT} \times 10\% \)
3. Compensatory effect: \( \Delta p_{POT1} + \Delta p_{POT2} > \Delta p_{POT} + \Delta p_{POT} \times 10\% \)

FIGURE 1   Exemplary dual-atom analog pair and classification. In (a), an exemplary dual-atom analog pair (N-C & O-C) is shown (top) for which two single-atom analogs were available (bottom). Atom modifications are colored red. In (b), a four-compound data structure is introduced including a dual-atom analog pair and its single-atom analogs, which is used to assess different potency effects, as detailed in the text [Colour figure can be viewed at wileyonlinelibrary.com]
analogs. To avoid boundary effects and account for experimental variance, a tolerance interval of $+/- 10\%$ of observed potency differences was applied and differences were only considered significant if they fell outside the tolerance interval. Three different potency effects were distinguished by comparing the sum of single-atom replacement contributions to the potency difference calculated for a dual-atom AC:

(i) Additive effect: the sum of potency increases of single-atom replacements was comparable to the potency difference between the dual-atom AC compounds.

(ii) Synergistic effect: the sum of potency increases was smaller than the potency difference captured by the dual-atom AC.

(iii) Compensatory effect: the sum of potency increases was larger than the potency difference of the dual-atom AC.

The analysis scheme was previously applied in the characterization of ACs with R-group replacements at multiple sites (Stumpfe et al., 2019). Figure 1b illustrates the classification scheme.

2.4 | X-ray structures of complexes with compounds from dual-atom activity cliffs

To aid in the rationalization of potency differences associated with dual-atom ACs, we searched the RCSB Protein Data Bank (PDB) (Berman et al., 2000) for X-ray structures of complexes between dual-atom AC compounds and their targets. Initially, ChEMBL target identifiers (IDs) were mapped to UniProt IDs (UniProt Consortium, 2015) to identify all available PDB entries for each dual-atom AC target. Then, crystallographic ligands were searched for dual-atom AC compounds and single-atom analogs. The search calculations were performed using KNIME protocols (Berthold et al., 2008). Ligand-target interactions in relevant complex structures were visualized and analyzed with the Molecular Operating Environment (MOE) (https://www.chemcomp.com/).

3 | RESULTS AND DISCUSSION

3.1 | Dual-atom analog pairs and activity cliffs

From the pool of 287,000 compounds with high-confidence activity data, a total of 3525 dual-atom APs were extracted.

---

### Table 1: Target families covering more than 10 dual-atom activity cliffs

| Target families                      | No. dual-atom ACs | No. AC targets |
|--------------------------------------|-------------------|---------------|
| G-protein coupled receptor 1 family  | 227               | 67            |
| Tyr protein kinase family            | 80                | 23            |
| CMGC Ser/Thr protein kinase family   | 49                | 11            |
| Cyclic nucleotide phosphodiesterase family | 32            | 3             |
| Peptidase S1 family                  | 25                | 7             |
| Transient receptor (TC 1.A.4) family| 25                | 1             |
| Nuclear hormone receptor family      | 23                | 10            |
| Sugar transporter (TC 2.A.1.1) family| 22                | 3             |
| Unclassified                         | 21                | 11            |
| AGC Ser/Thr protein kinase family    | 18                | 9             |
| G-protein coupled receptor 3 family  | 18                | 2             |
| Cytochrome P450 family               | 17                | 8             |
| CAMK Ser/Thr protein kinase family   | 16                | 6             |
| PDK/BCKDK protein kinase family      | 16                | 1             |
| P2X receptor family                  | 15                | 2             |
| TKL Ser/Thr protein kinase family    | 14                | 4             |
| PI3/PI4-kinase family                | 12                | 5             |
| Potassium channel family             | 12                | 2             |
| Sodium:solute symporter (SSF) (TC 2.A.21) family | 12            | 2             |

