

RSC Advances



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. This *Accepted Manuscript* will be replaced by the edited, formatted and paginated article as soon as this is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Paper**Extension of three-dimensional activity cliff information through systematic mapping of active analogs**

Ye Hu,¹ Norbert Furtmann,^{1,2} and Jürgen Bajorath^{1*}

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

²Pharmaceutical Institute, University of Bonn, An der Immenburg 4, D-53121 Bonn, Germany.

*To whom correspondence should be addressed:

Tel: +49-228-2699-306, Fax: +49-228-2699-341, E-mail: bajorath@bit.uni-bonn.de

Abstract

Activity cliffs are formed by pairs or groups of structurally similar or analogous compounds with large potency differences against a given target. Three-dimensional activity cliffs (3D-cliffs) are obtained by comparison of binding modes of ligands in complex X-ray structures. Currently, 630 high-confidence 3D-cliffs are available for a total of 61 human targets, which provide a knowledge base for structure-based SAR exploration and compound design. In this work, a systematic search for structural analogs of 3D-cliff compounds was carried out applying a variant of the matched molecular pair (MMP) formalism to further extend structure-activity relationship (SAR) information associated with 3D-cliffs. In many instances, series of active analogs were successfully mapped to 3D-cliffs. Compound relationships were explored in network representations and key compounds involved in the formation of multiple 3D-cliffs and structural relationships identified. In addition, superposition of analogs onto 3D-cliffs helped to rationalize distinguishing interactions and potency variations. In total, 1980 analogs were identified for 268 cliff compounds active against 50 human targets, which further extended 414 3D-cliffs and provided a variety of SAR environments for further study. A database comprising currently available 3D-cliffs and assigned analogs has been generated and is made freely available.

Key words: Activity cliffs, X-ray structures, ligand binding modes, matched molecular pairs, analog series, high-confidence activity data, compound mapping, networks.

Introduction

Activity cliffs are formed by pairs or groups of structurally similar or analogous compounds with large potency variations and represent primary focal points of structure-activity relationship (SAR) analysis.^{1,2} For activity cliff definition, molecular similarity and potency difference criteria must be taken into consideration.¹ Activity cliffs can be assessed on the basis of two- (2D) or three-dimensional (3D) molecular representations.² They have often been analyzed using molecular graph-based descriptors and 2D similarity methods,^{2,3} especially molecular fingerprints and Tanimoto similarity calculations.^{3,4} As an alternative to whole-molecule similarity calculations for activity cliff assignment, which can be difficult to interpret,³ transformation size-restricted matched molecular pairs (MMPs) have also been introduced, which yield a structurally conservative assessment of cliffs.⁵ An MMP is defined as a pair of compounds that are only distinguished by a structural change at a single site,⁶ i.e., the exchange of a substructure, termed a chemical transformation.⁷ Transformation size-restricted MMPs generally limit such transformations to commonly observed small chemical changes such as R-group replacements.⁵ Hence, transformation size-restricted MMPs are typically formed by structural analogs, which are chemically intuitive. Our preferred definition of a 2D activity cliff refers to a transformation-size restricted MMP formed by two compounds sharing the same specific activity but having an at least 100-fold difference in potency,⁸ yielding a so-called MMP-cliff.^{5,8}

In addition, activity cliffs can also be identified in three dimensions by comparing compound conformations⁴ or binding modes^{9,10} using 3D similarity measures. Activity cliffs assessed by ligand binding mode comparison on the basis of complex X-ray structures have been termed 3D-cliffs.⁹ To identify 3D-cliffs, X-ray structures of a given target in complex with different ligands were optimally superposed and an atomic property density function-based 3D similarity measure was applied taking conformational, positional, and atomic property

differences of bound ligands into account.⁹⁻¹² The preferred definition of 3D-cliffs focuses on pairs of crystallographic ligands that display at least 80% 3D similarity and an at least 100-fold difference in potency.^{9,10} The study of 3D-cliffs often reveals ligand-target interactions that are implicated in large-magnitude potency alterations and critical for SAR progression. Therefore, 3D-cliffs have been categorized according to observed interaction differences.¹⁰ It has also been shown that 3D-cliffs can often not be reconciled on the basis of 2D similarity calculations.⁹ Thus, 2D- and 3D-cliff information is complementary in nature. Given the abundance of bioactive compound information, there are many more 2D- than 3D-cliffs available. From currently available bioactive compounds, more than 20,000 high-confidence 2D-cliffs have been isolated¹³ most of which are formed by groups of analogs (as so-called coordinated activity cliffs⁸), giving rise to activity cliff clusters in network representations.¹³ However, 3D-cliffs add an additional layer of ligand-target interaction information to SAR analysis, which might best be exploited by combining 2D- and 3D-ligand information.

In our accompanying study, we have systematically searched for 3D-cliffs on the basis of currently available X-ray and high-confidence activity data and identified a total of 630 well-defined 3D-cliffs. From a structural perspective, this pool of 3D-cliffs provides a large knowledge base for SAR analysis and other follow-up investigations. However, given the large body of available small molecule activity data, we have taken the analysis another step forward to further extend 3D-cliff information through mapping of active compounds. The assignment of active analogs to 3D-cliffs makes it possible to bridge between 3D and 2D ligand information and helps to explore SARs. Proof-of-concept for this type of approach has recently been provided by a study of 3D-cliffs formed by kinase inhibitors that were complemented with inhibitor analogs.¹⁴ Herein, we report a systematic search for active analogs of all currently available 3D-cliffs. Our large-scale analysis has led to the extension of a total of 414 3D-cliffs involving 268 unique cliff compounds with activity against 50 human targets and further increased the SAR information associated with these cliffs.

Material and Methods

3D-cliffs

The identification of 3D-cliffs is described in detail in our accompanying study. Briefly, the UniProt¹⁵ target accession IDs (UniProtIDs) were used to link high-confidence compound activity data for human targets available in ChEMBL (release 19)¹⁶ to protein-ligand complex structures deposited in the Protein Data Bank (PDB).¹⁷ Three potency measurement-dependent data sets of X-ray structures were generated including an equilibrium constant (K_i value)-based set, an IC_{50} value-based set, and a combined K_i/IC_{50} -based set. A total of 3083 complex structures involving 340 human targets for which high-confidence compound activity data was available were obtained and searched for 3D-cliffs. Therefore, X-ray structures of a target for which multiple crystallographic ligands were available were optimally superposed and 3D binding mode similarity was determined.¹⁰ A pair of bound ligands qualified as a 3D-cliff if they displayed at least 80% 3D similarity and an at least 100-fold difference in potency. From the K_i , IC_{50} , and K_i/IC_{50} data sets, 236, 292, and 595 3D-cliffs were extracted that were associated with 26, 43, and 58 targets, respectively. Taken together, the three data sets yielded a total of 630 unique 3D-cliffs that involved 580 different compounds with activity against 61 human targets. The 3D-cliffs were subjected to visual inspection and categorized according to ligand-target interaction differences.

Collection of active compounds

Bioactive compounds with reported direct interactions with human targets, for which 3D-cliffs were identified, were collected from ChEMBL (release 19), the major public repository for compounds from medicinal chemistry sources. Only compounds with explicitly defined K_i and/or IC_{50} values were considered. Compounds with multiple measurements against the same target were discarded if these potency values differed by more than one order of

magnitude. Qualifying active compounds were organized by targets and assigned to the K_i value-based, IC_{50} value-based, and combined K_i/IC_{50} -based data set. For each target and data set, crystallographic ligands forming 3D-cliffs and all other qualifying active compounds (termed 2D-ligands) were combined.

