RESEARCH ARTICLE

Check for updates

Introducing a new category of activity cliffs combining different compound similarity criteria

Huabin Hu and Jürgen Bajorath [ID] *

Activity cliffs (ACs) are pairs of structurally similar or analogous active compounds with large differences in potency against the same target. For identifying and analyzing ACs, similarity and potency difference criteria must be determined and consistently applied. This can be done in various ways, leading to different types of ACs. In this work, we introduce a new category of ACs by combining different similarity criteria, including the formation of matched molecular pairs and structural isomer relationships. A systematic computational search identified such ACs in compounds with activity against a variety of targets. In addition to other ACs exclusively formed by structural isomers, the newly introduced category of ACs is rich in structure–activity relationship (SAR) information, straightforward to interpret from a chemical perspective, and further extends the current spectrum of ACs.

Introduction

Structurally similar active compounds with large potency differences form activity cliffs (ACs).[1,2] They can be detected in analog series during chemical optimization or extracted from compound data sets. ACs reveal small chemical modifications that significantly impact biological activity and are thus of high interest in structure–activity relationship (SAR) analysis.[2,3] In the practice of medicinal chemistry, ACs might be subjectively assessed on a case-by-case basis when encountered during compound optimization efforts. However, for systematic identification and organization as well as consistent representation and evaluation of ACs, similarity and potency difference criteria must be clearly defined and consistently applied.[2,3] We note that similarity is generally considered as a subjective criterion but in chemistry and other scientific fields, different metrics and measures have been introduced to quantify similarity in reproducible ways.[4] In medicinal chemistry, this provides the foundation for establishing compound similarity relationships beyond subjective assessment and chemical intuition and enabling systematic SAR exploration.[4] For large-scale identification and analysis of ACs, computational methods play an important role.[2] The choice and combination of alternative similarity and potency difference criteria give rise to different categories of ACs having different characteristics.

Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Endenicher Allee 19c, D-53115 Bonn, Germany. E-mail: bajorath@bit.uni-bonn.de; Fax: +49 228 7369 100; Tel: +49 228 7369 100

The question if two compounds are sufficiently similar to form an AC can be addressed in different ways, for example, by calculating Tanimoto similarity on the basis of graph-based molecular representations or by applying substructure-based similarity concepts.[2–4] Substructure-based measures include, for example, the conservation of compound scaffolds,[4,5] formation of matched molecular pairs (MMPs),[6,7] or presence of analog relationships (*i.e.*, two compounds belong to the same analog series).[8,9] If numerical similarity metrics are applied, a similarity threshold for AC formation must be set, which is not only representation-dependent, but also subjective in nature.[3]

For substructure-based similarity assessment, MMPs have become increasingly popular. They are defined as pairs of compounds that only differ by a confined chemical change at a single site, which is termed a chemical transformation.[6] MMPs can be extracted from large compound collections in computationally efficient ways,[6] which supports large-scale analysis. Hence, the MMP concept can also be applied to computationally identify structural analogs[7] and series of analogs.[8] In addition to applying molecular graph-based similarity measures, ACs have also been determined on the basis of X-ray structures of ligand-target complexes.[10] This requires the calculation of three-dimensional (3D) similarity of experimentally observed compound binding modes, yielding so-called 3D-cliffs.[10]

The question when potency differences between analogs become sufficiently large to qualify a compound pair as an AC can also be addressed in different ways. For example, a constant potency difference threshold can be applied that reflects a statistically significant difference in potency across many compound data sets (activity classes, also termed

target sets).[2,3] Alternatively, target set-dependent potency difference thresholds can be determined, which take set-specific potency value distributions into account.[11] As a constant potency difference threshold, a potency difference of at least two orders of magnitude between similar candidate compounds has often been applied.[3,7] For comparison, potency difference thresholds of target set-dependent ACs frequently range from 1.5 to 2.5 orders of magnitude[11] and are thus comparable.

The application of different similarity measures and potency difference thresholds characterize subsequent generations of ACs, beginning with ACs that were defined on the basis of numerical similarity measures and constant potency difference thresholds[1,2] and leading to ACs based upon substructure-based similarity criteria and target set-dependent potency difference thresholds.[9,11]

For large scale-analysis of ACs across many different compound classes, MMP-cliffs[6] have been particularly useful,[12,13] given their computationally efficient generation and chemically intuitive nature. For MMP-cliffs, the similarity criterion that must be met by candidate compounds is the formation of a transformation size-restricted MMP and a constant potency difference of at least two orders of magnitude is required.[7]

Herein, we introduce a new category of ACs by assessing similarity in a previously unconsidered manner. For the first time, different similarity criteria are applied in combination to define ACs, leading to the identification of new ACs with high SAR information content for a variety of pharmaceutical targets.

Materials and methods

Compounds and activity data

Bioactive compounds were extracted from ChEMBL version 24.1.[14] For our analysis, the following selection criteria were applied. Only compounds with direct interactions (target relationship type "D") with human target proteins at the highest assay confidence level (ChEMBL confidence score 9) and available numerically specified equilibrium constants (K_i values) were selected. Approximate measurements such as those indicated by "<", ">" or "~" were not considered. On the basis of these criteria, a total of 73 965 unique compounds with activity against 915 targets were obtained and divided into 915 target sets.

Systematic compound fragmentation

Following the MMP fragmentation scheme,[6] exocyclic bonds in test compounds were subjected to systematic single-cut fragmentation (i.e., a single bond was cleaved per iteration), which produced two substructures (core and substituent). The following size restrictions were applied:[7] the size of the core (number of non-hydrogen atoms) was required to be at least twice the size of the substituent and the size of the substituent was limited to at most 13

non-hydrogen atoms. The fragmentation protocol was applied to systematically generate MMPs and identify structural isomers (see below).