Note: Reported are the target families with more than 10 dual-atom ACs. Target families are designated according to the UniProt classification scheme. For each target family, the number of targets and ACs is given.
that were active against a total of 501 targets. About 75% of these APs consisted of compounds with comparable potency, whereas 852 dual-atom APs captured an at least 10-fold difference in potency. These APs were classified as dual-atom ACs and further analyzed. The 852 dual-atom ACs comprised a total of 1335 unique compounds with activity against a total of 272 targets. On the basis of high-confidence activity data, target annotations in addition to the primary AC target were identified for a subset of ACs. The 852 dual-atom ACs included 421 instances where both partner compounds were exclusively active against the primary target, 112 cases where one of the two compounds was annotated with one to 12 additional targets, and 319 instances where both compounds were active against one to 12 distinct targets. The 272 targets belonged 80 different UniProt families (UniProt Consortium, 2015) and included 11 non-classified targets. The target family distribution of dual-atom ACs is reported in Table 1. With 67 targets covering 227 dual-atom ACs, the G-protein coupled receptor 1 family was mostly frequently found, followed by the tyrosine kinase family with 23 targets of 80 ACs.

In addition, the topological scaffold diversity of the 1335 AC compounds was analyzed by extracting cyclic skeletons (CSKs) from compounds. In CSKs, heteroatoms in scaffolds are replaced with carbon atoms and bond orders are set to one. Atom replacements in rings formally define new scaffolds, but compounds from a dual-atom AC are represented by the same CSK. The AC compounds yielded a total of 405 CSKs, corresponding to a compound-to-scaffold ratio of only 3.3:1, hence indicating a high degree of topological diversity. Table 2 reports the top 11 CSKs representing most dual-atom AC compounds (with 11 to 22 compounds per skeleton).

As shown in Figure 2a, (N-C & N-C) ACs clearly dominated the distribution, with 589 instances, followed by (N-C & O-C) ACs (227 instances). By contrast, ACs with (O-C & O-C) and (N-O & N-O) replacement combinations were only rarely detected. The 852 dual-atom ACs contained 169 instances capturing an at least 100-fold potency difference. Figure 2b shows representative dual-atom ACs. Notably, dual-atom ACs were not identified thus far. There was no overlap between the set of dual-atom ACs and ~3800 ACs containing two R-group replacements, as identified previously on the basis of systematic compound fragmentation (Stumpf et al., 2019).

### 3.2 Dual-atom activity cliffs with single-atom analogs

For each dual-atom AC, a search for single-atom analogs was carried out. For 622 ACs (73%), at least one analog with corresponding single-atom replacement was detected including 192 ACs for which both single-atom analogs were available (Figure 3). Hence, a large number of single-atom analogs was identified, which made it possible to

### Table 2 Carbon skeletons from dual-atom activity cliff compounds

| Carbon skeletons | No. AC compounds |
|------------------|------------------|
| ![Carbon skeleton](image1) | 22 |
| ![Carbon skeleton](image2) | 20 |
| ![Carbon skeleton](image3) | 18 |
| ![Carbon skeleton](image4) | 12 |
| ![Carbon skeleton](image5) | 11 |
| ![Carbon skeleton](image6) | 11 |
| ![Carbon skeleton](image7) | 11 |
| ![Carbon skeleton](image8) | 11 |
| ![Carbon skeleton](image9) | 11 |

Note: Eleven CSKs representing most (more than 10) dual-atom AC compounds are presented.
analyze potency effects associated with the formation of many dual-atom ACs in detail.

Figure 4 shows an exemplary dual-atom AC capturing a 1000-fold difference in potency for which both single-atom analogs are available and illustrates a data structure used for the analysis of AC-associated potency differences (Stumpf et al., 2019). Here, the two dual-atom AC compounds on the left and right form an analog quartet with the two single-atom analogs, which enables the assessment of potency differences resulting from all individual atom replacements.

We first assessed the potential redundancy of dual-atom ACs on the basis of available single-atom analogs. A dual-atom AC was considered redundant if the associated potency difference was essentially caused by one of the dual-atom replacements. Figure 5 shows an exemplary redundant AC.