MMP analysis

For each 3D-cliff compound, a search for structural analogs sharing the same activity was carried out by generating all possible transformation size-restricted MMPs involving the cliff compound and 2D-ligands within the same target set. Accordingly, these MMPs were termed “3D-2D-MMPs”. Each 3D-2D-MMP identified a structural analog of a 3D-cliff compound (termed “2D-analog”). All MMPs were calculated using an in-house implementation of the algorithm by Hussain and Rea⁷ that utilized the OpenEye toolkit.¹⁸

Network analysis

Relationships between 3D-cliff compounds and their 2D-analogs in different potency measurement-dependent data sets were analyzed using network representations that combined 3D-cliff and 3D-2D-MMP information. In extended 3D-cliff networks, differently colored nodes represented 3D-cliff compounds or 2D-analogs. Two types of edges were used to indicate the formation of 3D-cliffs or 3D-2D-MMP relationships. On the basis of these network representations, 3D-cliffs were distinguished according to the presence of different cliff extension patterns. Networks were drawn with Cytoscape.¹⁹

Superpositions

2D-analogs were superposed on 3D-cliff forming compounds using the OpenEye OEOmega²⁰ and OEShape²¹ toolkits. Calculated superpositions were analyzed using the Molecular Operating Environment (MOE).¹²

Results and Discussion

Compound statistics

Table 1 reports the number of 3D-cliff compounds in the potency measurement-dependent data sets, qualifying 2D-ligands, and the number of targets these compounds were active against. As detailed in our accompanying study, the K_i , IC_{50} , and K_i/IC_{50} data sets yielded 236, 292, and 595 3D-cliffs, respectively, amounting to a total of 630 unique 3D-cliffs. These cliffs were formed by 163, 336, and 559 3D-ligands, respectively (**Table 1**), which were active against 26, 43, and 58 targets.

For these targets, a search for compounds with high-confidence activity data (2D-ligands) was carried out in this study. For the K_i , IC_{50} , and K_i/IC_{50} data sets, 6010, 19,332, and 27,518 2D-ligands with relevant activity were identified in ChEMBL (**Table 1**) and provided the compound pool for further analysis.

3D-2D-MMP search

A systematic 3D-2D-MMP search was carried out in three data sets to identify 2D-analogs of 3D-cliff compounds. Initial proof-of-principle for this analog mapping procedure was established previously by assigning analogs to 3D-cliffs associated with 13 kinase targets.¹⁴ Herein, we have carried out a global search for all human targets for which 3D-cliffs are currently available.

As reported in **Table 2**, a total of 1017, 1113, and 2608 3D-2D-MMPs were obtained for the K_i , IC_{50} , and K_i/IC_{50} data sets, respectively. **Fig. 1** shows the percentage of 3D-ligands forming increasing numbers of 3D-2D-MMPs. There were 12 (~16%; K_i set), 29 (~19%; IC_{50}) and 41 (~16%; K_i/IC_{50}) 3D-ligands that formed only one 3D-2D-MMP. By contrast, ~47% (K_i), ~25% (IC_{50}) and ~34% (K_i/IC_{50}) of all 3D-ligands were found to form 3D-2D-MMPs with more than 10 2D-analogs. Therefore, for many 3D-cliff compounds, corresponding

analog series were detected. The 3D-2D-MMPs identified 697 (K_i), 887 (IC_{50}), and 1915 (K_i/IC_{50}) 2D-analogs for 77 (K_i set), 148 (IC_{50}), and 257 (K_i/IC_{50}) 3D-ligands, respectively (**Table 2**). For the K_i , IC_{50} , and K_i/IC_{50} sets, these 2D-analogs were active against 18, 35, and 48 targets, respectively. In total, 1980 unique 2D-analogs were identified for 268 3D-ligands active against 50 different targets that formed a total of 414 3D-cliffs. Thus, newly identified 2D-analogs of 3D-ligands provided a substantial extension of 3D-cliff information, as further discussed in the following.

Extended 3D-cliff networks

By combining all 3D-cliffs and 3D-2D-MMP relationships between cliff compounds and 2D-ligands, extended 3D-cliff networks were generated for all three data sets. As a representative example, the network of the IC_{50} set is shown in **Fig. 2**. In the network, nodes represented 3D-cliff partners (blue) and their 2D-analogs (gray) sharing the same biological activity. Edges between blue nodes indicated the formation of 3D-cliffs (thick red lines) and edges between blue and gray nodes 3D-2D-MMPs (black lines). The network representation revealed how clusters of 3D-cliffs were complemented by structural relationships with active analogs. The majority of 3D-cliff compounds were involved in varying numbers of 3D-2D-MMPs, which determined network topology. Differently sized clusters comprising 3D- and 2D-ligands were observed. A number of densely connected clusters contained 3D-ligands involved in many 3D-cliffs and/or 3D-2D-MMPs.

Cliff extension patterns

In extended 3D-cliff networks, different cliff extension patterns emerged, as highlighted in **Fig. 2** and depicted in detail in **Fig. 3**. In **Fig. 3a**, a 3D-cliff cluster is shown that contained three inhibitors of mitogen-activated protein kinase 14 (A, B and C) forming two coordinated 3D-cliffs (A-B and B-C). A total of 27 and 20 2D-analogs and a single analog were identified

for cliff partner A, B and C, respectively. Because 2D-analogs were detected for all (highly and weakly potent) cliff compounds, such 3D-cliff extensions in networks were designated “two-partner matches”. In the cluster shown in **Fig. 3b**, 2D-analogs were identified for only one of two TGF-beta receptor type I antagonists (D) that formed 11 3D-2D-MMPs, hence representing a “one-partner match” for this 3D-cliff (D-E). Furthermore, (single or coordinated) 3D-cliffs that were not further extended by 2D-analogs (“no match”) were also observed, as illustrated in **Fig. 3c**. Thus, network analysis revealed recurrent cliff extension patterns. 3D-cliff clusters with multiple 2D-analogs typically have high SAR information content and can be easily identified and prioritized in extended 3D-cliff networks.

Table 3 reports the number of 3D-cliffs that represented different cliff extension patterns for the three data sets. A total of 49 (K_i set), 82 (IC_{50}), and 154 (K_i/IC_{50}) 3D-cliffs with two-partner matches were detected and 114 (K_i), 92 (IC_{50}) and 231(K_i/IC_{50}) cliffs with a one-partner match. In total, 2D-analogs were identified for 414 of 630 currently available 3D-cliffs. Thus, the majority of 3D-cliffs were successfully extended through 3D-2D-MMP analysis.

3D-cliff hubs

Another prominent feature of the extended cliff networks was the emergence of individual 3D-ligands that formed multiple 3D-cliffs and 3D-2D-MMP relationships. **Fig. 4** reports the statistics for the three data sets. For 3D-ligands, there was no apparent correlation between the number of 3D-cliffs and 3D-2D-MMPs they participated in. In **Fig. 4**, two threshold values are marked indicating the formation of at least five 3D-cliffs and five 3D-2D-MMPs. 3D-ligands involved in fewer than five 3D-cliffs often had many 2D-analogs, which further emphasized the principal idea of cliff extension. **Fig. 4** also reveals that 10, six, and 19 3D-ligands were identified in the K_i , IC_{50} , and K_i/IC_{50} sets, respectively, that formed five or more

3D-cliffs and 3D-2D-MMPs. These 3D-ligands were designated “3D-cliff hubs” and emerged as key compounds from our analysis.

Fig. 5 shows two coordinated 3D-cliffs formed by three thrombin inhibitors that were 3D-cliff hubs in the K_i set. The highly potent cliff partner (PDB identifier 3RML) was involved in a total of 13 coordinated 3D-cliffs and formed MMPs with seven 2D-analogs. The two weakly potent cliff partners (2ZC9, 2ZHQ) participated in 12 and nine 3D-cliffs and had 19 and 20 2D-analogs, respectively. Despite large potency variations between the 3D-ligands and their 2D-analogs (of up to four orders of magnitude), structural changes distinguishing the analogs were small and chemically interpretable (as they were detected via transformation size-restricted MMPs). Thus, the exemplary compound series in **Fig. 5** provided substantial information for structure-oriented SAR exploration through cliff extension.

