Analysing Multitarget Activity Landscapes using Protein-Ligand Interaction Fingerprints: *Interaction cliffs*

Oscar Méndez-Lucio,[1,] Albert J. Kooistra,[2] Chris de Graaf,[2] Andreas Bender[1] and José L. Medina-Franco[3,*]*

[1] Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom.

[2] Division of Medicinal Chemistry, Faculty of Sciences, Amsterdam Institute for Molecules, Medicines and Systems (AIMMS), VU University Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

[3] Facultad de Química, Departamento de Farmacia, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, Mexico

ABSTRACT

Activity landscape modelling is mostly a descriptive technique that allows rationalizing continuous and discontinuous SARs, however the interpretation, especially of activity cliffs, is not straightforward. As the nature of activity cliffs depends on the ligand and the target, information regarding both should be included in the analysis. A specific way to include this information is using protein-ligand interaction fingerprints (IFPs). In this paper we report the activity landscape modelling of 507 ligand-kinase complexes (from KLIFS database) adding IFP, which facilitates the analysis and interpretation of activity cliffs. To this end, we introduce the structure-activity-interaction similarity (SAIS) maps that incorporate information of ligand-target contact similarity. We also introduce the concept of *interaction cliffs* defined as ligand-target complexes with high structural and interaction similarity, but a large potency difference of the ligands. Moreover, the specific interaction information allowed the identification of activity cliff hot spots, which help to rationalize activity cliffs from the target point of view. In general, the information provided by IFPs helps to get a better understanding when modelling an activity landscape. This paper shows examples of analyses that can be carried out when IFPs are added to the activity landscape model.

Keywords: Activity landscape, Activity cliffs, Interaction cliffs, Protein-ligand interaction fingerprints, Kinase inhibitors, SAS maps.

INTRODUCTION

The analysis of structure-activity relationships (SARs) is a fundamental tool to understand and design new bioactive molecules. In this context, activity landscape modelling arises as a descriptive technique that allows to rationalize continuous and discontinuous SARs, but also helps to systematically analyse and characterize large data sets.[1] An activity landscape is defined as "biological response surfaces in chemical space that are obtained by adding an activity dimension to this space"[2] or as "any representation that integrates the analysis of the structural similarity and potency differences between compounds sharing the same biological activity".[1] Numerous methods have been proposed to study the activity landscape of large datasets, for example structure-activity similarity (SAS) maps,[3, 4] structure multiple-activity similarity (SmAS) maps,[5] dual and triple activity difference (DAD/TAD) maps,[6, 7] and network-like similarity graphs (NSG)[8] to name a few (more methods have been reviewed[9] recently). Moreover, activity landscape modelling is suitable for the identification of activity cliffs, which are defined as a pair of structurally similar molecules that have large changes in potency.[10, 11]

Despite the fact that activity landscape modelling has been extensively used to characterize different target[12-14] and multitarget[5, 15, 16] data sets, the interpretation of the activity landscape, especially for activity cliffs, is not straightforward.[17] In order to rationalize the formation of activity cliffs our group has used molecular docking[18] and the concept of activity cliff generators[19] to explore the causes associated with the potency difference. Other attempts to rationalize activity cliffs include the concepts of structure-based activity cliffs and activity cliff hot spots proposed by Seebeck et al., which evaluate the frequency with which a protein atom is involved in the formation of an activity cliff taking into account the interaction energies of protein-ligand complexes.[20]

As the nature of an activity cliff depends on the ligand and target, whenever possible, information regarding both should be included in the analysis. One way to accomplish this is to include explicit information regarding the target (*i.e.* sequences and sequence similarities) as previously reported for a set of kinase inhibitors.[16] A more specific alternative is to capture information of how the ligand interacts with the protein, for example, using molecular interaction fingerprints (IFPs) of protein-ligand complexes. IFPs conveniently simplify the interactions between proteins and ligands by coding them in a 1D representation.[21, 22] IFPs have been successfully used for post-processing of ligand docking poses according to known interaction patterns for protein targets in structure-based virtual screening studies[23, 24] and allow systematic mining of protein-ligand interaction space to identify conserved and selective protein interaction hot spots.[25-27] In this work we report an approach to integrate IFPs in the activity landscape modelling process, in order to identify regions in the target protein that are associated with activity cliffs. This method is particularly useful when analysing a large number of ligand-target complexes, given that it does not require of large amount of computing time or resources. We also introduce the structure-activity-interaction (SAIS) maps, which are a natural extension of the SAS maps initially developed to characterize the SAR of screening data sets. The use of IFPs facilitates the analysis of activity cliffs and the identification of scaffold hops. Moreover, adding specific interaction information allowed the identification of activity cliff hot spots, which help to gain a deeper insight in the formation of activity cliffs. To exemplify the approach, we analysed KLIFS, a recently developed and publicly available database that has information of 1,734 crystallographic structures covering 190 human kinases.[27] It is important to mention that this paper demonstrates only some of the analyses that can be carried out when IFPs are added to the activity landscape model. Another example is the analysis that was published by Furtmann *et*

al.,[28] which was published during the final preparations of this study. They have analysed 3D-cliffs using this same data set, denoting the interest of the scientific community in using crystal structures for activity landscape modelling.

MATERIALS AND METHODS

Data set

The structural data used in this study was extracted from KLIFS.[27] KLIFS is a curated database that contains 1,734 aligned crystal structures, covering 190 different human kinases, from which 1,252 are co-crystalized with a ligand. In order to facilitate the analysis of the crystal structures present in KLIFS, the ligands and the binding pockets (containing 85 amino acids) were separated in this database. More detailed information regarding the curation, alignment and preparation of KLIFS can be found elsewhere.[27] It should be noted that the pocket residue numbering and nomenclature reported in the KLIFS publication is used throughout this manuscript.

All experimental binding data was extracted from Binding MOAD database,[29, 30] which contains information from the primary reference of each PDB entry. Using this database, 409 out of the 1,252 KLIFS entries were annotated with pIC_{50} values (ranging from 3 to 10.52), pK_i values were found for 70 complexes (4.46 - 9.96) and 28 structures were annotated with pK_d values (4.43 - 8.68). It is noteworthy that the wide activity range presented in these data sets make them suitable for activity landscape analysis and activity cliffs detection. These three data sets were used for further analysis.

Structure, activity and interaction similarity

Same as in previous studies,[7, 19] a set of eleven broadly used 2D fingerprints was calculated for these datasets with MayChemTools.[31] These fingerprints include atom neighbourhoods,[32] atom

types, electrotopological state indices (EStateIndices),[33] extended connectivity (ECFP4),[34] MACCS (322 bits),[35] path length, topological atom pairs (TopAtomPairs),[36] topological atom torsions (TopAtomTorsions),[37] topological atom triplets (TopAtomTriplets), topological pharmacophore atom pairs (TopPh4Pairs),[38] and topological pharmacophore atom triplets (TopPh4Triplets).[39] The Tanimoto coefficient[40-42] was employed to assess the structural similarity using each of these 2D descriptors. ComboScore, computed with the Rapid Overlay of Chemical Structures (ROCS) module of OpenEye Scientific Software,[43] was also used to evaluate the 3D similarity among the different compounds. To compute the 3D similarity, the coordinates of the bioactive conformation of each ligand were taken from each crystal structure and used to calculate ComboScore as implemented in ROCS. In order to maintain the same range in molecular similarity measurements (0 to 1) across all fingerprints, ComboScore similarity was scaled dividing it by two. Although this is a purely ligand-based approach, the information of ligand positions in the binding pocket is encoded in the IFP of each compound (see below).