Generation of matched molecular pairs

An MMP is defined as a pair of compounds that only differ by a chemical change at a single site. For MMP generation, the size restrictions specified above were complemented by applying an additional rule, i.e., the size difference between exchanged fragments (representing a chemical transformation) was limited to at most eight non-hydrogen atoms. The application of these rules yielded transformation size-restricted MMPs.[7]

Identification of structural isomers

Structural isomers are compounds that have the same chemical composition formula but are topologically distinct. Herein, structural isomers were identified that only differed in the core position of the substituent fragment, corresponding to sets of analogs in which the same substituent fragment occurred at different positions. To systematically identify and classify such structural isomers in target sets, generalized cores were constructed with the aid of the OpenEye Chemistry toolkit,[15] in which each attachment site of a substituent fragment was substituted with a hydrogen atom. All structural isomers originating from a target set that were represented by the same generalized core and fragment were then combined into an isomer set.

Activity cliff criteria

Three types of ACs were investigated herein, consistently requiring an at least 100-fold difference in potency between cliff compounds. First, standard MMP-cliffs were extracted from target sets. In addition, "isomers cliffs" were defined to be formed by two structural isomers from the same set, also having an at least 100-fold difference in potency. A separate search for isomer cliffs was carried out. We note that isomer cliffs, as defined herein, are related to "topology cliffs" that were reported previously applying a scaffold-based similarity criterion.[4] Furthermore, "isomer/MMP-cliffs" were introduced. As described in more detail below, in isomer/MMP-cliffs, one MMP-cliff compound was replaced by a structural isomer. Hence, searching for isomer/MMP-cliffs required combining MMP- and structural isomer-based similarity assessment. Therefore, as a pre-requisite of identifying isomer/MMP-cliffs, MMP-cliff compounds were determined that also belonged to isomer sets. The corresponding ACs were termed "MMP-cliffs with isomer extension". If multiple MMP-cliffs were found to be associated with the same isomer set, only the MMP-cliff with the largest potency difference was retained for further exploration of isomer/MMP-cliffs, thus avoiding potential AC redundancy.

Results and discussion

Systematic exploration of structural relationships

Fig. 1 illustrates different structural relationships investigated in this work. Bioactive compounds with high-confidence activity data were systematically searched for transformation size-restricted MMPs. In parallel, a search was carried out for sets of structural isomers. Then, MMP-cliff compounds were identified that also participated in isomer sets, thus combining MMPs and isomer sets for AC analysis.

Extending the current spectrum of activity cliffs

Three types of ACs investigated herein are depicted in Fig. 2. As a standard, MMP-cliffs were systematically identified. In addition, isomer sets were independently identified and searched for pairs of isomers with an at least 100-fold difference in potency, yielding isomer cliffs. In these ACs, compounds were distinguished by the position of a given substituent (resulting from molecular

fragmentation). So-called "chirality cliffs"[5] or "chiral cliffs"[16] in which compounds with large potency differences are only distinguished by the configuration at a single stereo center have been described previously.[5,16] By contrast, isomer cliffs as defined herein have not been introduced before (but -as stated above- are related to scaffold-based topology cliffs). Moreover, isomer/MMP-cliffs also shown in Fig. 2 represent a novel category of ACs. We reasoned that adding structural isomers to MMP-cliffs would further extend their SAR information content. By definition, structural analogs forming MMP-cliff are distinguished by a substitution at one and only one site. However, replacing an MMP-cliff compound by a structural isomer adds another substitution site. Hence, compounds forming an isomer/MMP-cliff are distinguished by different substituents (R-groups) at two sites, which can be accounted for following MMP terminology as H ↔ R transformations. Combining different similarity criteria is a characteristic feature of isomer/MMP-cliffs setting them apart from other AC categories.
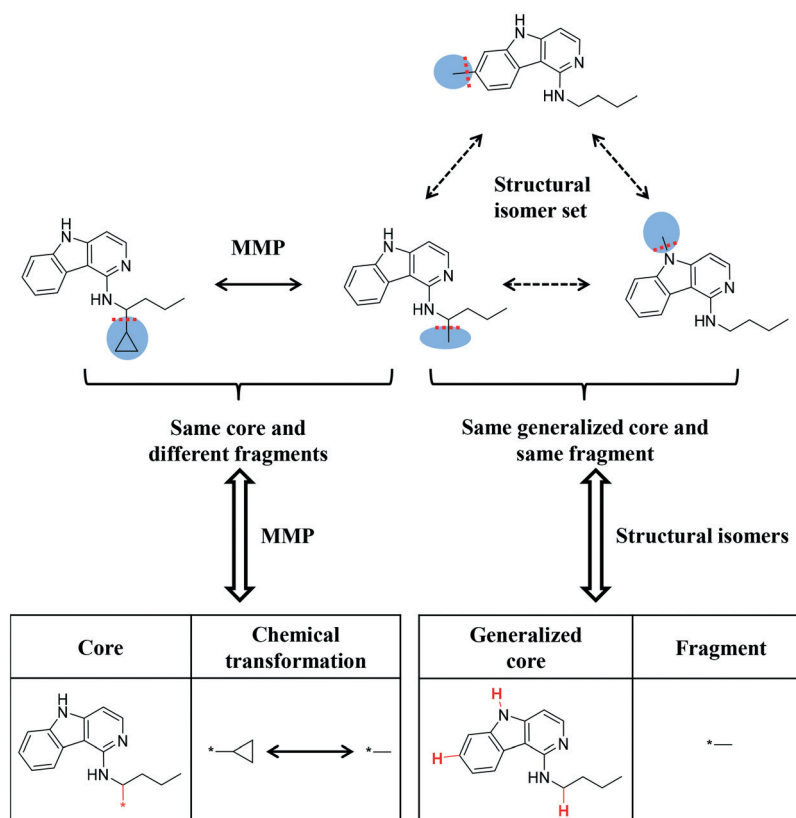