Superposition of 2D-analogs on 3D-cliffs

2D-analogs can also be analyzed in three dimensions by superposing them onto corresponding cliff compounds. This represents another feature of 3D-cliff extension. **Fig. 6** shows two exemplary 3D-cliffs for which 2D-analogs were identified for both cliff partners. For a 3D-cliff formed by inhibitors of mitogen-activated protein kinase 14 (**Fig. 6a**), corresponding 2D-analogs were superposed onto the weakly potent (PDB identifier 2ZB1) and highly potent cliff partner (3RIN). The methyl oxadiazole group of the weakly potent cliff compound was replaced by a thiophene ring with an amide linker in the corresponding analog. The superposition indicated that the introduction of this linker resulted in further improved shape complementary between the substituted ring and a lipophilic pocket in the binding site, which provided a plausible rationale for the 100-fold improved potency of the 2D-analog relative to the weakly potent cliff partner. By contrast, the highly potent cliff partner and its superposed analog only differed by the introduction of an additional fluorine atom. This hydrogen-

fluorine exchange at the phenyl ring did not modify observed ligand-target interactions and had no notable influence on compound potency.

Fig. 6b shows another exemplary 3D-cliff formed by beta-secretase 1 inhibitors. The 2D-analog of the weakly potent cliff partner (3RTN) also displayed higher potency. The superposition of these compounds suggested that the increase in linker length between the amide group and the cyclohexane ring of the analog resulted in further improved shape complementarity with the active site (similar to the example discussed above). Furthermore, the highly potent 3D-cliff compound (3RSV) and its analog differed by an inverted stereo center and the exchange of a cyclohexane ring with a dimethyl oxacyclohexane group. The inverted stereo center led to a different orientation of a part of the analog, which prevented the formation of a possible hydrogen bond between the amide group of the analog and the enzyme. This hydrogen bond was observed in the X-ray structure of the highly potent cliff partner and thus provided a plausible explanation for the higher potency of the cliff compound compared to its analog. Assignment of 2D-analogs to 3D-cliffs provided many such opportunities to further explore SARs on the basis of superpositions.

Targets, cliffs, and analogs

Of the 61 targets for which 3D-cliffs were available, 2D-analogs were identified for 3D-cliff compounds associated with 50 targets. The number of 3D-cliffs and corresponding 2D-analogs varied substantially across these targets. In **Table 4**, the targets were ranked according to the number of 3D-2D-MMPs their 3D-cliff compounds formed (reflecting the degree of cliff extension). For each target, the number of 3D-cliffs and their unique 2D-ligands is also reported. The list contains a variety of targets including different enzymes and receptors. Many top-ranked targets were proteases such as coagulation factor Xa and thrombin from the serine protease family or beta-secretase 1 and renin from the aspartate

protease family. The summary in **Table 4** represents the current spectrum of 3D-cliffs that were extended with 2D-analogs.

Concluding remarks

In this work, we have systematically mapped active compounds to three-dimensional activity cliffs. All currently available 3D-cliffs and other compounds active against the same targets were taken into consideration, and only high-confidence activity data were used. The large-scale analysis was facilitated by a variant of the matched molecular pair formalism designed for activity cliff assessment. The assignment of active analogs to 3D-cliffs bridged between 3D- and 2D-ligand information. Of 630 high-confidence 3D-cliffs that are currently available, 414 cliffs were successfully extended, and the resulting 3D-cliff environments were typically enriched with SAR information. In total, nearly 2000 2D-analogs were identified for 268 3D-cliff compounds associated with 50 human targets, which provide many opportunities for follow-up investigations. Extended 3D-cliffs were analyzed in network representations combining activity cliff information and analog relationships, which identified a number of cliff compounds with notable hub character. In addition, comparison of 3D-cliff partners and corresponding 2D-analogs has been shown to provide attractive starting points for structure-oriented SAR exploration. All currently available 3D-cliffs, 2D-analogs, and detailed compound information will be made freely available upon publication in a deposition on the open access ZENODO platform under the authors' names (www.zenodo.org).

Acknowledgement

N.F. was supported by a fellowship from the Jürgen Manchot Foundation, Düsseldorf, Germany. We are grateful to OpenEye Scientific Software, Inc., for the free academic license of the OpenEye Toolkits.

References

1. D. Stumpfe and J. Bajorath, *J. Med. Chem.* 2012, **55**, 2932-2942.
2. D. Stumpfe, Y. Hu, D. Dimova and J. Bajorath, *J. Med. Chem.* 2014, **57**, 18-28.
3. G. Maggiora, M. Vogt, D. Stumpfe and J. Bajorath, *J. Med. Chem.* 2014, **57**, 3186-3204.
4. J. L. Medina-Franco, K. Martínez-Mayorga, A. Bender, R. M. Marín, M. A. Giulianotti, C. Pinilla and R. A. Houghten, *J. Chem. Inf. Model.* 2009, **49**, 477-491.
5. X. Hu, Y. Hu, M. Vogt, D. Stumpfe and J. Bajorath, *J. Chem. Inf. Model.* 2012, **52**, 1138-1145.
6. E. Griffen, A. G. Leach, G. R. Robb and D. J. Warner, *J. Med. Chem.* 2011, **54**, 7739-7750.
7. J. Hussain and C. Rea, *J. Chem. Inf. Model.* 2010, **50**, 339-348.
8. D. Stumpfe, A. de la Vega de León, D. Dimova and J. Bajorath, *F1000Res.* 2014, **3**, 75.
9. Y. Hu and J. Bajorath, *J. Chem. Inf. Model.* 2012, **52**, 670-677.
10. Y. Hu, N. Furtmann, M. Gütschow and J. Bajorath, *J. Chem. Inf. Model.* 2012, **52**, 1490-1498.
11. L. Peltason and J. Bajorath, *Chem. Biol.* 2007, **14**, 489-497.
12. Molecular Operating Environment (MOE), 2014.09; Chemical Computing Group Inc.: Montreal , Quebec, Canada.
13. D. Stumpfe, D. Dimova and J. Bajorath, *J. Chem. Inf. Model.* 2014, **54**, 451-561.
14. N. Furtmann, Y. Hu and J. Bajorath, *J. Med. Chem.* 2015, **58**, 252-264.
15. UniProtConsortium. *Nucleic Acids Res.* 2010, **38**, D142-D148.

16. A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani and J. P. Overington, *Nucleic Acids Res.* 2012, **40**, D1100-D1107.
17. H. Berman, K. Henrick, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.* 2000, **28**, 235-242.
18. OEChem, version 1.7.7, OpenEye Scientific Software, Inc., Santa Fe, NM, USA. 2012. <http://www.eyesopen.com>.
19. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res.* 2003, **13**, 2498-2504.
20. OEOmegaTK, v. Feb2014; OpenEye Scientific Software Inc: Santa Fe, NM, 2014. <http://www.eyesopen.com>.
21. OEShapeTK, v. Feb2014; OpenEye Scientific Software Inc: Santa Fe, NM, 2014. <http://www.eyesopen.com>.

Figure legends

Fig. 1. 3D-cliff compounds forming 3D-2D-MMPs. Shown is the percentage of 3D-cliff compounds that were found to form increasing numbers of 3D-2D-MMPs with 2D-analogs for the (a) K_i , (b) IC_{50} , and (c) K_i/IC_{50} value-based data sets.

Fig. 2. Extended 3D-cliff network. Shown is the extended 3D-cliff network for the IC_{50} data set. In the network, nodes represent 3D-cliff partners (blue) and their 2D-analogs (gray) sharing the same biological activity. Edges between blue nodes indicate the formation of 3D-cliffs (thick red lines) and edges between blue and gray nodes 3D-2D-MMPs (black lines). Three clusters (1-3) of ligands with different cliff extension patterns are marked.

Fig. 3. Cliff extension patterns. In (a)-(c), compounds forming three exemplary clusters highlighted in the network in Fig. 2 are shown that represent different cliff extension patterns. Ligands forming 3D-cliffs are presented on a light blue background. For each 3D-cliff compound, representative 2D-analogs are shown (if available). Structural differences between 3D-cliff compounds and their 2D-analogs are highlighted in red. For each (3D or 2D) ligand, the negative logarithmic IC_{50} value is given.