Potency differences were used to assess the activity relationship between two compounds. Activity differences were calculated as follows:

$$|\Delta pA(\mathrm{R})_{i,j}| = \left|pA(\mathrm{R})_i - pA(\mathrm{R})_j\right|$$

where $pA(R)_i$ and $pA(R)_j$ are the activities (pIC_{50}, pK_i and pK_d) of the *i*th and *j*th molecules ($j > i$) against each receptor R. In this paper R can be the same or different kinase.

Interaction similarity was assessed using IFPs extracted from KLIFS for each complex used in this study. IFPs in these datasets were calculated using the interaction fingerprints developed by Marcou and Rognan,[44] encoding seven types of interactions for each amino acid *i.e.* seven binary bits per amino acid depending if the interaction is present or absent. The seven bits correspond to the following interactions: hydrophobic contact, face-to-face aromatic interactions, face-to-edge

aromatic interactions, protein H-bond donor, protein H-bond acceptor, protein cationic interactions, protein anionic interactions (calculation details are listed in Table S1). A total of 595 bits were obtained for each complex corresponding to the 85 aligned residues that form the ATP binding pocket as defined by van Linden *et al*.[27] The interaction similarity between two complexes was also calculated using the Tanimoto coefficient,[40-42] although other similarity measures can be used as well.

Activity landscape modelling

In this study, SAIS maps were developed to analyse the multitarget activity landscape of these data sets. In a SAIS map, which is based on the structure of the SAS maps, each point represents a pairwise comparison between two protein-ligand complexes, localized by plotting the structure similarity between the two ligands in the X-axis against the absolute potency difference in the Y-axis. Data points are color-coded by protein-ligand interaction similarity using a continuous scale from more similar (red) to less similar (green). Similar to SAS maps,[4, 14, 45, 46] SAIS maps can analyse the multitarget activity landscape against one or more targets as they incorporate protein information through the measure of interaction similarity. This approach is similar to other multitarget methods in which explicit protein information is given by sequence similarity,[16] however the use of interaction similarity allows the identification of important amino acids for the formation of activity cliffs.

SAIS maps can be roughly divided in four different regions, namely regions I-IV. Region I contains pairs of molecules that have low potency difference and low structural similarity, thus they are generally considered as scaffold hops.[47] Data points located in region II are characterized by similar activity (low potency difference) and high structural similarity; therefore these pairs of compounds present continuous SARs. Region III contains data pairs with different

structures and low potency similarity (large potency difference). Finally, pairs of molecules located in region IV exhibit a discontinuous SAR denoted by high structural similarity, but high potency difference and therefore are associated with activity cliffs.

The activity landscape characterization, as described before, was achieved by dividing each plot using potency difference and structure similarity thresholds along the Y- and X-axis, respectively.[19] An additional threshold was used for the interaction similarity in order to identify differences in the binding modes. In this work, a threshold of 1 log unit of absolute difference was used to distinguish between compounds with high and low potency difference. However, assigning a threshold for structural similarity is not straightforward as different criteria to impose thresholds can be employed.[4, 48] In this work, the thresholds for structural and interaction similarity were set as the mean similarity value plus two standard deviations and were calculated individually for each data set. A graphical representation of the density distribution for each data set is shown in Figure 1.

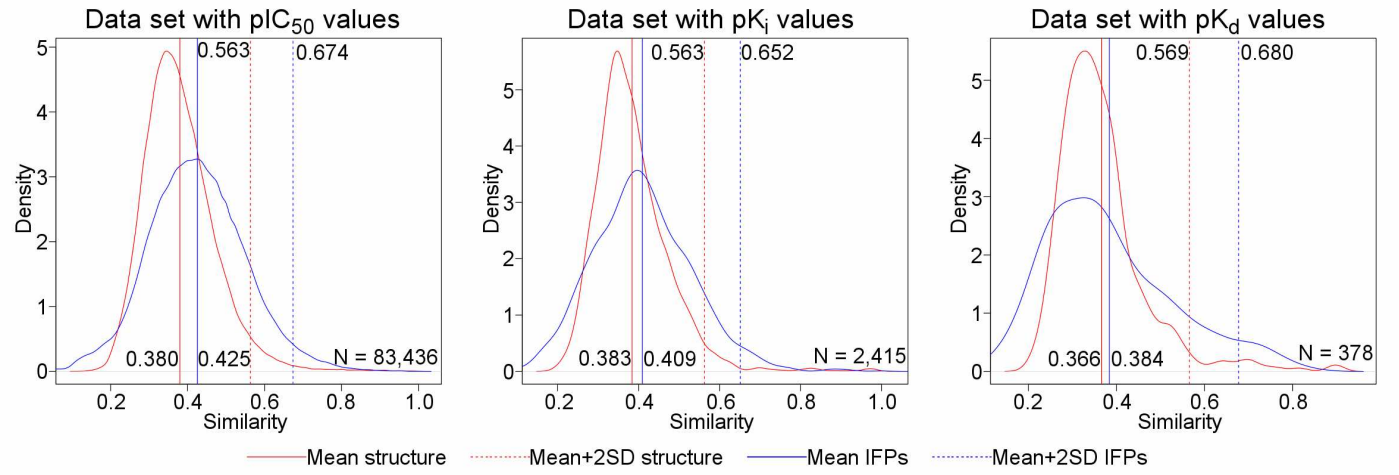


Figure 1. Density distribution of structural and interaction similarity for the three data sets (pIC_{50}, pK_i, pK_d) used in this study evaluating 83,436, 2,415 and 378 molecular pairs, respectively. The mean values of structure and interaction similarity for these data sets (using ComboScore/2 and IFPs, respectively) are shown with continuous lines.

RESULTS AND DISCUSSION

Activity landscape modelling

Figures 2A-C show the SAIS maps for each of the three data sets used in this study (pIC_{50}, pK_i, pK_d) containing 83,436, 2,415 and 378 points, respectively, and Table 1 summarizes the distribution of the data points in each region. These maps encode the relationships between the 3D binding conformation of the ligands, how they interact with the target, and the associated biological activity. In general, ligands in all data sets have different binding conformations represented by low structural similarity values. Figure 1 shows that the mean structural similarity in the three data sets, represented here by the ComboScore/2 value, ranges from 0.366 to 0.383. This low structural similarity is accompanied of large potency differences, even reaching 7.5 log units of difference. Because of these structural and potency differences, more than 50% of all the data points are located in region III of the three SAIS maps (Table 1). This distribution and the low structural similarity are associated with the large size and flexibility of the ATP binding site in kinases, which is able to accommodate very diverse compounds (e.g. type I and type II inhibitors) that bind to different subregions within the ATP binding site. The remaining data points are distributed in the following manner: region I comprises around 40% of data points, whereas region II and IV only contain less than 2% each. As expected, data points in region IV (activity cliffs) represent a small fraction of all pairs in SAIS maps, although it is known that this region provides the most information on SARs.[46]

Table 1. Distribution of data points across different regions of the SAIS maps generated for the three datasets. The percentage of points in each region is indicated in parenthesis.