Fig. 1 Structural relationships. The schematic representation illustrates structural relationships that were systematically identified. For this purpose, a small compound (CPD) set with four analogs is used. For this exemplary set, MMP and structural isomer relationships are shown. Initially, compounds from all qualifying target sets were subjected to fragmentation of exocyclic single bonds to detect MMPs. Cleaved bonds are indicated by dashed red lines and the resulting fragments are shown on a blue background. For each MMP, the chemical transformation was recorded. The MMP fragmentation scheme was also adapted to identify structural isomers (that share the same composition formula, but are topologically distinct). Therefore, a search was carried out for structurally distinct (unique) compounds that yielded the same fragment and core of the same composition. Such compounds were represented by the same fragment and generalized core, in which fragmentation sites were hydrogen substituted (shown in red), and combined into an isomer set. Finally, MMP compounds were identified that also belonged to isomer sets, thereby combining different structural relationships.
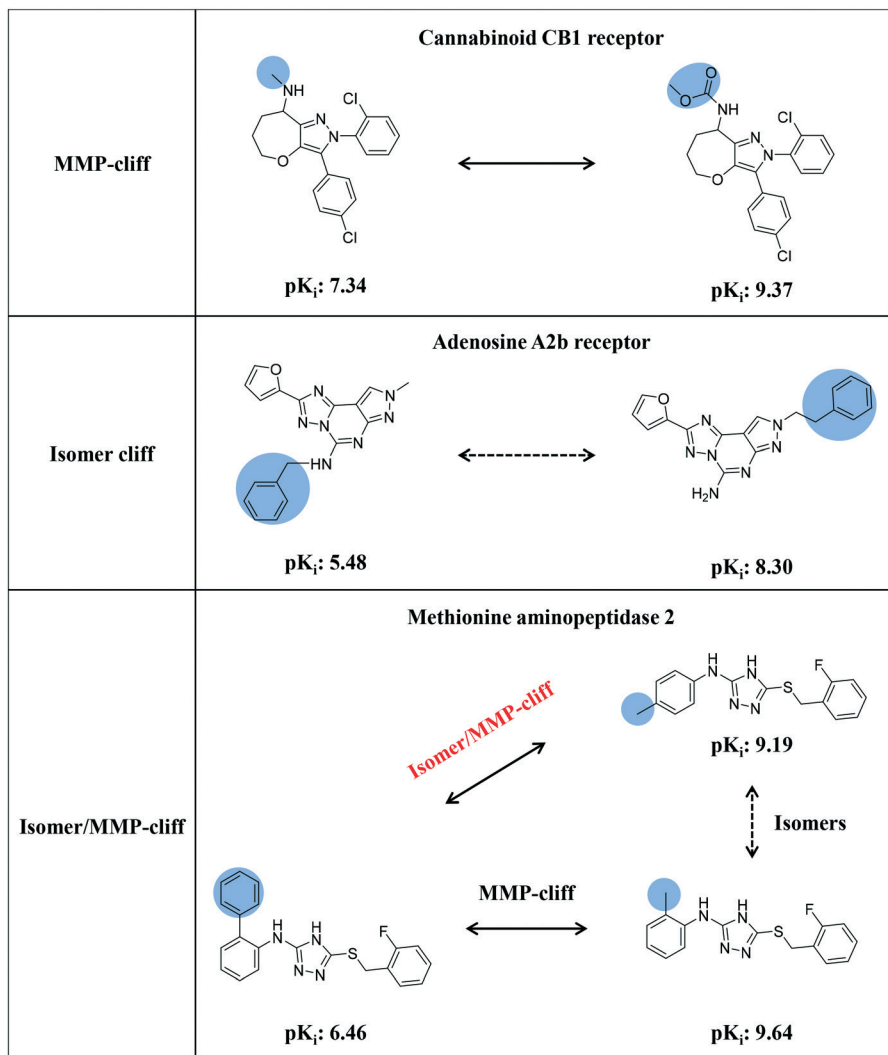
Fig. 2 Activity cliff categories. Shown are exemplary ACs belonging to different categories including (from the top to the bottom) an MMP-cliff, isomer cliff, and isomer/MMP-cliff. In each case, the target of the AC compounds is given.

Searching for activity cliffs

A systematic search for the three types of ACs was carried out in ChEMBL, as summarized in Fig. 3. From nearly 74 000 compounds with qualifying activity data for 915 targets, more than 600 000 MMPs were extracted that yielded 26 966 MMP-cliffs originating from 351 target sets. These MMP-cliffs involved 14 008 unique compounds. In addition, 10 571 different isomer sets (with the median and maximum size of two and eight isomers, respectively) were identified comprising 13 867 unique compounds from 412 target sets. These isomer sets contained 16 314 isomer pairs that yielded 493 isomer cliffs for 124 different targets. Fig. 4a shows that in only 425 (4.0%) of all isomer sets, the potency difference threshold for AC formation was met. By contrast, in 5706 isomer sets, maximal pairwise potency differences were close to zero and in more than 8000 sets, they fell within one order of magnitude. Hence, structural

isomers mostly had similar potency against their targets. Although the absolute number of isomer cliffs was much smaller than of MMP-cliffs, the percentage of isomer cliffs among isomer pairs (3.0%) was comparable to the proportion of MMP-cliffs among MMPs (4.4%). Fig. 4b shows two isomer sets with three isomers each in which isomer cliffs were formed.