Fig. 4. 3D-cliffs versus 3D-2D-MMPs. Shown are scatter plots that report the number of 3D-cliffs and 3D-2D-MMPs for all 3D-cliff compounds in the (a) K_i , (b) IC_{50} , and (c) K_i/IC_{50} data set. Each dot represents one or more 3D-cliff compounds forming varying numbers of 3D-cliffs and 3D-2D-MMPs. Black lines delineate a region where 3D-cliff compounds are involved in at least five 3D-cliffs and five 3D-2D-MMPs. The number of 3D-cliff compounds (3D-ligands) falling into this region is reported. 3D-ligands with five or more 3D-cliffs and 3D-2D-MMPs were designated “3D-cliff hubs”.

Fig. 5. 3D-cliff hubs. Shown are three thrombin inhibitors that qualified as 3D-cliff hubs and two coordinated 3D-cliffs they formed. For each cliff, the crystallographic binding site view is shown. Carbon atoms of the highly potent cliff partner are colored in cyan and carbons of the weakly potent cliff partners in magenta. For each 3D-cliff compound, four exemplary 2D-analogs are displayed. Structural differences are highlighted in red and negative logarithmic K_i values are provided for all compounds.

Fig. 6. Superposition of 3D-cliff compounds and 2D-analogs. At the top, exemplary 3D-cliffs with “two-partner matches” of 2D-analogs are shown formed by inhibitors of (a) mitogen-activated protein kinase 14 and (b) beta-secretase 1. Carbon atoms of the highly potent cliff partners are colored in cyan and carbons of the weakly potent partners in magenta. Active site regions of the enzymes are depicted as surfaces. In the middle, superpositions of 2D-analogs with varying potency (carbon atoms colored in green or orange) onto individual 3D-cliff partners bound to their target are shown. The locations of structural differences between 3D-cliff compounds and 2D-analogs are encircled (red). At the bottom, molecular graphs of 3D-cliff partners and corresponding 2D-analogs are presented (with structural differences highlighted in red) and their negative logarithmic IC_{50} values are provided. For 3D-cliff compounds, PDB identifiers of their X-ray complex structures are given.