Dataset	Protein-Ligand structures	Molecular pairs	Ligand structure similarity threshold[a]	I	II	III	IV
pIC_{50}	409	83,436	0.563	32,749 (39.25%)	1,276 (1.53%)	47,948 (57.47%)	1,463 (1.75%)
pK_i	70	2,415	0.563	833 (34.49%)	35 (1.45%)	1,512 (62.61%)	35 (1.45%)
pK_d	28	378	0.569	164 (43.39%)	5 (1.32%)	201 (53.17%)	8 (2.12%)

[a]Ligand structure similarity is calculated using ComboScore as implemented in ROCS

As discussed before,[49] chemical space, and hence activity landscape models, is highly dependent from the chemical representation used to describe molecules. In contrast with the strategy used in previous work where the combination of different molecular representation was needed (consensus activity landscapes),[7, 19, 48, 49] here, only the 3D similarity (*i.e.* the ComboScore, which combines shape and pharmacophore similarity) was used due to the availability of the bioactive 3D conformation of the ligands, although 3D techniques other than ROCS could have been used in addition as well. Tables S2-S4 show matrices with Pearson's correlation coefficients between all the pairwise similarities for each molecular representation used in this study. The ComboScore showed low correlation (<0.55) with the scores obtained from the other eleven 2D fingerprints used to assess molecular similarity in all data sets. Interestingly, the scores from the Extended Connectivity (in pIC_{50}) and AtomNeighborhoods (in pK_i and pK_d) presented the highest correlation with the ROCS ComboScore with correlation coefficients of 0.243, 0.408 and 0.530 for pIC_{50}, pK_i and pK_d data sets, respectively. This low

correlation suggests that the 2D descriptors computed in this work would not be appropriate to capture information related to 3D conformations, such as different binding poses that could be related to a particular interaction profile between the ligand and the target protein.

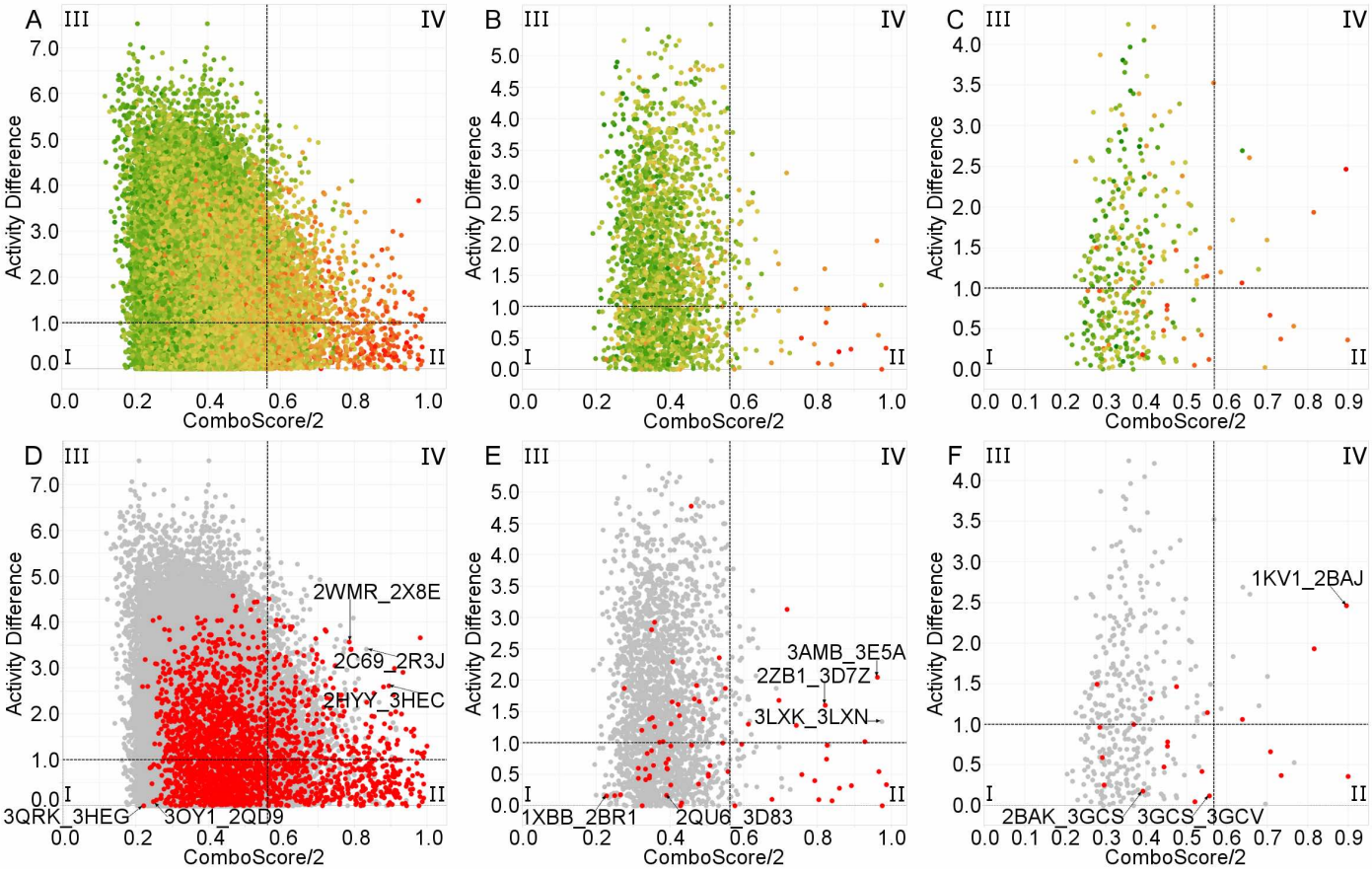


Figure 2. Panels A-C show Structure-Activity-Interaction Similarity (SAIS) maps for the three datasets used in this study (pIC_{50}, pK_i and pK_d) containing 83,436, 2,415 and 378 data points, respectively, resulting from the pairwise comparisons. Data points are color-coded by interaction similarity using a continuous scale from red for very similar interactions to green for molecular pairs that form different interactions with the target. Regions (I-IV) are labelled in each SAIS map. Panels D-F highlight in red those molecular pairs with high interaction similarity, that is, two standard deviations above mean similarity for each data set. Detailed information can be found in Table 2.

Interaction cliffs

Interaction similarity calculated using the IFPs as a representation of protein-ligand interactions, was used in order to gain a deeper understanding at the structural level of the activity cliffs present in these data sets. Alike to structural similarity, the three data sets also exhibit low interaction similarity, with mean values ranging from 0.384 to 0.425. Despite the analogous distribution in both measures, a very low correlation (< 0.45) was found between interaction similarity and structural similarity. It is worth mentioning that ComboScore showed the highest correlation with interaction similarity in all data sets compared to the 2D descriptors (Tables S2-S4). The lack of correlation between ligand and interaction similarities has been observed in large-scale studies reported before, but it has also been noted that interaction similarity can be correlated with binding site similarity.[21]

Figures 2D-F show the three SAIS maps highlighting those molecular pairs with high interaction similarity relative to the data set, that is, two standard deviations above the corresponding mean similarity for each data set. As depicted in these figures, not only those compounds with high molecular similarity show high interaction similarity, but also those molecular pairs with different chemical structures can present similar ligand-target interactions. On average, only 33% of the molecular pairs categorized as highly similar (regions II and IV) showed similar interactions. Moreover, most of the compounds with high interaction similarity are located in regions I and III, which can be related to scaffold hops (as they can retain the same pharmacophore even presenting different shape/conformation). Details regarding the distribution of molecular pairs with high interaction similarity can be found in Table 2. Interestingly, less than 40% of the pairs located in region IV present high interaction similarity, suggesting that high similar compounds do not always interact in a similar manner with the binding site of

different kinases. Those pairs of ligands that present high structural and high interaction similarity, but also a large potency/affinity difference (*i.e.* activity cliffs with high interaction similarity) can provide information regarding the specific interactions or chemical features that are directly associated with the increase or decrease of potency and will be referred to as *interaction cliffs* from now on in this work.