The 430 dual-atom ACs with only one available single-atom analog included 213 redundant instances where AC formation was essentially due to the single-atom replacement represented by the available analog. Moreover, the 192 dual-atom ACs for which both single-atom analogs were available included 115 redundant and 77 non-redundant ACs. Hence, the formation of nearly 60% of these ACs originated from only one of the dual-atom replacements, which was an unexpected finding. In addition, we detected 171 redundant dual-atom ACs where a single-atom analog had higher potency than the highly potent dual-atom AC compound. Hence, there was substantial SAR heterogeneity within this AC population and
in many instances, individual atom replacements were responsible for AC formation.

### 3.3 Differential potency effects

Using the subset of 77 non-redundant dual-atom ACs for which both single-atom analogs were available, we next investigated the potency effects accompanying cliff formation. The analysis revealed 22 additive effects as well as 21 synergistic and 34 compensatory effects of varying magnitude (Figure 6). Thus, for non-redundant dual-atom ACs where both atom replacements contributed to AC formation, there were a variety of potency effects reflecting the presence of different SAR characteristics.

Figure 7 illustrates different potency effects. In Figure 7a, an additive effect of individual atom replacements is observed leading to the formation of the dual-atom AC.
In Figure 8a, an AC formed by thrombin inhibitors is shown for which only one single-atom analog was available. This single-atom analog was nearly as potent as the highly potent AC compound. In this case, a complex structure containing this single-atom analog was available. Interaction analysis revealed the presence of two hydrogen bonds including an intramolecular hydrogen-bonding contact in the ligand between its piperidinium and pyridine nitrogen atoms as well as a hydrogen bond between the ligand and thrombin residue Gly216. These hydrogen-bonding interactions apparently stabilized the bound conformation of the oxazolo[4,5-c]pyridine moiety and positioned it for π-stacking interactions with residue Tyr60A. The highly potent cliff compound was also capable of forming the intramolecular hydrogen bond. By contrast, this hydrogen bond could not be formed by the weakly potent AC compound, given its (N-C & N-C) replacements, which would likely destabilize the binding mode. To our knowledge, this is one of the first (if not the first) example of an AC where an intramolecular interaction in the potent cliff compound may be critical for AC formation.

Figure 8b shows a rare example of a dual-atom AC for which X-ray structures with both AC compounds were available. This AC was formed by phosphodiesterase 10A inhibitors. Comparison of these structures revealed similar binding modes of the two AC compounds and several conserved interactions in the active site: (1) the pyridine or pyrazine ring of the inhibitors formed π-stacking interactions with residue Phe719; (2) the pyridine nitrogen or one of pyrazine nitrogen atoms were involved in hydrogen-bonding interactions with residue Gln716; (3) the planar bound conformation of the phenyl keto-benzimidazole moiety was stabilized by a hydrogen-bonding interaction between one of the nitrogens on the benzimidazole ring and Tyr683. The ~35-fold higher potency of the AC compound with (N-C & N-C) replacements was likely at least partly attributable to additional hydrogen-π interactions with Phe719, which further improved the fit of the compound into the binding site compared to the less potent analog (as also indicated by minor conformational differences between the bound conformations). This possible explanation was consistent with the intermediate potency of both single-atom analogs with one of the two N-C substitutions and the presence of a minor synergistic potency effect.

**3.4 Structure-based analysis**

For 34 dual-atom ACs (distinct from the 34 compensatory ACs discussed above), one or more X-ray structures of protein complexes with AC compounds or single-atom analogs were identified, which included a total of 20 different targets. For 11 of these ACs (covering eight targets), both single-atom analogs were available. We analyzed ligand-target interactions in these structures attempting to formulate hypotheses for rationalizing AC formation or associated potency effects. Although only structures with one AC compound or single-atom analog were available for most ACs, the formation of ACs could be rationalized in some cases. Figure 8 shows exemplary dual-atom ACs and corresponding X-ray structures.