Fig. 1

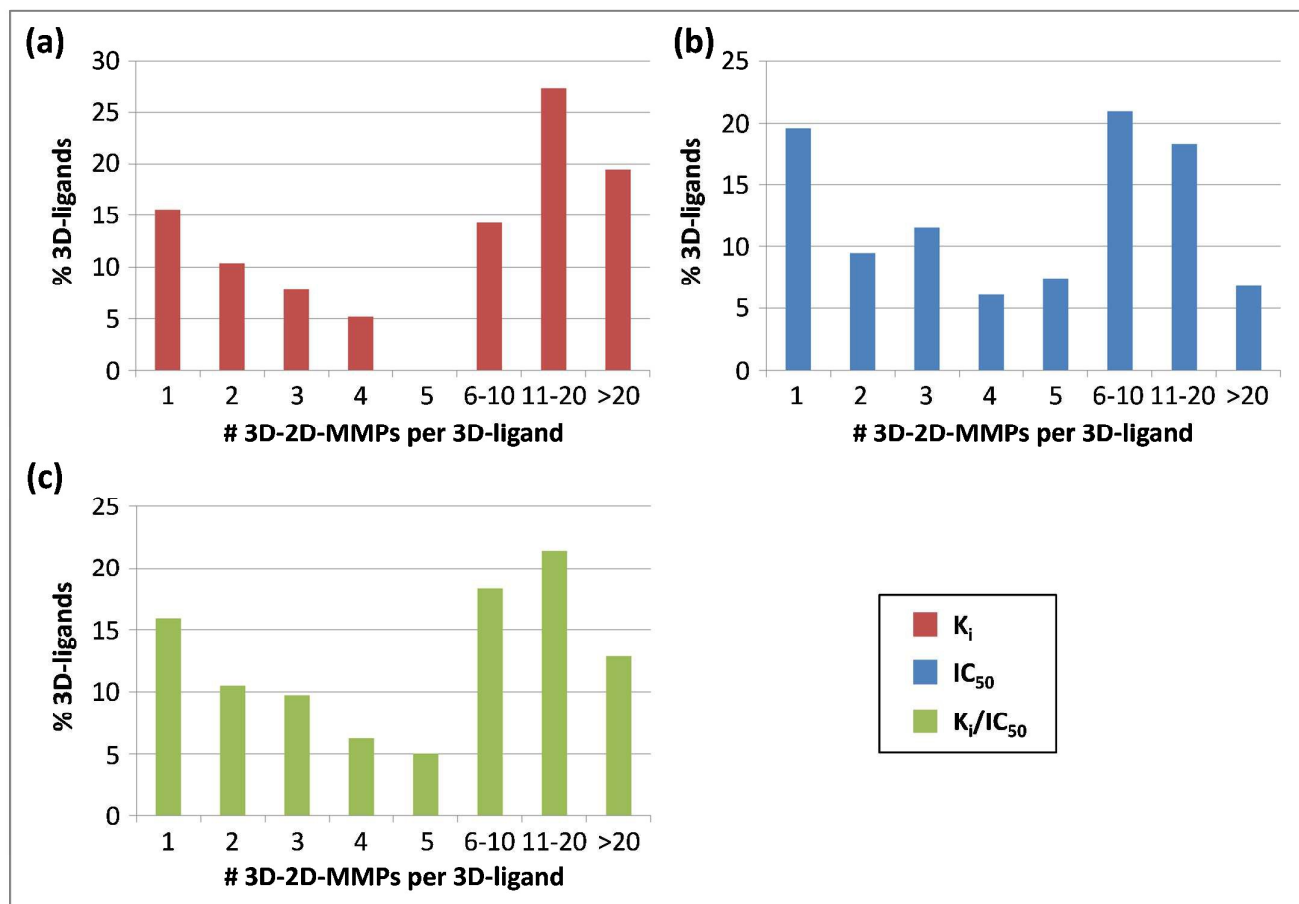


Fig. 2

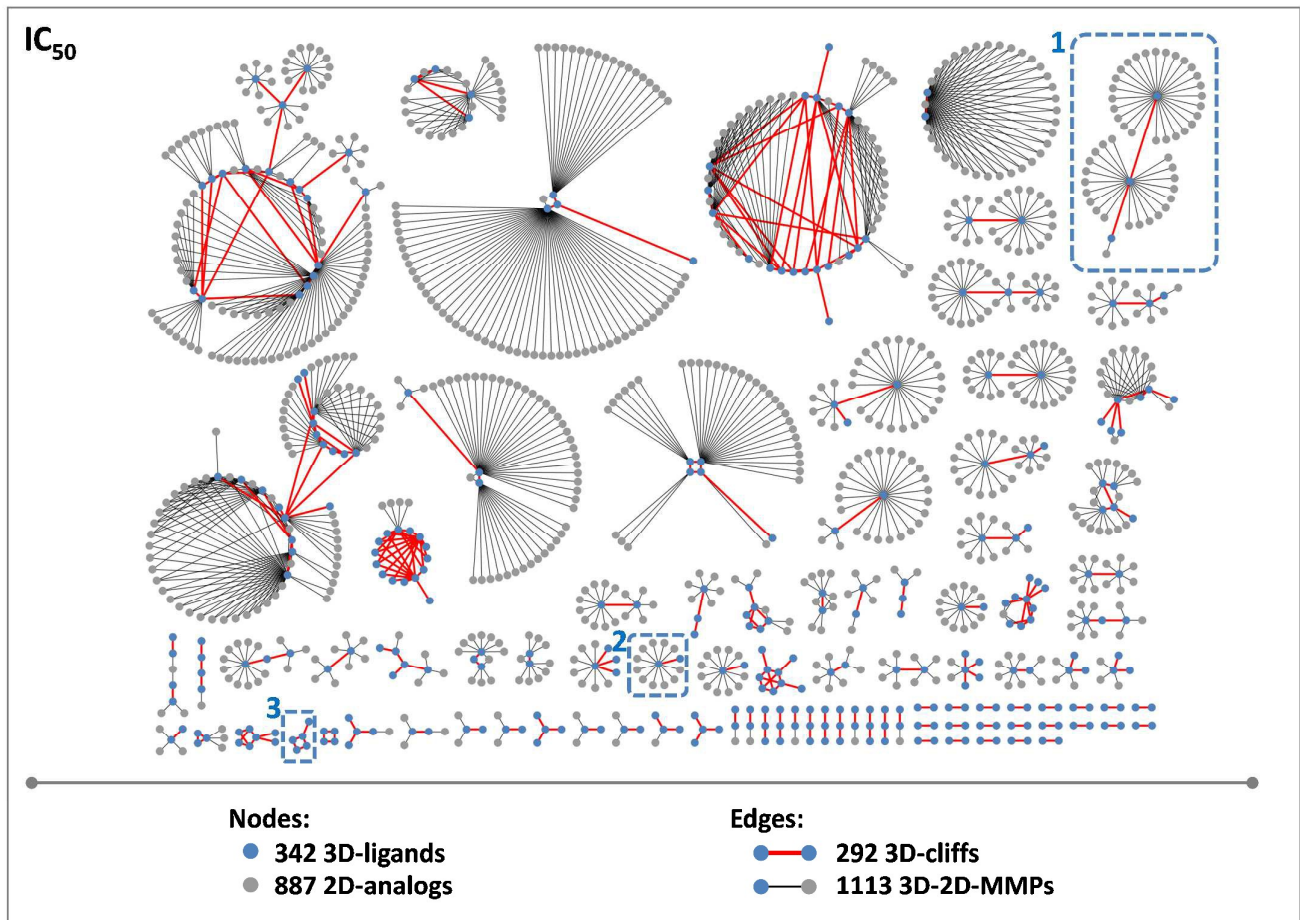
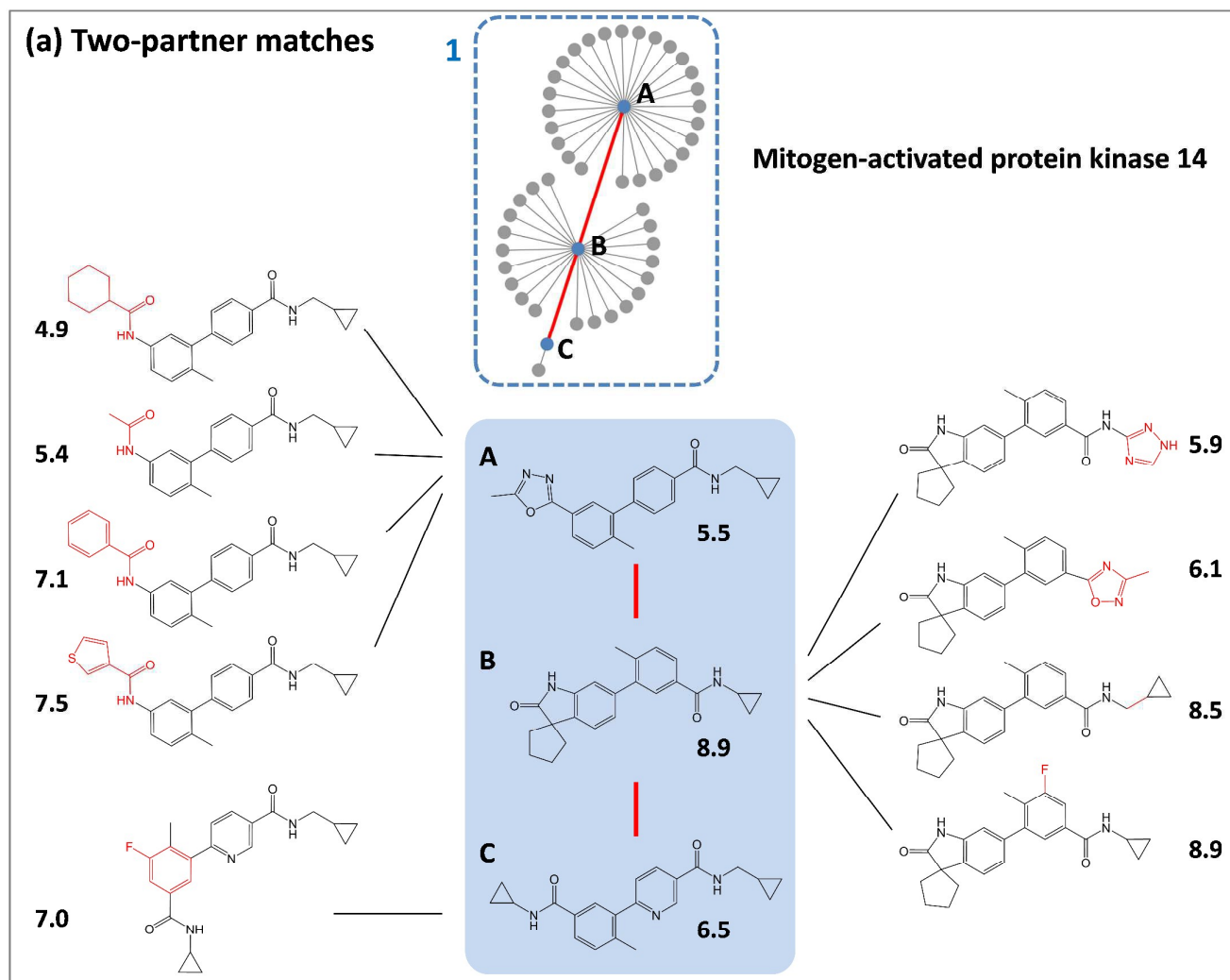