Table 2. Distribution of data points with high interaction similarity across different regions of the SAIS maps. The percentage of points in each region is indicated between parentheses.

Dataset	Interaction similarity threshold[a]	Molecular pairs	I	II	III	IV
pIC_{50}	0.674	83,436	838 (1.00%)	372 (0.45%)	959 (1.15%)	280 (0.34%)
pK_i	0.652	2,415	29 (1.20%)	14 (0.58%)	22 (0.91%)	7 (0.29%)
pK_d	0.680	378	11 (2.91%)	3 (0.79%)	4 (1.06%)	3 (0.79%)

[a]Interaction similarity is calculated by Tanimoto similarity of IFPs

It is noteworthy that 80% of the interaction cliffs in the pIC_{50} data set include crystal structures of kinases in two different branches of the kinome, namely CMGC and TK. From the 280 interaction cliffs, 136 are formed by two complexes of the CMGC group, 51 include two complexes of the TK group and 36 by one complex of the CMGC and one of the TK group. It is important to mention that this high percentage is influenced by the number of crystal structures with kinases from these groups (177 and 109 for CMGC and TK, respectively). In the cases where the interaction cliffs are formed by two complexes involving the same kinase, the amino acid sequences are identical in both proteins and thus, the interaction and potency differences are caused mainly by small structural differences in the ligand that can form new interactions with

the target. On the other hand, when interaction cliffs are formed by complexes containing kinases from different groups (or protein with low sequence similarity), then the potency and interaction difference could also be caused by changes in the target sequence, and thus protein structure, that impact ligand binding.

In order to compare this multitarget study with an equivalent single target approach we generated a SAIS map with 97 ligand-CDK2 complexes from the pIC_{50} dataset (Figure S1). This SAIS maps contains 4,656 pairwise comparisons, but only 691 of them with high interaction similarity. Interestingly, the Pearson correlation between the interaction and structural similarity was 0.383 for these CDK2 complexes. This represents a slightly increment compared to the same correlation including all the pIC_{50} data points (which leads to a value of 0.369). In general, the low correlation is caused by pairs of ligand-target complexes in which the structural similarity is lower than the interaction similarity; in other words, different ligands that maintain similar or key interactions with the target. Another differences between the single and the multitarget approach is the fraction of activity cliffs that also are interaction cliffs. For instance, for the CDK2 complexes we identified 267 activity cliffs from which 119 correspond to interaction cliffs (44.6%), whereas in the complete pIC_{50} dataset we found 1,463 activity cliffs and only 280 interaction cliffs (including the 119 of CDK2). However, this difference is not caused by an increased number of interaction cliffs, but can be ascribed to a *decreased number of activity cliffs with low interaction similarity*.

Interpretation of activity and interaction cliffs

Figure 3 shows examples of activity and interaction cliffs in the three data sets. As mentioned above, the inclusion of interaction similarity facilitates the analysis of activity cliffs and helps to

find specific characteristic in the compounds or targets that can drive to more active or selective compounds. From the 280 interaction cliffs found in the data set with pIC_{50} values, pair 2WMR_2X8E is a representative example of similar compounds (ComboScore/2 value of 0.79) that form almost the same interactions with the binding site (interaction similarity of 0.714), but present a potency difference of more than 3.5 log units. Figure 3A shows that in the complex 2WMR the pyrimidinone moiety of **1** forms a hydrogen bond with the $E85^{hinge.46}$ and $C87^{hinge.48}$ amino acids located in the hinge area of CHK1 and also has hydrophobic contact with the gatekeeper amino acid $L84^{GK.45}$ (Figure 3D).[50] The same hydrogen bonds are present in the complex 2X8E by the triazolone of **2**, but in this complex the 4-pyridyl ring is located in the solvent channel of the kinase domain forming hydrophobic interactions with $S88^{linker.49}$ and $G89^{linker.51}$.[51] This difference helps to provide a structure-based hypothesis of the very large potency difference.

Another example of an interaction cliff in this data set is pair 2HYY_3HEC (Figure 3B), which contain the same kinase inhibitor (imatinib **3**) bound to ABL1 and p38α, respectively. Despite the fact that both crystal structures contain the same ligand, the interaction similarity of this pair is 0.75 and present a potency difference of 2.6 log units. In both crystal structures imatinib forms hydrogen bonds between the pyrimidine and the main chain of $M^{hinge.48}$, the NH linker with the $T^{GK.45}$, and the amide moiety of **3** with the side and main chains of $E^{\alpha C.24}$ and $D^{xDFG.81}$, respectively. In addition to these interactions, the complex with ABL1 (2HYY)[52] presents an extra hydrogen bond between the piperazine of **3** and the backbone of $I360^{VI.67}$, and face-to-face π stacking with the $F317^{hinge.47}$. Figure 3D also shows that imatinib forms more hydrophobic interactions in 2HYY compared to 3HEC, which reflects the variation in solvent

accessible surface area reported for ligand bound to ABL1 and p38α (35.4 and 89.6 Å, respectively) that correlates with the potency difference.[53]

Figure 3C shows an example of activity cliff in the same data set with pIC_{50} values. Inhibitors **4** and **5** that form the pair 2R3J_2C69 have very high structural similarity (ComboScore/2 value of 0.83) and a large potency difference of 3.4 log units against CDK2, however the interaction similarity (0.652) is lower than the threshold used to identify interaction cliffs. In this case, differences in interaction patterns are mainly caused by the hydrophobic contacts between the 5-bromophenyl group of **4** and $E12^{g.l.5}$ as well as the hydrogen bond formed between the pyrazolopyrimidine moiety of **4** and the basic amine group of $K33^{III.17}$ in 2C69, that is not formed with pyrazolopyrimidine **5** in 2R3J.[54] Previous studies have shown that the cavity that accepts the 5-bromophenyl and the 5-phenyl present in these compounds is large enough to contain bulky substituents.[55] However, experimental data suggest that the bromide at position 3 of the pyrazolopyrimidine of **5** is responsible of the potency difference as it fills a small hydrophobic cavity formed at the back of the gatekeeper amino acid ($F80^{GK.45}$).[55]