Next, we searched for MMP-cliffs with isomer extension, *i.e.*, MMP-cliff compounds that also belonged to isomer sets. For 1182 MMP-cliffs (4.4%) originating from 147 target sets, structural isomers were identified, as reported in Fig. 3. From these MMP-cliffs with isomer extension, a total of 597 isomer/MMP-cliffs were extracted, which consisted of 636 unique compounds with activity against 80 different targets. Thus, 39.8% of MMP-cliffs with isomer extension represented isomer/MMP cliffs and provided informative AC constellations for further analysis, as discussed below. First, we take a closer look at MMP-
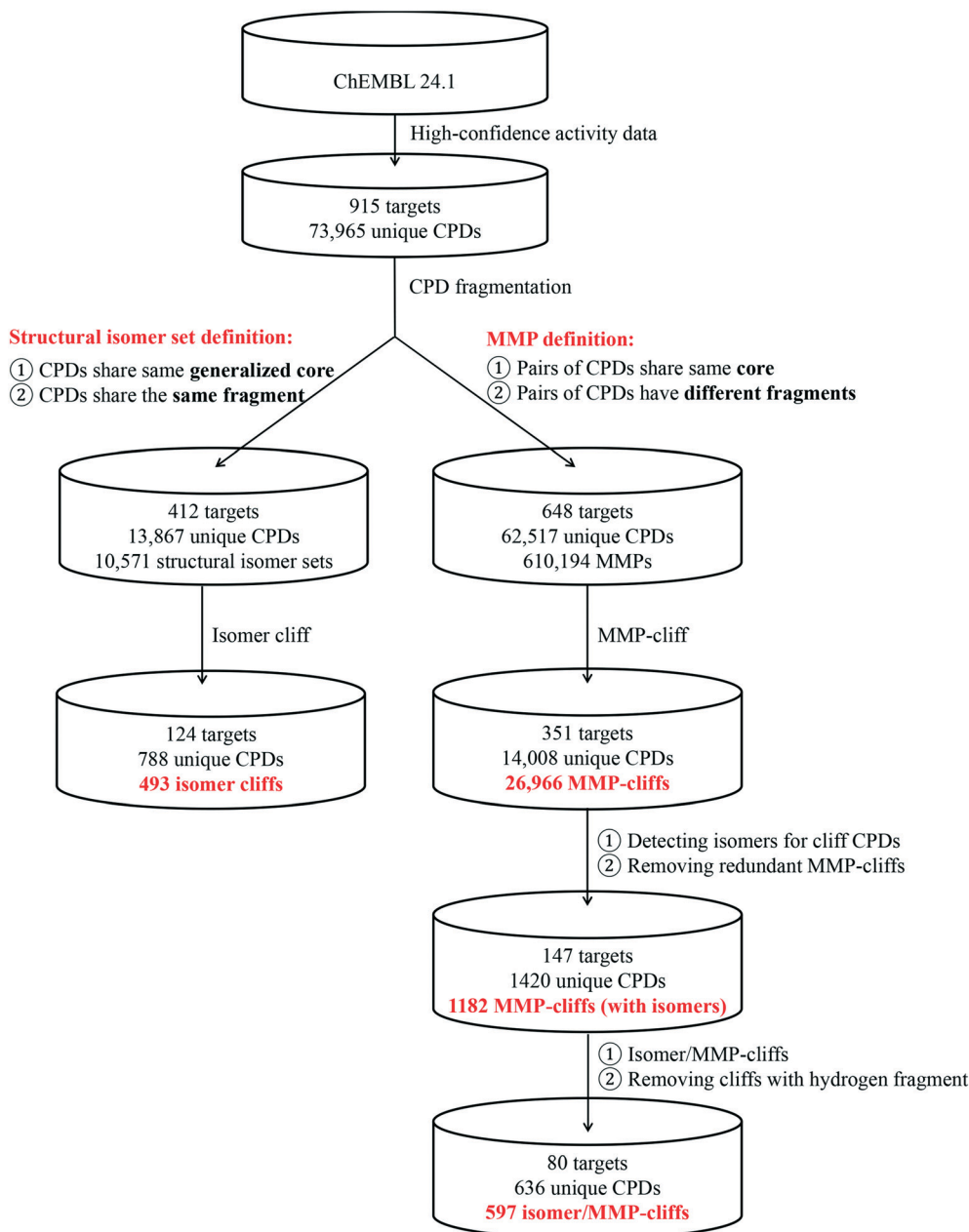
Fig. 3 Identification of different activity cliffs. The workflow chart summarizes the identification of MMP-cliffs, isomer cliffs, MMP-cliffs with isomer extension, and isomer/MMP-cliffs across different target sets.

cliffs with isomer extension and the chemical transformations they contained.

#### Chemical transformations in extended MMP-cliffs

We systematically analyzed chemical transformations associated with MMP-cliffs. For the 1182 MMP-cliffs with isomer extension, 676 unique chemical transformations were detected. Interestingly, small transformations involving hydrogen atom replacements were among the most frequently observed chemical changes in MMP-cliffs with isomer extension, as shown in Fig. 5a. The replacement of a

hydrogen atom by a methyl group occurred most frequently (and with similar frequency as in all MMP-cliffs), followed by hydrogen replacements with a methoxy group and chlorine atom, respectively. For extended MMP-cliffs with hydrogen atom replacements, broad distributions of maximum potency differences between MMP-cliff compounds and structural isomers were observed, as reported in Fig. 5b, often with median values around two orders of magnitude.