Fig. 3



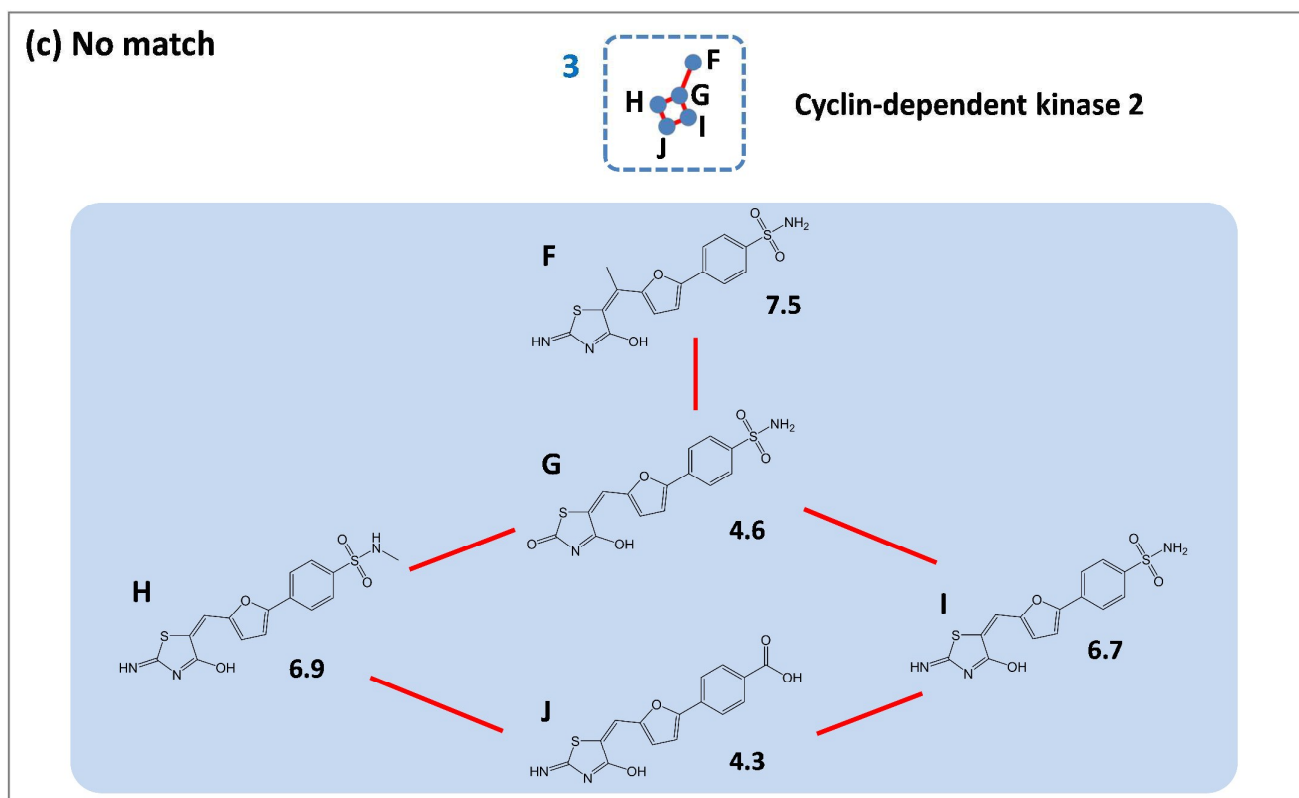
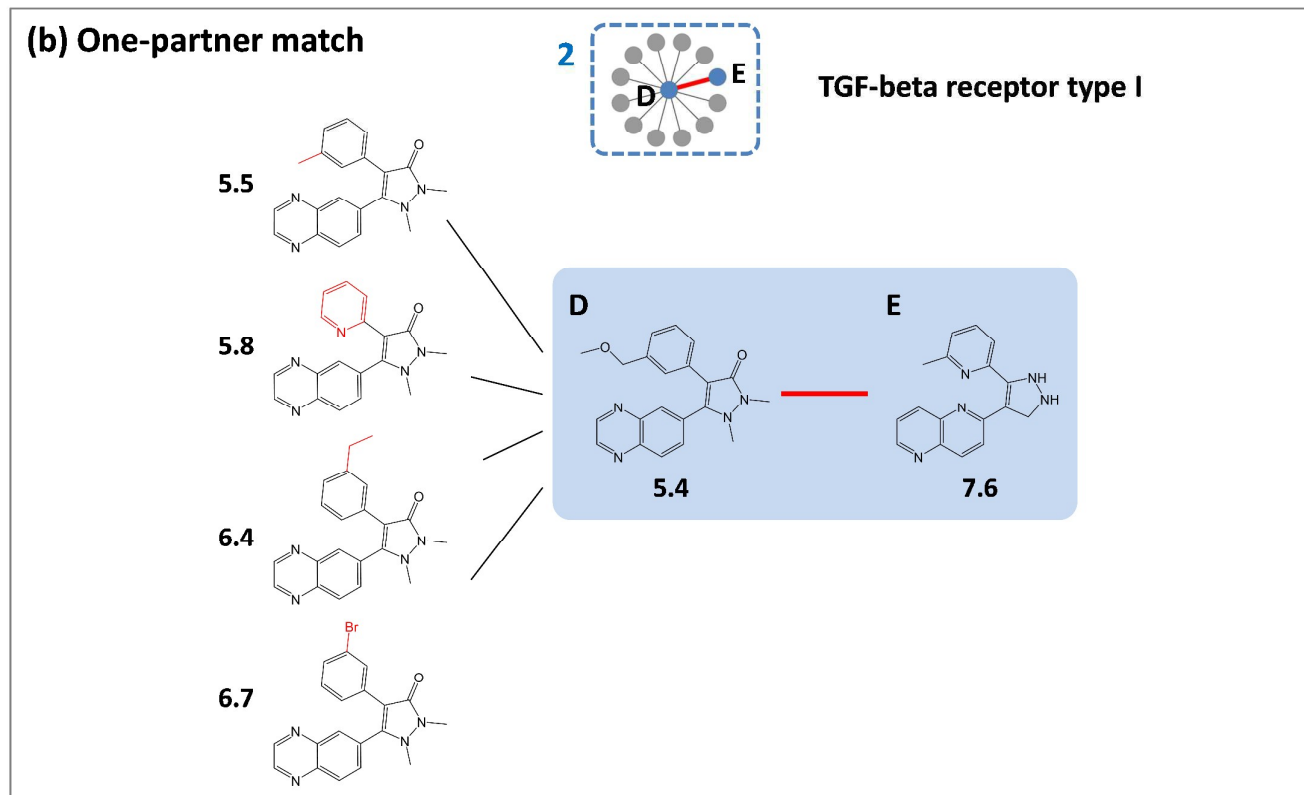