**4 CONCLUSIONS**

In this work, we have systematically analyzed APs with dual-atom replacements and corresponding single-atom
Figure 7 Different potency effects. Exemplary dual-atom ACs with two available single-atom analogs are shown to illustrate different potency effects accompanying AC formation including an (a) additive, (b) compensatory, and (c) synergistic effect [Colour figure can be viewed at wileyonlinelibrary.com]
FIGURE 8  X-ray structures with dual-atom activity cliff compounds or single-atom analogs. In (a), an AC formed by thrombin inhibitors is shown for which an X-ray structure with the single-atom analog was found. For this dual-atom AC, only one single-atom analog was available, which helped to rationalize AC formation. In the bound ligand, an intra-molecular interaction was formed between the piperidinium nitrogen and the aminopyridine ring nitrogen atom. In (b), an AC of phosphodiesterase 10A inhibitors with a synergistic potency effect is depicted for which X-ray structures with both AC compounds were available. Their binding modes are aligned at the bottom and minor conformational differences are encircled. X-ray structures are represented by coloring ligand and protein carbon atoms cyan and orange, respectively. A surface representation of the binding site is shown and residues involved in apparent key interactions are labeled. Dashed green and magenta lines indicate the formation of hydrogen bond and hydrogen-π interactions, respectively. For each X-ray structure, the PDB code is provided [Colour figure can be viewed at wileyonlinelibrary.com]
analogs. The analysis led to the identification of previously unconsidered ACs and revealed the presence of varying atom contributions and differential potency effects associated with AC formation. The AC-based analog ensembles reported herein are rich in SAR information and should be of interest for medicinal chemistry projects. In addition, they provide excellent test cases for evaluating computational methods. This is the case because the analog ensembles represent well-defined combinations of subtle chemical modifications causing significant potency alterations and different types of potency effects. To these ends, they are especially useful in combination with available X-ray structures of AC compounds or closely related analogs. For example, crystallographic and/or modeled complexes with these analogs enable benchmarking and fine-tuning of docking approaches and scoring functions as well as binding energy methods including free energy perturbation calculations. All dual-atom ACs reported herein and associated target data are made freely available in an open access deposition (http://doi.org/10.5281/zenodo.5634280).

ACKNOWLEDGMENTS

H.H. is supported by the China Scholarship Council (CSC). Open access funding enabled and organized by ProjektDEAL.

CONFLICT OF INTEREST

The authors state no conflict of interest and have received no payment for preparation of this manuscript.

REFERENCES

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, 28, 235–242. https://doi.org/10.1093/nar/28.1.235

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., & Wiswedel, B. (2008). KNIME: the konstanz information miner. In C. Preisach, H. Burkhart, Y., Nicolotti, O., Cordeiro, M. N. D., & Borges, F. (Eds.), *Studies in classification, data analysis, and knowledge organization* (pp. 319–326). Springer.

Cruz-Monteagudo, M., Medina-Franco, J. L., Perez-Castillo, Y., Nicolotti, O., Cordeiro, M. N. D., & Borges, F. (2014). Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? *Drug Discovery Today*, 19, 1069–1080. https://doi.org/10.1016/j.drudis.2014.02.003

Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazzikani, B., & Overington, J. P. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Research*, 40, D1100–D1107. https://doi.org/10.1093/nar/gkr777

Hu, H., & Bajorath, J. (2020). Activity cliffs produced by single-atom modifications of active compounds: systematic identification and rationalization based on x-ray structures. *European Journal of Medicinal Chemistry*, 207, 112846. https://doi.org/10.1016/j.ejmech.2020.112846

Hu, H., Stumpfe, D., & Bajorath, J. (2019). Second-generation activity cliffs identified on the basis of target set-dependent potency difference criteria. *Future Medicinal Chemistry*, 11, 379–394. https://doi.org/10.4155/fmc-2018-0299

Hu, X., Hu, Y., Vogt, M., Stumpfe, D., & Bajorath, J. (2012). MMP-cliffs: Systematic identification of activity cliffs on the basis of matched molecular pairs. *Journal of Chemical Information and Modeling*, 52, 1138–1145. https://doi.org/10.1021/ci3001138

Hu, Y., & Bajorath, J. (2012). Extending the activity cliff concept: Structural categorization of activity cliffs and systematic identification of different types of cliffs in the ChEMBL database. *Journal of Chemical Information and Modeling*, 52, 1806–1811. https://doi.org/10.1021/ci300274c