Fig. 6a shows exemplary MMP-cliffs with an H $\leftrightarrow$ CH_3 transformation for which structural isomers of weakly potent cliff compounds were available. These extended MMP-cliffs
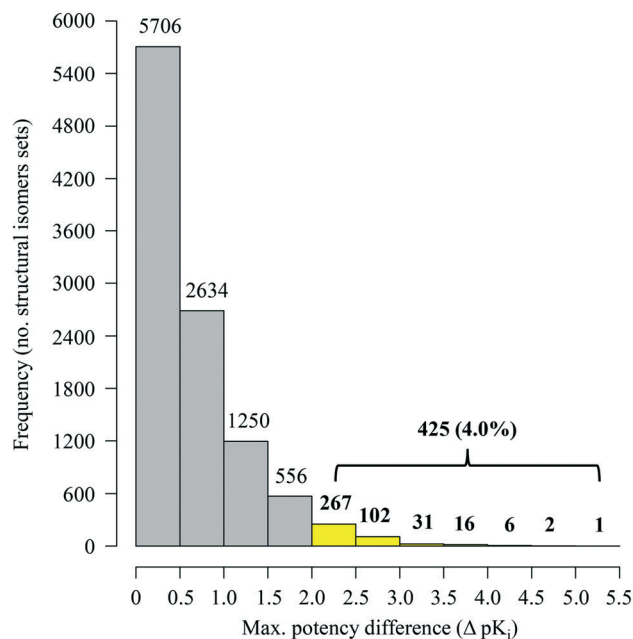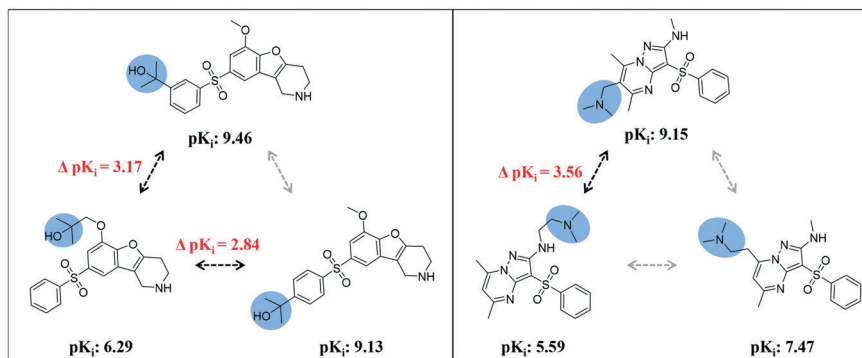
Fig. 4 Potency differences in isomer sets and isomer cliffs. (a) The distribution of maximum pairwise potency differences (ΔpK_i) in isomer sets is reported. In only 4% of all isomers sets (yellow bars), the 100-fold potency difference threshold for AC formation was met. (b) Shown are exemplary isomer cliffs (with ΔpK_i values given in red) for two isomer sets formed by serotonin 6 (5-HT6) receptor ligands.

nicely illustrate "magic methyl" effects as a consequence of positional variations. Fig. 6b depicts corresponding examples of extended MMP-cliffs with an H ↔ F transformation, which revealed potency effects of fluorine substitutions at varying positions.

Among the 1182 MMP-cliffs with isomer extension, structural isomers of weakly potent, highly potent, or both cliff compounds were detected for 589, 496 and 97 MMP-cliffs, respectively (with a median value of one isomer per MMP-cliff). Fig. 7 shows examples of MMP-cliffs with different chemical transformations for which structural isomers of both weakly and highly potent cliff compounds were available. These examples illustrate various effects of methyl to phenyl or chlorine to bromine replacements at different positions.

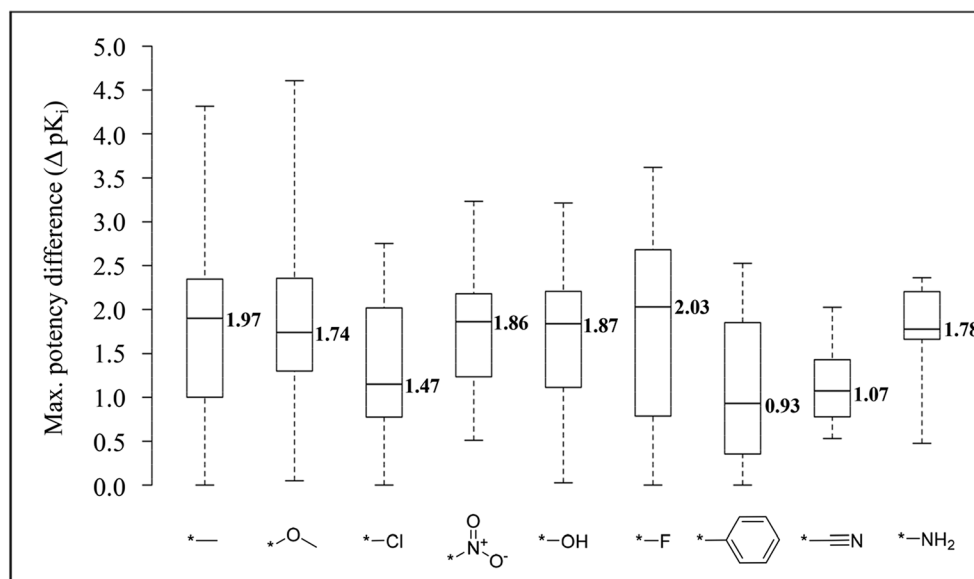Taken together, the representative examples discussed above reveal that extension of MMP-cliffs with structural isomers was SAR-informative even in cases where the AC potency difference threshold was not reached and no formally defined isomer/MMP-cliffs were obtained. However, isomer extension generally resulted in an increase in relevant compound relationships and positional effects of substitutions provided additional SAR information for nearly 1200 MMP-cliffs with activity against 147 targets (Fig. 3).