Fig. 4

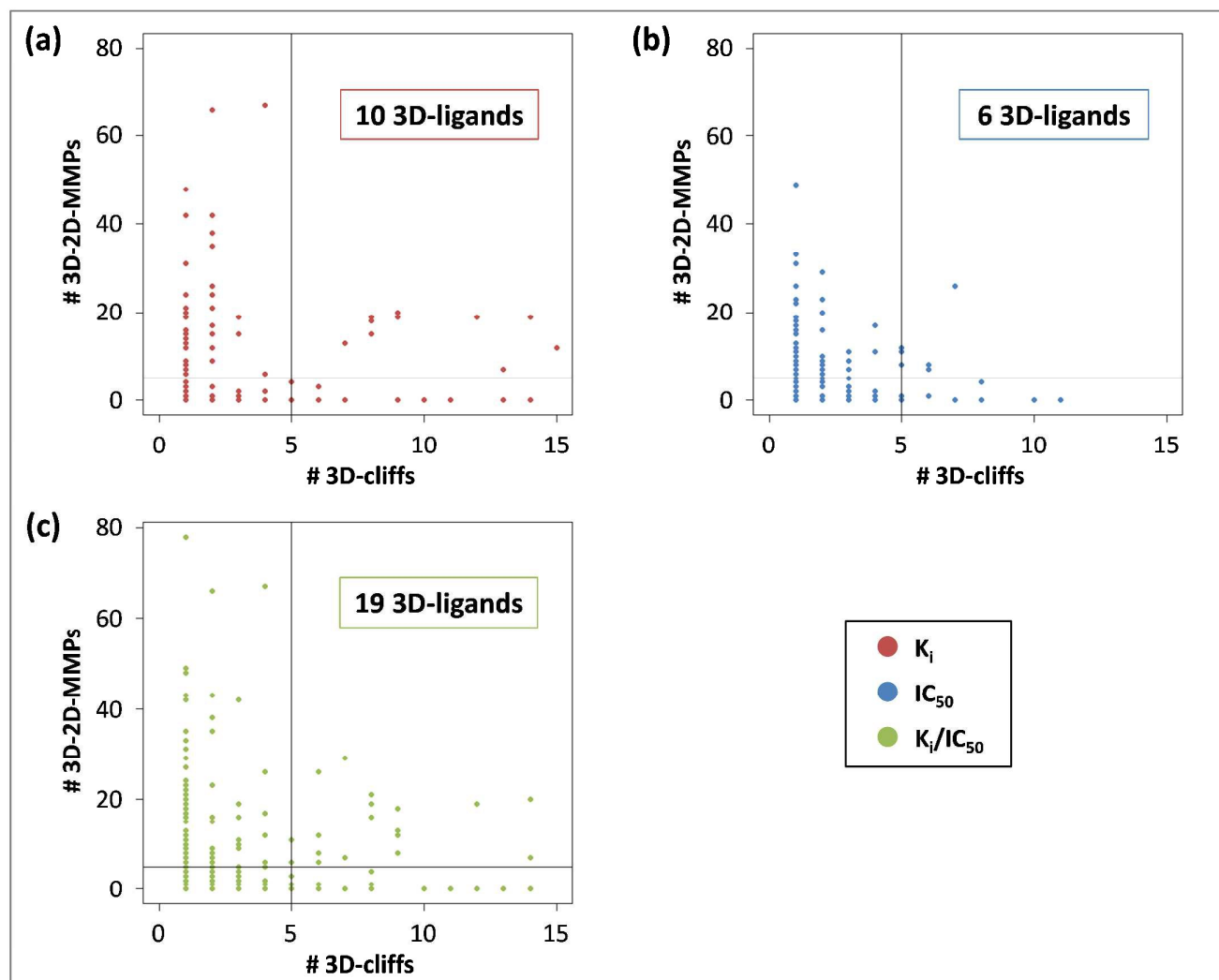


Fig. 5

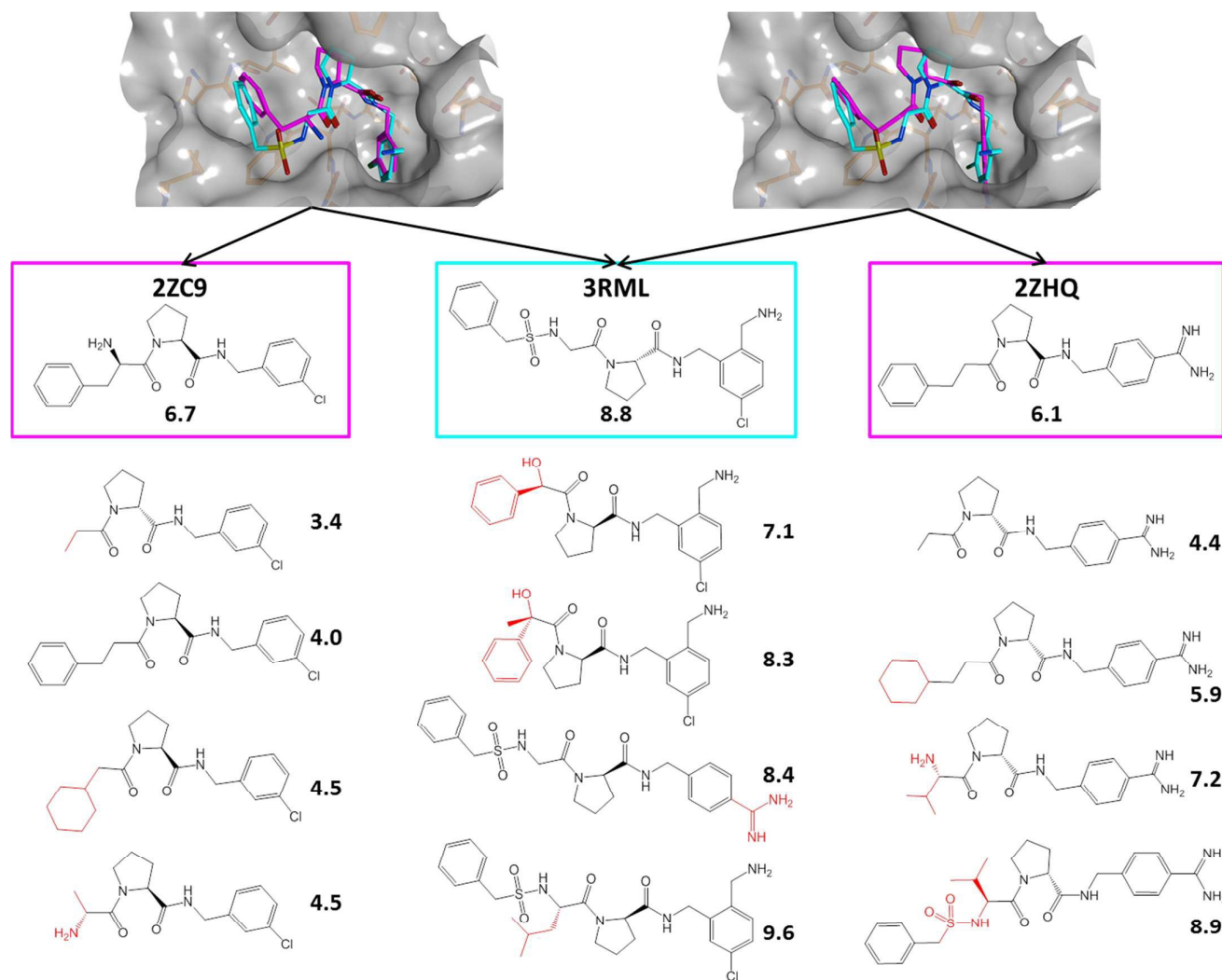
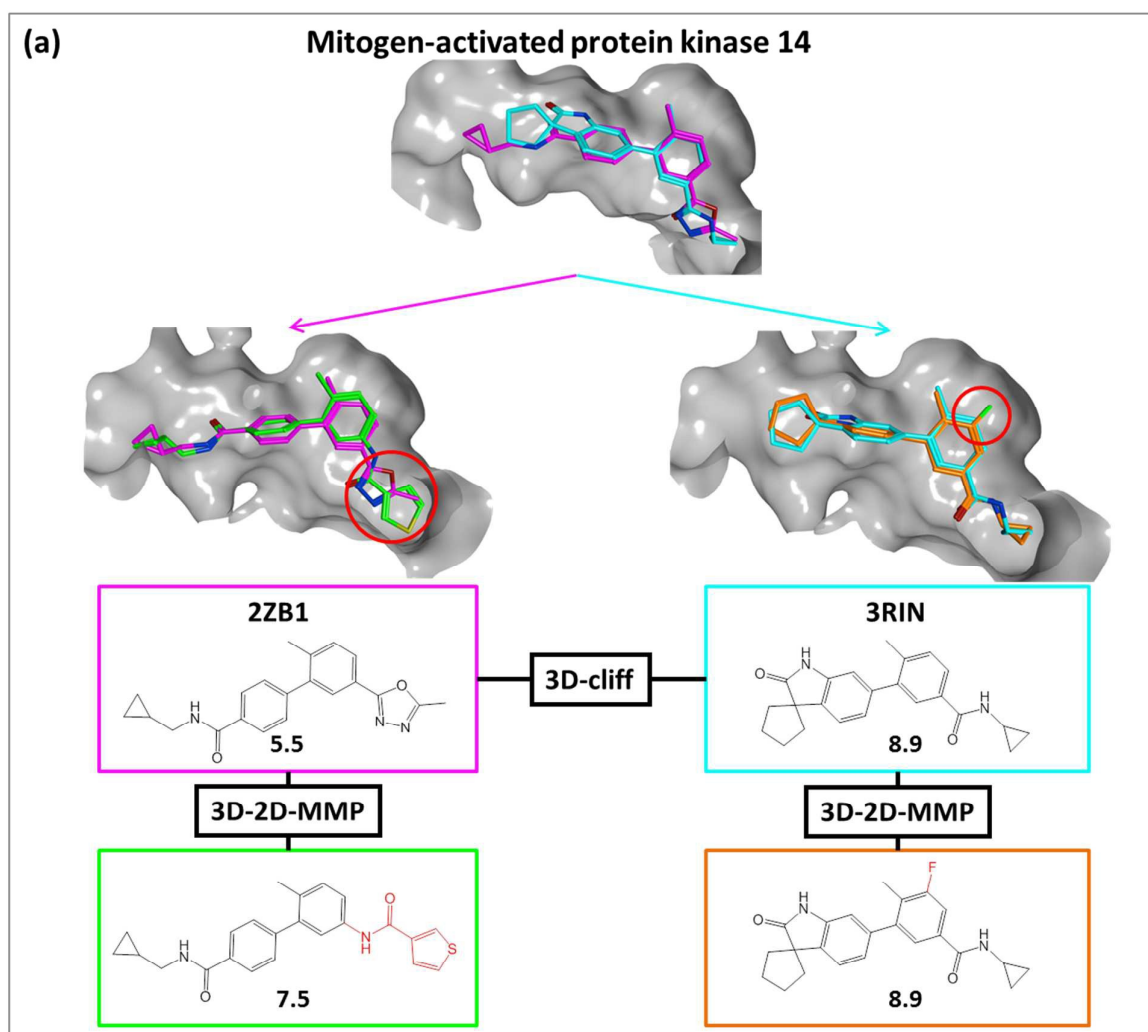