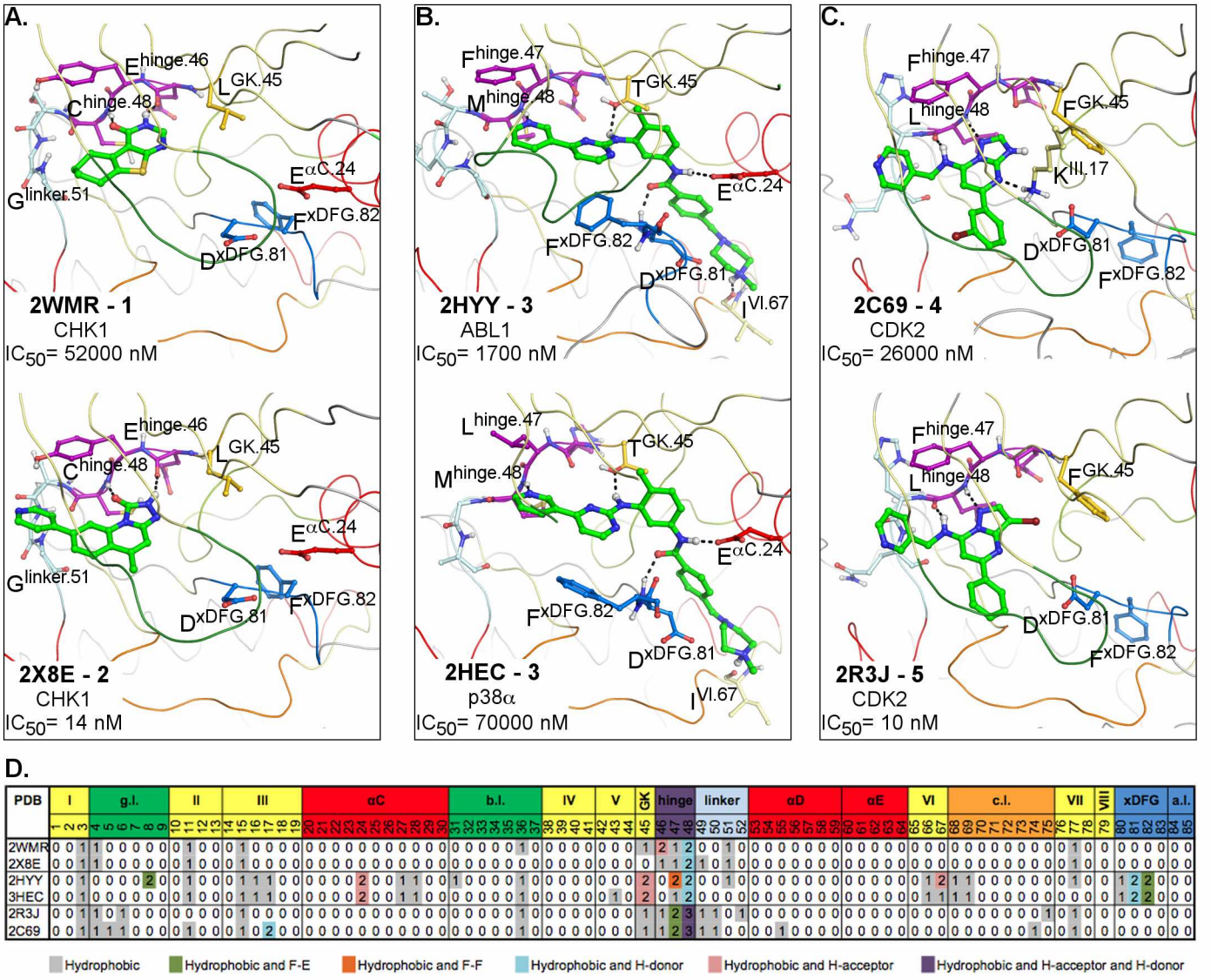


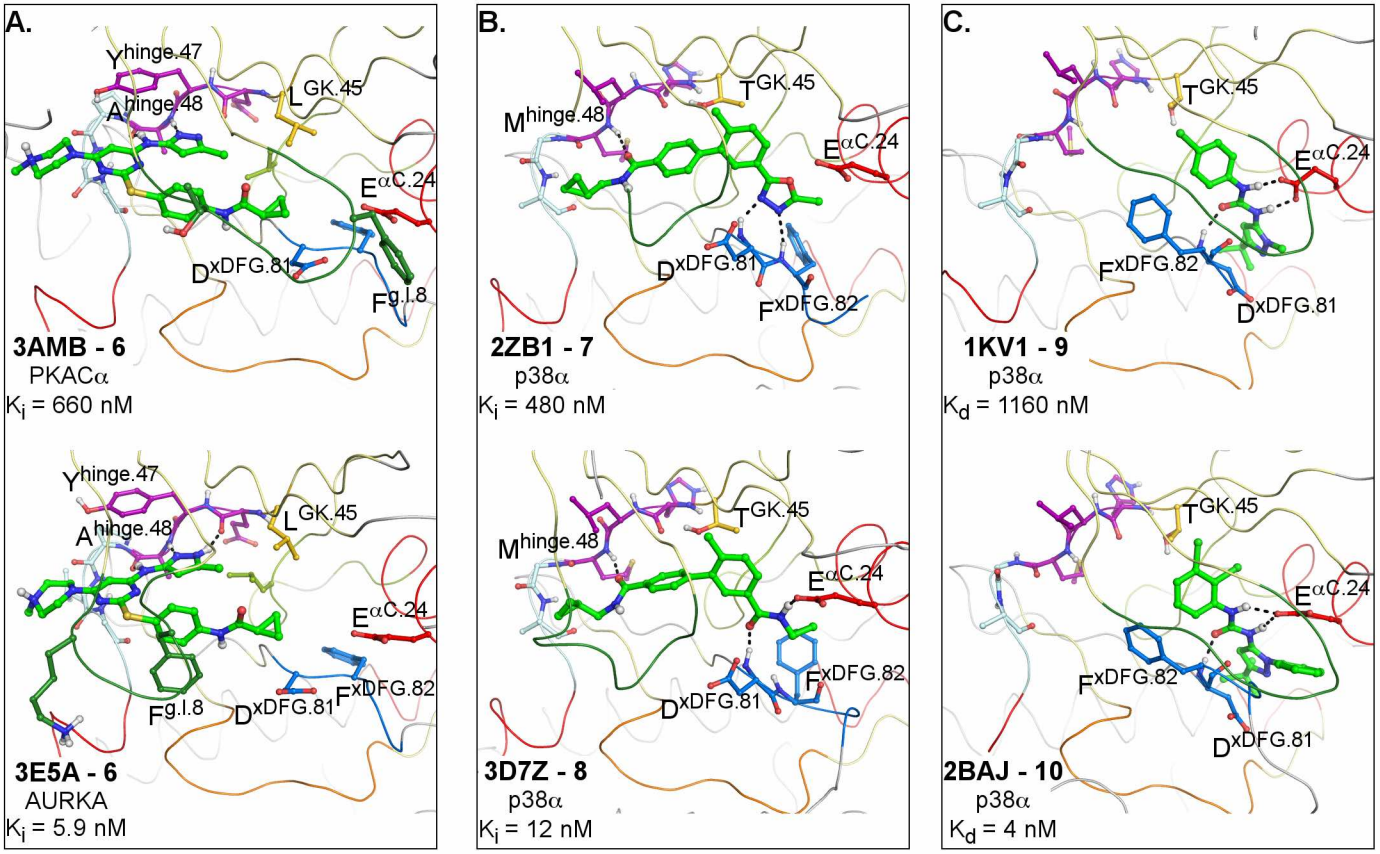
Figure 3. 3D representation and activity data for representative kinase-ligand complexes that form activity and interaction cliffs in the data set with pIC_{50} values: (A) Pyrimidininone **1**[50] and triazolone **2**[51] with CHK1; (B) Imatinib **3** with ABL1[52] and p38α[53]; (C) pyrazolopyrimidines **4**[54] and **5**[55] with CDK2. Panel D shows the number of interactions between the ligand and each of the 85 amino acids in the binding site numbered as defined in KLIFS.[27] Per residue, the total number of interacting bits is reported and color-coded according to (the combination of) interaction feature(s). The position of each pair in SAIS maps is depicted in Figure 2D.

Figures 4A and 4B depict two representative examples of the seven interaction cliffs identified in the data set with pK_i values. The first example (Figure 4A) shows the kinase inhibitor Tozasertib (VX-680) **6** bound to two different kinases, namely aurora kinase (AURKA) and a mutant of cAMP-dependent protein kinase (PKACα). In both crystal structures, **6** forms the same hydrogen bond interactions with the two targets, AURKA (PDB ID: 3E5A)[56] and PKACα (PDB ID: 3AMB).[57] Nevertheless, the π-staking interaction with F144$^{g.l.8}$ and hydrophobic contacts with two more amino acids (G140$^{g.l.4}$ and L194$^{b.l.36}$) in the binding site of AURKA increase the potency by two log units. This result highlights the importance of F$^{g.l.8}$ for the binding of Tozasertib, which can be considered as a "hot spot" to take into account during the design of new AURKA inhibitors.