Isomer/MMP-cliffs

Our analysis yielded a total of 597 isomer/MMP-cliffs, which represented the subset of MMP-cliffs with isomer extension having largest potency effects. Compared to MMP-cliffs, the small number of currently available isomer/MMP-cliffs indicates that SARs involving isomers of specific substitutions are only little explored. This also

Chemical transformation	No. MMP-cliffs with isomer extension	No. all MMP-cliffs	No. all target sets
*— ⟷ H	68	289	114
*-O- ⟷ H	33	125	53
*—Cl ⟷ H	24	86	53
*-phenyl ⟷ *—	19	67	43
*-NO2 ⟷ H	19	34	12
*—OH ⟷ H	17	99	56
*—F ⟷ H	16	50	32
*-cyclohexyl ⟷ H	14	113	64
*-O- ⟷ *—	10	16	13
*—≡N ⟷ H	10	32	25
*-O- ⟷ *—Cl	9	19	17
*-O- ⟷ *—F	9	25	17
*—NH2 ⟷ H	9	45	25
*—≡N ⟷ *—Cl	8	14	13
*—CF3 ⟷ *—F	7	15	14

(a)

(b)

Fig. 5 Transformations in MMP-cliffs with isomer extension. (a) Listed are the top 15 most frequent chemical transformations in MMP-cliffs with isomer extension. Nine transformations representing hydrogen atom replacements are highlighted using a gray background. (b) For these nine chemical transformations, the distribution of maximum pairwise potency differences between MMP-cliff compounds and structural isomers is shown in boxplots. In each case, the median value is reported.
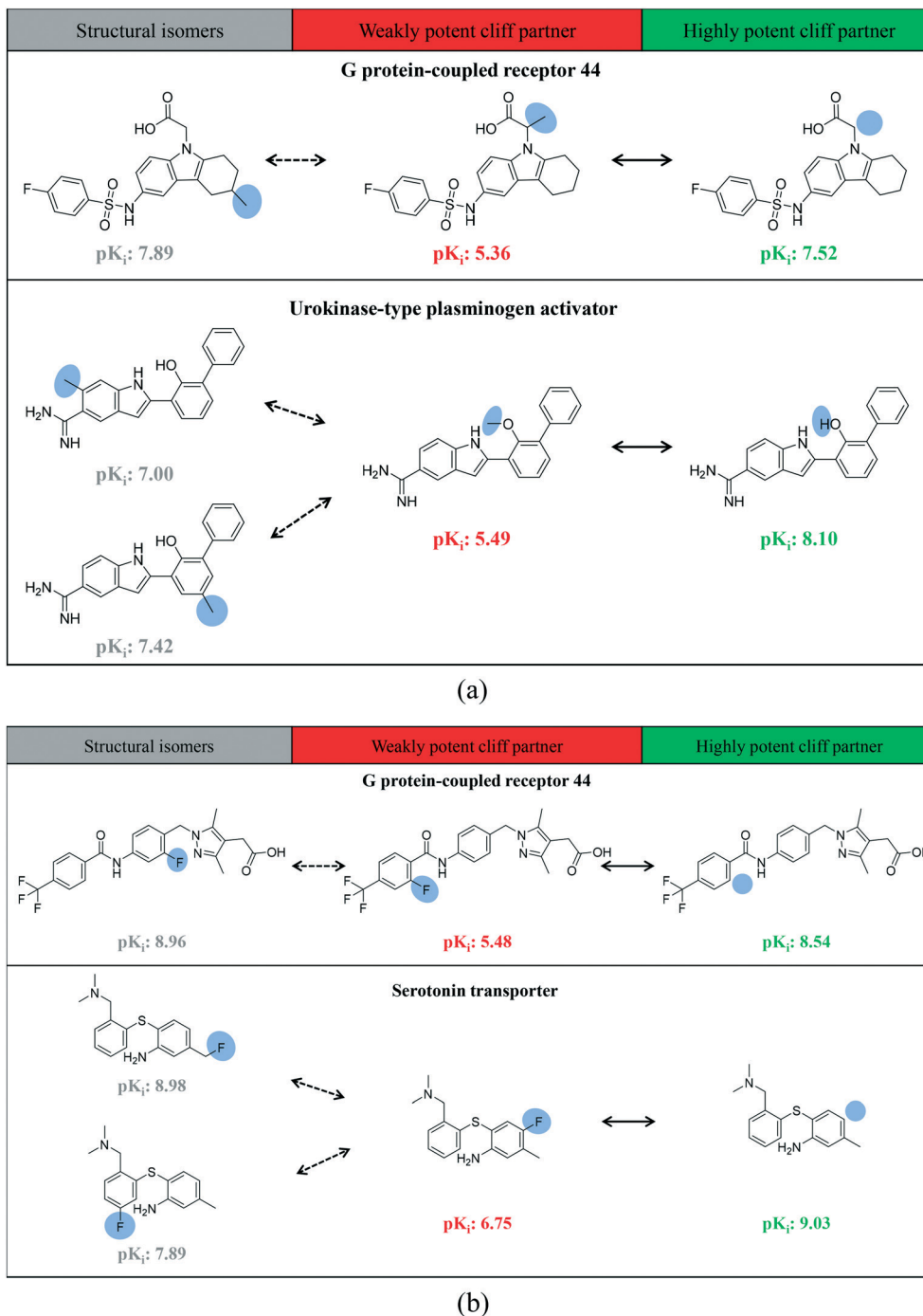
(a)

(b)

Fig. 6 MMP-cliffs with smallest transformations and isomer extension. Shown are exemplary MMP-cliffs with isomer extension that captured the smallest possible chemical transformation including the replacement of a hydrogen atom with a (a) methyl group (H ↔ CH$_3$) and (b) fluorine atom (H ↔ F). Compound targets are given.

indicates that newly introduced substituents at a given sites are only infrequently considered at other positions in active compounds. This has implications for practical medicinal chemistry and suggests further analog design strategies such as the introduction of new substituents at proximal yet distinct sites (yielding an MMP with isomer extension).