Fig. 6



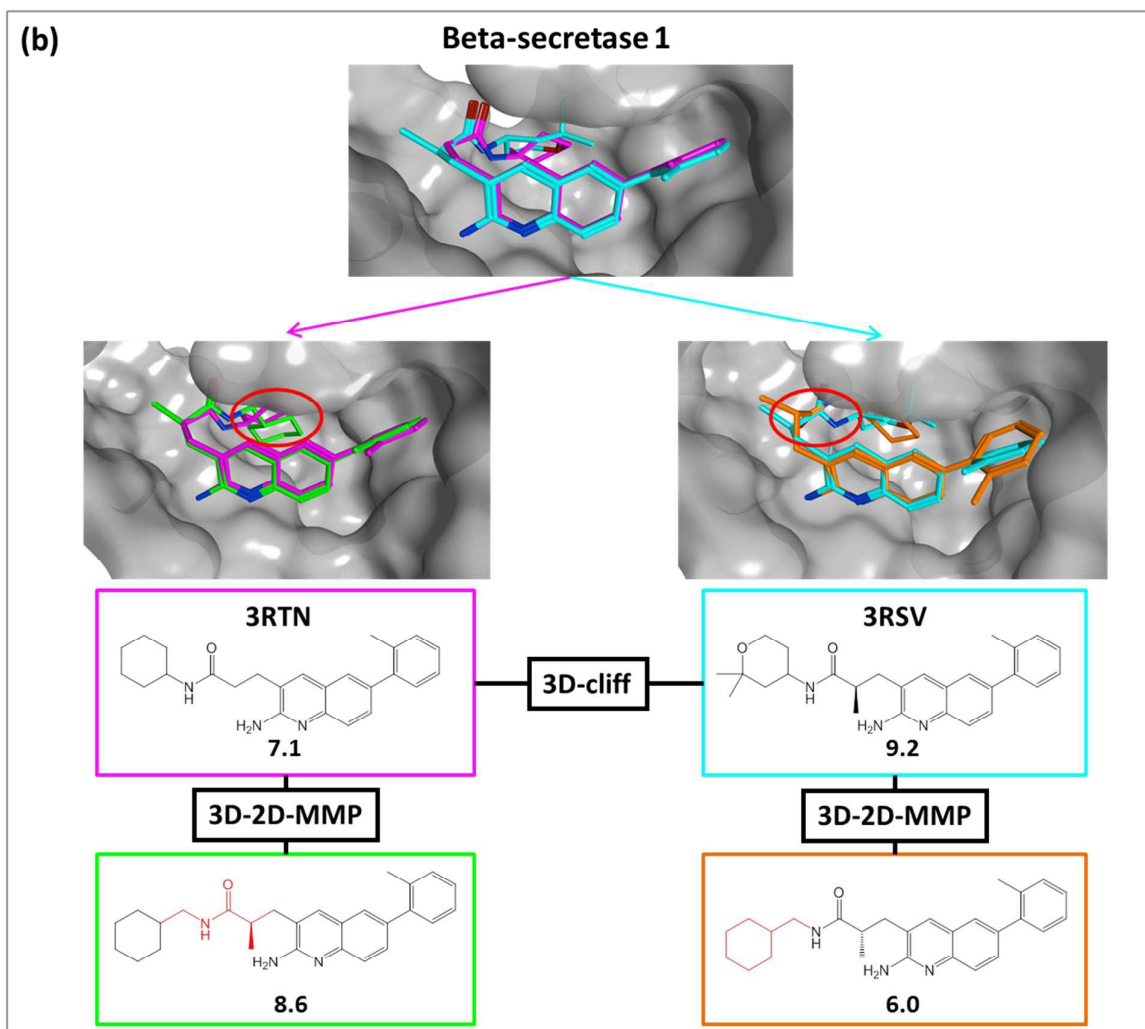


Table 1. Data sets^a

Number of	K_i	IC₅₀	K_i + IC₅₀
3D-ligands	163	336	559
2D-ligands	6010	19,332	27,518
Targets	26	43	58

^aFor the K_i, IC₅₀, and K_i/IC₅₀ value-based data sets, the total number of targets for which 3D-cliffs were obtained, 3D-ligands (from X-ray structures) involved in the formation of 3D-cliffs, and corresponding 2D-ligands (from ChEMBL) is reported.

Table 2. 3D-2D-MMP statistics^a

Number of	K_i	IC₅₀	K_i/IC₅₀
3D-2D-MMPs	1017	1113	2608
3D-ligands	77	148	257
2D-analogs	697	887	1915
Targets	18	35	48

^aFor the K_i, IC₅₀, and K_i/IC₅₀ data sets, the number of 3D-2D-MMPs formed between 3D-cliff compounds and active 2D-analogs is reported. In addition, the number of unique 3D- and 2D-ligands involved in MMP formation and the number of targets these ligands were active against is given.

Table 3. Extension of 3D-cliffs^a

	Number of 3D-cliffs (%)		
	K_i	IC_{50}	K_i/IC_{50}
Two-partner matches	49 (20.8%)	82 (28.1%)	154 (25.9%)
One-partner match	114 (48.3%)	92 (31.5%)	231 (38.8%)
No match	73 (30.9%)	118 (40.4%)	210 (35.3%)
Total	236	292	595

^aFor each data set, the number (percentage) of 3D-cliffs for which both cliff partners formed 3D-2D-MMPs with 2D-ligands (“Two-partner matches”), only one cliff partner formed MMPs (“One-partner match”), or for which no 2D-analogs were identified (“No match”) is reported.

Table 4. Targets with extended 3D-cliffs^a

Target name	Number of		
	3D-cliffs	3D-2D-MMPs	2D-analogs
Coagulation factor Xa	30	559	468
Beta-secretase 1	68	424	309
Thrombin	174	323	143
Carbonic anhydrase II	24	172	85
Renin	8	153	135
Leukotriene A4 hydrolase	37	93	37
Glutamate carboxypeptidase II	2	91	43
Dipeptidyl peptidase IV	5	87	54
Mitogen-activated protein kinase 14	14	75	74
Estrogen receptor alpha	10	74	68
Hepatocyte growth factor receptor	5	60	60
Protein-tyrosine phosphatase 1B	8	51	47
Peroxisome proliferator-activated receptor alpha	1	43	43
Hypoxanthine-guanine phosphoribosyltransferase	1	40	23
Serine/threonine-protein kinase Chk1	10	34	34
Peroxisome proliferator-activated receptor gamma	2	32	29
Kinesin-like protein 1	4	31	31
Phosphodiesterase 4B	2	29	29
Phenylethanolamine N-methyltransferase	6	28	27
Histone-lysine N-methyltransferase, H3 lysine-79 specific	2	28	17
Vascular endothelial growth factor receptor 2	4	23	23
Methionine aminopeptidase 2	4	21	18
Glycogen synthase kinase-3 beta	2	19	17
Estrogen receptor beta	2	18	18
Dihydroorotate dehydrogenase	4	16	15
Fructose-1,6-bisphosphatase	1	15	15
Serine/threonine-protein kinase PIM1	7	15	15
Tyrosine-protein kinase ABL	2	14	14
Dihydrofolate reductase	5	12	12
Heat shock protein 90-alpha	45	12	11
TGF-beta receptor type I	1	11	11
Urokinase-type plasminogen activator	13	10	9
Matrix metalloproteinase 3	2	10	6
Coagulation factor VII	4	9	9
Adenosine kinase	1	8	8

Matrix metalloproteinase 7	1	6	3
Matrix metalloproteinase 12	2	6	5
Serine/threonine-protein kinase Aurora-A	4	4	4
PI3-kinase p110-gamma subunit	2	4	4
Dual specificity mitogen-activated protein kinase kinase 1	11	4	1
Angiotensin-converting enzyme	2	3	3
Quinone reductase 2	2	3	3
Matrix metalloproteinase 8	3	3	3
Phosphodiesterase 4D	2	3	3
Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1	4	3	3
Beta-glucocerebrosidase	1	2	2
Tyrosine-protein kinase SRC	2	2	2
Ephrin type-B receptor 4	4	2	2
Hematopoietic prostaglandin D synthase	3	1	1
Serine/threonine-protein kinase NEK2	4	1	1

^aListed are 50 targets that were ranked according to the number of available 3D-2D-MMPs formed by their crystallographic ligands and 2D-analogs. For each target, the total number of unique 3D-cliffs, 3D-2D-MMPs, and unique 2D-analogs (for all three data sets) is reported.

Graphical Contents Entry

Shown is a superposition of two analogs (green and orange) on crystallographic ligands (magenta and cyan) forming a three-dimensional activity cliff (MMP stands for Matched Molecular Pair).