Figure 4B depicts another example of an interaction cliff, corresponding to the pair 2ZB1_3D7Z (**7** and **8**). In this case, two similar compounds (ComboScore/2 = 0.82) bind to p38α with a similar binding mode (interaction similarity of 0.741), but with large potency difference (1.6 log units). The only difference between ligands **7** and **8** is the substitution of the oxadiazol ring in 2ZB1[58] by a (N-cyclopropyl)carboxamide in 3D7Z.[59] This structural change disturbs the hydrogen bonding pattern with p38α, that is, whereas inhibitor **7** in 2ZB1 forms hydrogen bonds with the backbones of D168$^{xDFG.81}$ and F169$^{xDFG.82}$ of the DFG motif, the compound **8** in 3D7Z forms hydrogen bonds with backbone of D168$^{xDFG.81}$ and with the carboxylate moiety of E71$^{\alpha C.24}$ resulting in an increased potency. In addition, the cyclopropyl group in 3D7Z increases the number of hydrophobic contacts with the binding site.

Finally, only three interaction cliffs were detected in the data set with pK_d values, one of those is the pair 1KV1_2BAJ (**9** and **10**) shown in Figure 4C. This example shows two highly similar pyrazolourea inhibitors **9** and **10** (ComboScore/2 = 0.90) that present similar interactions (0.84)

with p38α, but a potency difference of 2.5 log units. The most important chemical differences between **9** and **10** are the methyl group on the pyrazole ring of **9** in 1KV1[60] versus the phenyl group of **10** in 2BAJ[61] and the 4-chlorophenyl of **9** in 1KV1 versus the 2,3-dichlorophenyl of **10** in 2BAJ. As observed before, a bulky group at this position increments the contacts with the hydrophobic portion of the side chain of E71$^{\alpha C.24}$ in the helix αC. This extra phenyl group targets the hydrophobic back pocket V[27, 62] resulting in an increased activity against p38α.[60]



D.

PDB	I	g.l.	II	III	αC	b.l.	IV	V	GK	hinge	linker	αD	αE	VI	c.l.	VII	VIII	xDFG	a.l.
	1 2 3	4 5 6 7 8 9	10 11 12 13	14 15 16 17 18 19	20 21 22 23 24 25 26 27 28 29 30	31 32 33 34 35 36 37	38 39 40 41	42 43 44	45	46 47 48	49 50 51 52	53 54 55 56 57 58 59	60 61 62 63 64	65 66 67	68 69 70 71 72 73 74 75	76 77 78	79	80 81 82 83	84 85
3E5A	0 0 1	1 0 0 0 2 0	0 1 0 0	0 1 0 0 0	0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 1 0	0 0 0 0	0 0 0	1	1 2 3	1 1 1 1	0 0 0 0 0 0 0	0 0 0 0 0	0 0 0	0 0 0 0 0 0 0 0	0 1 0	0	1 1 0 0	0 0
3AMB	0 0 1	0 0 0 0 0 0	0 1 0 0	0 1 0 1 0 0	0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0	0 0 0	1	1 2 3	1 1 1 1	0 0 0 0 0 0 0	0 0 0 0 0	0 0 0	0 0 0 0 0 0 0 0	0 1 0	0	1 1 0 0	0 0
2ZB1	0 0 0	0 0 0 0 2 0	0 1 0 0	0 1 1 1 0 0	0 0 0 0 1 0 0 0 1 0 0	0 0 0 0 0 1 0	0 0 0 0	0 1 0	1	0 0 1	1 1 0 1	0 0 0 0 0 0 0	0 0 0 0 0	0 0 0	0 0 0 0 0 0 0 0	0 1 0	0	1 2 3 0	1 0
3D7Z	0 0 0	0 0 0 0 1 0	0 1 0 0	0 1 1 1 0 0	0 0 0 0 2 0 0 1 1 0 0	0 0 0 0 0 1 0	0 0 0 0	0 1 0	1	0 1 2	1 1 0 1	0 0 0 0 0 0 0	0 0 0 0 0	0 0 0	0 0 0 0 0 0 0 0	0 1 0	0	1 2 1 0	1 0
1KV1	0 0 0	0 0 0 0 0 0	0 1 0 0	0 1 0 1 0 0	0 0 0 0 2 0 0 1 1 0 0	0 1 0 0 0 1 1 0	0 0 0 0	0 1 0	1	0 0 0	0 0 0 0	0 0 0 0 0 0 0	0 1 0 0 0	0 1 0	1 0 0 0 0 0 0 0	0 0 0	1	1 2 2 0	0 0
2BAJ	0 0 0	0 0 0 0 0 0	0 1 0 0	0 1 1 1 0 0	1 0 0 1 2 0 0 1 1 0 0	1 0 0 0 1 1 0	0 0 0 0	0 1 0	1	0 0 0	0 0 0 0	0 0 0 0 0 0 0	0 1 0 0 0	0 1 0	1 0 0 0 0 0 0 0	0 0 0	1	1 2 2 1	0 0

Hydrophobic | Hydrophobic and F-E | Hydrophobic and F-F | Hydrophobic and H-donor | Hydrophobic and H-acceptor | Hydrophobic and H-acceptor and H-donor

Figure 4. 3D representation and activity data for representative protein-ligand complexes that form activity and interaction cliffs in the data sets with pK_i and pK_d values: (A) Tozasertib **6** with

PKACα[57] and AURKA[56]; (B) Biphenyls **7**[58] and **8**[59] with p38α; (C) pyrazoloureas **9**[60] and **10**[61] with p38α. Panel D shows the number of interactions between the ligand and each of the 85 amino acids in the binding site numbered as defined in KLIFS.[27] Per residue, the total number of interacting bits is reported and color-coded according to (the combination of) interaction feature(s). The position of each pair in SAIS maps is depicted in Figure 2E-F. Figure S2 and Figure S3 show the 2D depiction of these complexes (generated using PoseViewWeb 1.97.0[63, 64]) and the superposed 3D structures of these compounds extracted from the aligned crystal structure, respectively.

Interaction cliff generators

Activity cliff generators are compounds highly associated with activity cliffs in the data set (e.g., above two standard deviations of the mean frequency of activity cliffs) and hence present a high probability to form activity cliffs with other structurally similar molecules tested in the same assay.[19] By analogy to activity cliff generators, interaction cliff generators are ligand-target complexes highly associated with interaction cliffs. A major difference between these two is that interaction cliff generators are more suitable for the analysis of multitarget activity landscapes since they are identified based on the ligand structure, its potency, and how it interacts with the target (which, to some extent, adds target information). In contrast, activity cliff generators do not include any information regarding to the target (only compound structure and potency) and are more suitable for a single-target activity landscapes. It is important to emphasize that the selection of cliff generators is based on comparisons with other compounds in the data set. Hence the larger the data set, the more reliable the generator is. In this study, the most reliable cliff generators are the ones present in the data set with pIC_{50} values, which contains information of 409 ligand-target complexes.

Both activity and interaction cliff generators were identified in the three data sets and are listed with their statistics in Tables S5 and S6. In the pIC_{50} data set, 20 and 21 ligand-target complexes were identified as activity and interaction cliff generators, respectively. One example of interaction cliff generator is complex 2VWV, which forms 10 interaction cliffs. As shown in Table 3, 2VWV contains a low potency ligand bound to EphB4. Interestingly, the ligand in this complex shows high structural and interaction similarity to other ligands bound to different kinases.