Fig. 8 shows exemplary isomer/MMP-cliffs for which qualifying structural isomers of highly potent (Fig. 8a) or weakly potent cliff compounds (Fig. 8b) were available. The examples illustrate how an isomer/MMP-cliff transforms a standard MMP-cliff into an AC with different substituents at two sites, thereby extending the MMP-formalism according to which substitutions are limited to
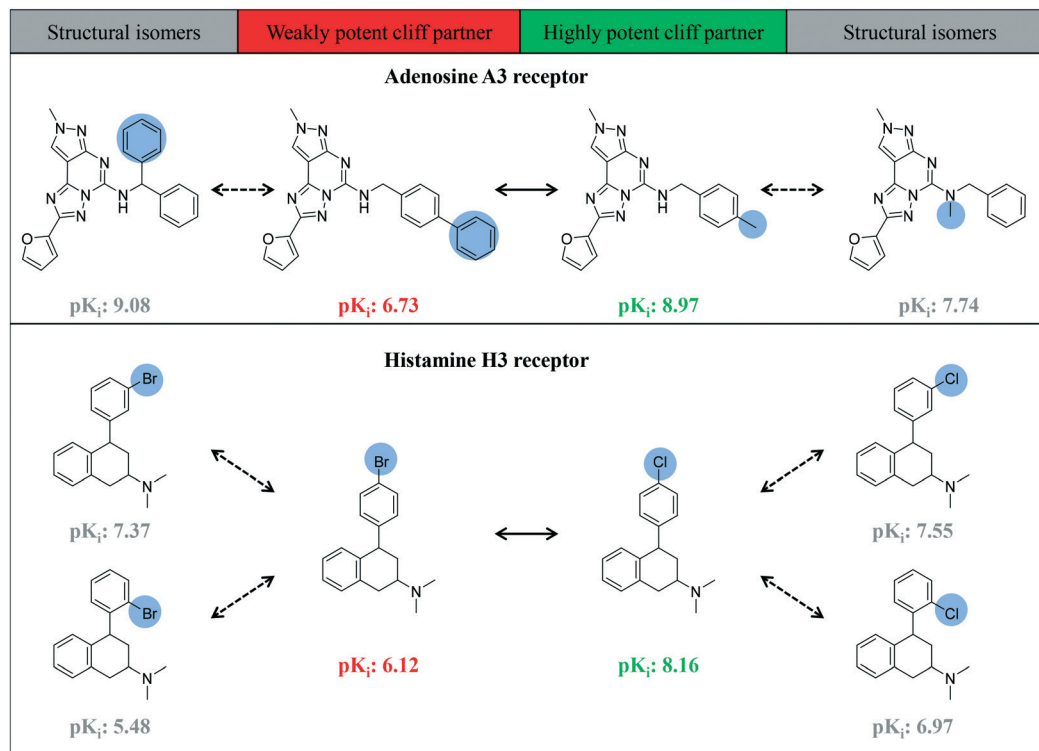
Fig. 7 Fully extended MMP-cliffs. Exemplary MMP-cliffs are shown for which structural isomers of both highly and weakly potent cliff compounds were available.

a single site. ACs with substitutions at multiple sites can also be obtained if they are extracted from analog series including computationally identified series.[9] In Fig. 8a, structural isomers of highly potent MMP-cliff compounds have comparable potency. In Fig. 8b, isomers of weakly potent cliff partners display larger potency variations, but the pre-defined AC potency difference threshold is met in both instances. Comparison of MMP-cliffs and corresponding isomer/MMP-cliffs makes it possible to better understand if the chemical nature of a substitution and/or its position might be more important for achieving high compound potency, which can be further assessed through the design of additional analogs. Since only a small proportion of isomer sets contained compounds with potency variations of large magnitude, as also shown herein, isomer/MMP-cliffs are also likely to indicate regions in potent compounds where key substituents might be positioned in different ways, as illustrated in Fig. 8, hence providing alternatives for chemical synthesis.

Conclusions

In this study, we have introduced a new category of ACs by associating MMP-cliffs with structural isomers, leading to the definition of isomer/MMP-cliffs. These ACs uniquely combine different similarity criteria and transform MMP-cliffs into ACs with different substituents at two sites. Through large-scale compound data analysis, the presence of isomer cliffs