Hussain, J., & Rea, C. (2010). Computationally efficient algorithm to identify matched molecular pairs (MMPs) in large data sets. *Journal of Chemical Information and Modeling*, 50, 339–348. https://doi.org/10.1021/ci900450m

Leung, C. S., Leung, S. S., Tirado-Rives, J., & Jorgensen, W. L. (2012). Methyl effects on protein–ligand binding. *Journal of Medicinal Chemistry*, 55, 4489–4500. https://doi.org/10.1021/jm3003697

Maggiora, G. M. (2006). On outliers and activity cliffs – why QSAR often disappoints. *Journal of Chemical Information and Modeling*, 46, 1535. https://doi.org/10.1021/ci60117s

Naveja, J. J., Vogt, M., Stumpfe, D., Medina-Franco, J. L., & Bajorath, J. (2019). Systematic extraction of analogue series from large compound collections using a new computational compound-core relationship method. *ACS Omega*, 4, 1027–1032. https://doi.org/10.1021/acsofme.8b03390

Pennington, L. D., Aquila, B. M., Choi, Y., Valiulin, R. A., & Muegge, I. (2020). Positional analogue scanning: an effective strategy for multiparameter optimization in drug design. *Journal of Medicinal Chemistry*, 63, 8956–8976. https://doi.org/10.1021/acs.jmedchem.9b02092

Pennington, L. D., & Moustakas, D. T. (2017). The necessary nitrogen atom: A versatile high-impact design element for multiparameter optimization. *Journal of Medicinal Chemistry*, 60, 3552–3579. https://doi.org/10.1021/acs.jmedchem.7b00407

Pérez-Villanueva, J., Méndez-Lucio, O., Soria-Arteche, O., & Medina-Franco, J. L. (2015). Activity cliffs and activity cliff generators based on chemotype-related activity landscapes. *Molecular Diversity*, 19, 1021–1035. https://doi.org/10.1007/s11030-015-9609-z

Schneider, N., Lewis, R. A., Fechner, N., & Ertl, P. (2018). Chiral cliffs: Investigating the influence of chirality on binding affinity. *ChemMedChem*, 13, 1315–1324. https://doi.org/10.1002/cmdc.201700798

Stumpfe, D., & Bajorath, J. (2012). Exploring activity cliffs in medicinal chemistry. *Journal of Medicinal Chemistry*, 55, 2932–2942. https://doi.org/10.1021/jm201706b

Stumpfe, D., & Bajorath, J. (2015). Monitoring global growth of activity cliff information over time and assessing activity cliff frequencies and distributions. *Future Medicinal Chemistry*, 7, 1565–1579. https://doi.org/10.4155/fmc.15.89

Stumpfe, D., Hu, H., & Bajorath, J. (2019). Evolving concept of activity cliffs. *ACS Omega*, 4, 14360–14368. https://doi.org/10.1021/acsomega.9b02221

Stumpfe, D., Hu, H., & Bajorath, J. (2019). Introducing a new category of activity cliffs with chemical modifications at multiple
sites and rationalizing contributions of individual substitutions. Bioorganic & Medicinal Chemistry, 27, 3605–3612. https://doi.org/10.1016/j.bmc.2019.06.045

Stumpf, D., Hu, Y., Dimova, D., & Bajorath, J. (2014). Recent progress in understanding activity cliffs and their utility in medicinal chemistry. Journal of Medicinal Chemistry, 57, 18–28. https://doi.org/10.1021/jm401120g

UniProt Consortium (2015). UniProt: A hub for protein information. Nucleic Acids Research, 43, D204–D212. https://doi.org/10.1093/nar/gku989

How to cite this article: Hu, H., & Bajorath, J. (2022). Systematic identification of activity cliffs with dual-atom replacements and their rationalization on the basis of single-atom replacement analogs and X-ray structures. Chemical Biology & Drug Design, 99, 308–319. https://doi.org/10.1111/cbld.13985