Table 3. List of ten ligand-target complexes that form interaction cliffs with the interaction cliff generator 2VWV (included in the list as reference).

PDB ID	Kinase	IC_{50} (nM) of co-crystal ligand	pIC_{50} Difference	Interaction Similarity	Ligand Structure Similarity[a]
2VWV	EphB4	16000	0	1	1
3EKK	INSR	2	3.90	0.739	0.623
3EKN	INSR	2	3.90	0.714	0.607
3CJF	KDR	6.3	3.40	0.750	0.790
3FQS	SYK	41	2.59	0.889	0.879
2NP8	Aurora A	42	2.58	0.739	0.600
1OI9	CDK2	69	2.37	0.696	0.600
3H3C	PYK2	140	2.06	0.682	0.580
2IW6	CDK2	140	2.06	0.696	0.597
2C5N	CDK2	220	1.86	0.714	0.670
2C6K	CDK2	730	1.34	0.680	0.610

[a]Ligand structure similarity is calculated using ComboScore as implemented in ROCS

Identification of activity cliff hot spots

Seebeck *et al*. define activity cliff hot spots as those specific regions or atoms in the target involved in key interactions with the ligand that lead to the formation of an activity cliff.[20] In this paper, activity cliff hot spots were identified in the pIC_{50} data set as the most frequent interactions that appear in the formation of an activity cliff. To this end, only the interactions presented by one of the two compounds forming the cliff were taken into account, discarding the interactions that are presented by both compounds in the pair. From the 595 interactions (seven per each of the 85 amino acids) encoded in the IFPs for this data set, only 116 were present at least once in the formation of activity cliffs. Hydrophobic contact was the most common interaction (frequency of 55), followed by cases where the amino acids are acting as hydrogen bond acceptors or donors (frequency of 21 and 15, respectively) or forming face-to-face/edge interactions (frequency of 7 and 11, respectively). The least frequent interactions were protein anionic and cationic interactions (frequency of five and two times, respectively).

Table 4 lists the interactions that were present in at least 20% of the 1,463 activity cliffs in the pIC_{50} data set. This table also shows the amino acid position, the type of interaction and its frequency either in the most active or least active compound of the molecular pair. Among the most frequent amino acids interacting in activity cliffs we can find those at positions 4-6 which correspond to the glycine-rich loop, position 17 located in β–sheet III at the gate area, positions 46-48 which are in the hinge region, 49-52 from the linker region and 80-81 from the xDFG motif. Interestingly, some of these amino acids are highly conserved, for example a glycine at position 4 and 6 (98% and 100%, respectively, of conservation in KLIFS structures), a lysine at position 17 (100%) and an aspartate at position 81 (99%).[27] It is important to note that some of these amino acids are known to be important for ligand binding, for example the orientation of

$D^{xDFG.81}$ side chain is commonly used (in combination with $F^{xDFG.82}$ and the backbone shift) to define if the kinase is in an active or inactive (DFG-in or DFG-out, respectively) conformation.[65] In this analysis, the most active compound in the activity cliff usually presents a hydrophobic interaction with the amino acid at this position. Another example of an important residue is the one at position 47 in the hinge region. For this residue it is known that many compounds have a face-to-face/edge interaction when there is a phenylalanine at this position, however the interaction is lost when the kinase has a leucine or tyrosine instead. The conserved $K^{III.17}$ also plays a very important role in ligand binding; it forms hydrogen bonds and hydrophobic interactions in 37% and 41% of the activity cliffs, respectively. This conserved amino acid, $K^{III.17}$, commonly forms a hydrogen bond with DFG-out binders, which are known to present improved selectivity and slower dissociative off-rate.[66] It is important to mention that no direct effect or influence of DFG-in or DFG-out target conformations was observed in these activity cliffs.

Table 4. List of most frequent interactions involved in the formation of at least 20% of activity cliffs. The total frequency is further divided in two parts to distinguish if the interaction is more frequent in the most active or least active compound of the molecular pair.

Amino acid position in KLIFS[27]	Type of interaction[a]	Percentage of activity cliffs	Frequency		
			Total	Most active compound	Least active compound
xDFG.81	HYD	45.73	669	515	154
hinge.47	FF	41.63	609	348	261
linker.50	HYD	41.63	609	256	353
linker.51	HYD	40.53	593	385	208

III.17	HYD	40.46	592	340	252
linker.52	HYD	40.33	590	314	276
g.l.4	HYD	39.30	575	375	200
xDFG.80	HYD	36.98	541	323	218
III.17	DON	36.57	535	332	203
linker.49	HYD	36.43	533	263	270
hinge.48	ACC	36.29	531	295	236
hinge.46	HYD	35.13	514	197	317
hinge.46	ACC	32.13	470	273	197
g.l.5	HYD	30.14	441	242	199
c.l.74	HYD	27.89	408	235	173
b.l.36	HYD	27.55	403	216	187
g.l.6	HYD	27.41	401	216	185
αD.55	HYD	26.04	381	140	241
c.l.75	HYD	24.74	362	254	108
hinge.48	DON	20.51	300	189	111
hinge.48	HYD	20.10	294	193	101

[a]HYD=hydrophobic, FF= face-face π-stacking, DON= H-bond donor, ACC= H-bond acceptor