and isomer/MMP-cliffs in different target sets was confirmed and a data set of MMP-cliffs with isomer extension was obtained. We have shown that isomer extension provides additional SAR information for MMP-cliffs, regardless of whether isomer/MMP-cliffs are formed or not, which depends on the chosen potency difference threshold. Hence, MMP-cliffs with isomer extension and isomer/MMP-cliffs might be considered to reveal a continuum of SARs and potency effects, rather than as discrete states. Regardless, the newly introduced data structure is highly SAR-informative. In some instances, very small chemical modifications such as the introduction of a methyl group at varying positions led to significant potency alterations in active compounds. In others, positional variation of larger substituents that were critical for high potency was readily tolerated. Such findings make this data structure interesting for SAR exploration in medicinal chemistry. From a computational perspective, isomer/MMP-cliffs are also thought to provide meaningful test cases for potency prediction methods. Hence, taken together, the extension of the MMP-cliff concept and new AC category introduced herein widen the current spectrum of AC and provide additional opportunities for SAR exploration. These opportunities also include complementary analysis of new two- and three-dimensional ACs,[17] which can be extended through structure-based predictive modeling.[18] For SAR analysis or other investigations, our data set of MMP-cliffs with isomer extension is freely available upon request.
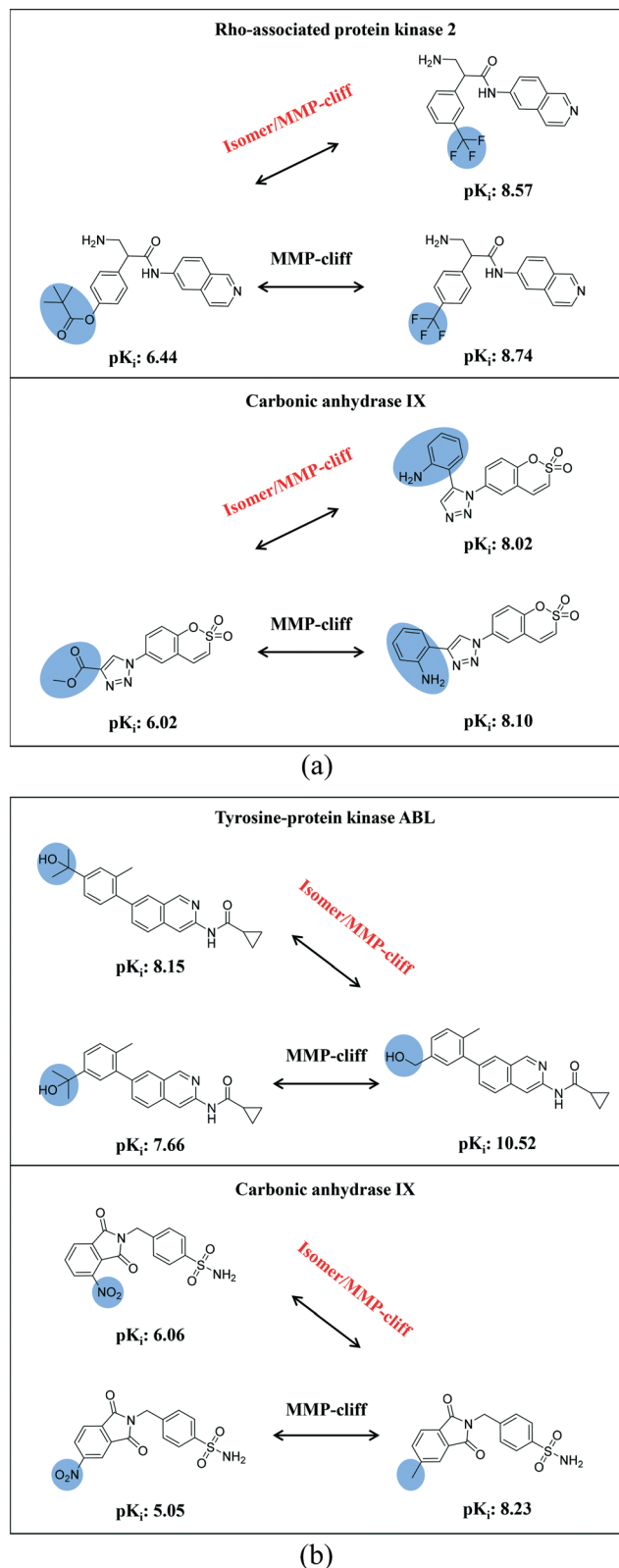
Fig. 8 Isomer/MMP-cliffs. Shown are exemplary isomer/MMP-cliffs where structural isomers replaced the (a) highly or (b) weakly potent MMP-cliff compounds.

Conflicts of interest

There is no conflict of interest to declare.

Acknowledgements

References

1 G. M. Maggiora, *J. Chem. Inf. Model.*, 2004, **46**, 1535.

2 D. Stumpfe and J. Bajorath, *J. Med. Chem.*, 2012, **55**, 2932–2942.

3 D. Stumpfe, D. Y. Hu, Y. D. Dimova and J. Bajorath, *J. Med. Chem.*, 2014, **57**, 18–28.

4 G. M. Maggiora, M. Vogt, D. Stumpfe and J. Bajorath, *J. Med. Chem.*, 2014, **57**, 3186–3204.

5 Y. Hu and J. Bajorath, *J. Chem. Inf. Model.*, 2012, **52**, 1806–1811.

6 J. Hussain and C. Rea, *J. Chem. Inf. Model.*, 2010, **50**, 339–348.

7 X. Hu, Y. Hu, M. Vogt, D. Stumpfe and J. Bajorath, *J. Chem. Inf. Model.*, 2012, **52**, 1138–1145.

8 D. Stumpfe, D. Dimova and J. Bajorath, *J. Med. Chem.*, 2016, **59**, 7667–7676.

9 D. Stumpfe, H. Hu and J. Bajorath, *Bioorg. Med. Chem.*, 2019, **27**, 3605–3612.

10 Y. Hu, N. Furtmann and J. Bajorath, *RSC Adv.*, 2015, **5**, 43006–43015.

11 H. Hu, D. Stumpfe and J. Bajorath, *Future Med. Chem.*, 2019, **11**, 379–394.

12 Y. Hu, D. Stumpfe and J. Bajorath, *F1000Research*, 2013, **2**, e199.

13 D. Stumpfe and J. Bajorath, *Future Med. Chem.*, 2015, **7**, 1565–1579.

14 A. Gaulton, A. Hersey, M. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magariños, J. P. Overington, G. Papadatos, I. Smit and A. R. Leach, *Nucleic Acids Res.*, 2017, **45**, D945–D954.

15 *OEChem TK*, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, 2012.

16 N. Schneider, R. A. Lewis, N. Fechner and P. Ertl, *ChemMedChem*, 2018, **13**, 1315–1324.

17 Y. Hu, N. Furtmann and J. Bajorath, *RSC Adv.*, 2015, **5**, 43006–43015.

18 J. Husby, G. Bottegoni, I. Kufareva, R. Abagyan and A. Cavalli, *J. Chem. Inf. Model.*, 2015, **55**, 1062–1076.