Identification of scaffold hops

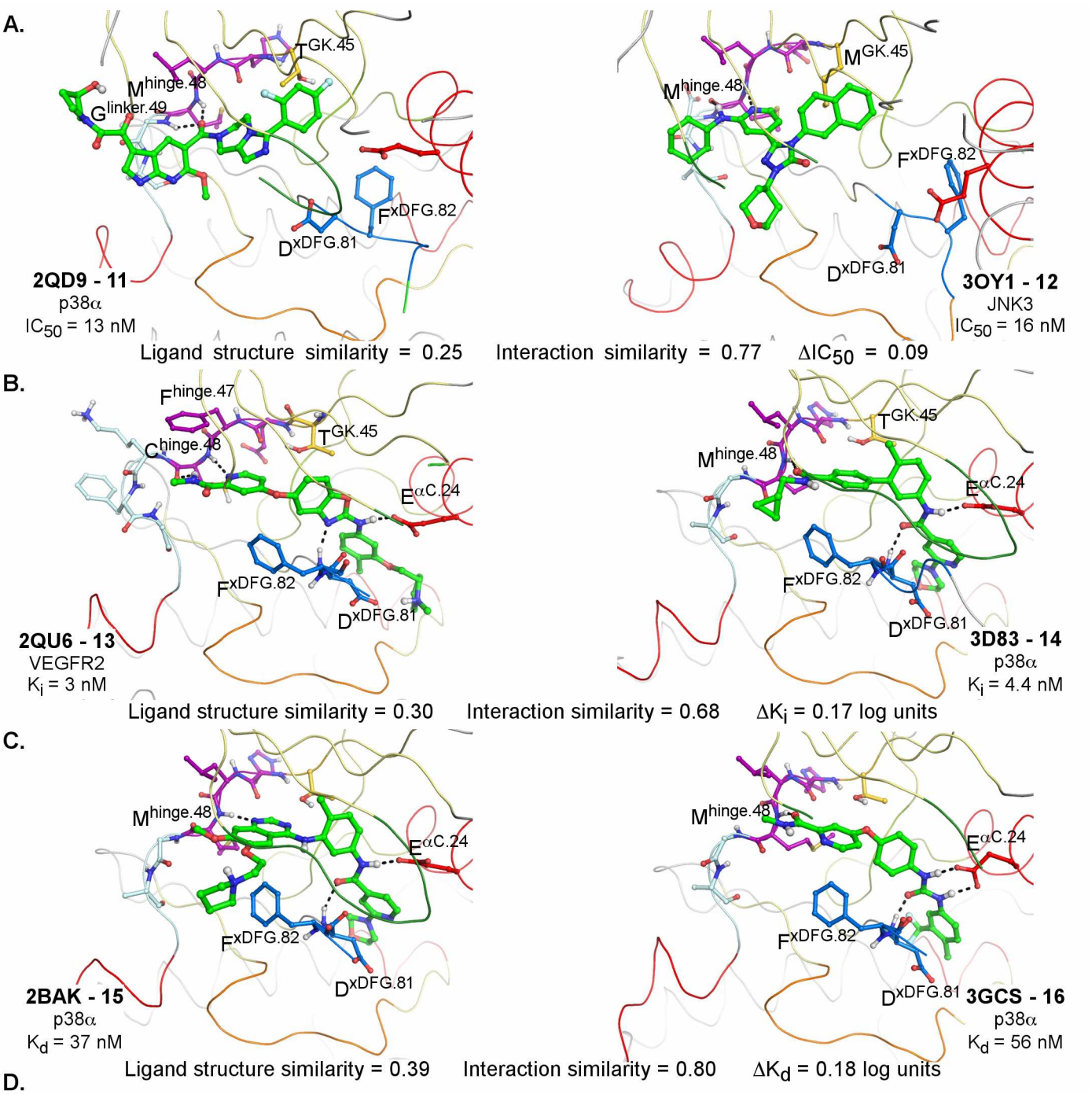
Using the same strategy outlined above, it was possible to identify pairs of compounds that have different chemical structures but similar protein-ligand interactions and similar potency. These pairs of compounds are the so-called *scaffold hops* as the ones presented in Figure 5. Despite the fact that region I in the SAIS maps contains around 40% of the data points, only a few of them (< 7%) have an interaction similarity above the thresholds previously defined (see Table 2 and the Methods section). This is not surprising taking into account the large size of

kinase ATP binding site and the different binding conformation that the inhibitors may have. The analysis of scaffold hops helps to identify conserved interactions that are important for the activity across many kinases to gain promiscuity and also to identify compounds that might bind to a different kinase with similar potency.

In the examples shown in Figure 5, compounds represented by the same data point form similar hydrogen bonds with amino acids at the same positions in both targets. Interestingly, most of the hydrogen bonds are formed with the amino acid backbone or with very conserved amino acids suggesting that these similar interactions can also be formed with other kinases. Remarkably, compounds in the same data pair present conserved hydrophobic interactions suggesting that the binding sites in both targets have similar shape. When looking at the 3D binding conformation of the compounds extracted from the aligned crystal structure it can be observed that the compounds in the same data point overlap to each other, specially at the pharmacophoric points involved in the formation of hydrogen bonds (Figure S4).

Interestingly, from the 878 scaffold hops with similar interaction patterns that were identified in the three data sets, only in 325 the two compounds of the data point target the same kinase (*e.g.* Figure 5C). These 325 scaffold hops correspond to only 24 kinases, where CDK2, p38α, and CHK1 account for 205, 55 and 11 of them, respectively. In the remaining 553 scaffold hops, the two compounds target different kinases with similar potency and presenting similar interaction patterns (*e.g.* Figure 5A and 5B). These cases involve 49 kinases, where CDK2 is again the most prevalent target participating in 179 of the 553 data points, followed by KDR and MET which take part in 76 and 66, respectively. The most common pairs of kinases are CDK2_JNK3 and CDK2_GSK3B, appearing 35 and 20 times, respectively, followed by CDK2_JAK2, KDR_MET and KDR_p38α which appear 15 times each. These five pairs of

kinases represent the 11% of the 878 scaffold hops with high interaction similarity present in the three data sets. It is important to note that the high frequency of CDK2 in these results is influenced by the large number of available crystal structures for this protein kinase (namely 257 in KLIFS of which 216 kinase-ligand complexes).



D.

PDB	I	g.l.	II	III	αC	b.l.	IV	V	GK	hinge	linker	αD	αE	VI	c.l.	VII	VIII	xDFG	a.l.
	1 2 3	4 5 6 7 8 9	10 11 12 13	14 15 16 17 18 19	20 21 22 23 24 25 26 27 28 29 30	31 32 33 34 35 36 37	38 39 40 41	42 43 44	45	46 47 48	49 50 51 52	53 54 55 56 57 58 59	60 61 62 63 64	65 66 67	68 69 70 71 72 73 74 75	76 77 78	79	80 81 82 83	84 85
3GCS	0 0 1	0 0 0 0 0 0	0 1 0 0	0 1 0 1 0 0	0 0 0 1 2 0 0 1 1 0 0	1 0 0 0 1 1 0	0 0 0 0	0 1 0	1	0 1 2	1 0 0 0	0 0 0 0 0 0 0	0 1 0 0 0	0 1 0	1 0 0 0 0 0 0 0	0 0 0	1	1 2 2 0	0 0
2BAK	0 0 1	1 0 0 0 0 0	0 1 0 0	0 1 1 1 0 0	0 0 0 2 0 0 1 1 0 0	1 0 0 0 1 1 0	0 0 0 0	0 1 0	1	1 1 2	0 0 0 0	0 0 0 0 0 0 0	0 1 0 0 0	0 0 0	1 0 0 0 0 0 0 0	0 0 0	1	1 2 2 0	0 0
2QU6	0 0 1	0 0 0 0 0 0	0 1 0 0	0 1 0 1 0 0	0 0 0 2 0 0 1 1 0 0	0 0 0 0 1 1 0	0 0 0 0	0 0 0	1	1 2 3	1 0 1 0	0 0 0 0 0 0 0	0 1 0 0 0	0 1 1	1 1 0 0 0 0 0 0	0 1 0	1	1 2 2 0	0 0
3D83	0 0 1	0 0 0 0 0 0	0 1 0 0	0 1 1 1 0 0	0 0 0 2 0 0 1 1 0 0	1 0 0 0 1 1 0	0 0 0 0	0 0 0	1	0 1 2	0 0 0 0	0 0 0 0 0 0 0	0 1 0 0 0	0 1 0	1 0 0 0 0 0 0 0	0 0 0	1	1 2 2 0	0 0
2QD9	0 0 1	1 0 0 0 0 0	0 1 0 0	0 1 1 1 0 0	0 0 0 0 0 0 0 0 0 1 0 0	0 0 0 0 0 1 0	1 0 0 0	0 1 1	1	0 1 2	2 1 1 0	0 0 0 0 0 0 0	0 0 0 0 0	0 0 0	0 0 0 0 0 0 0 0	1 0 1	1 0	0 1 0 0	0 0
3OY1	0 0 1	1 1 0 0 0 0	0 1 0 0	0 1 1 1 0 0	0 0 0 0 0 0 0 0 0 1 0 0	0 0 0 0 0 1 0	1 0 0 0	0 1 1	1	0 1 3	1 1 0 1	0 0 0 0 0 0 0	0 0 0 0 0	0 0 0	0 0 0 0 0 0 0 0	1 0 1	0	1 0 0 0	0 0

Hydrophobic | Hydrophobic and F-E | Hydrophobic and F-F | Hydrophobic and H-donor | Hydrophobic and H-acceptor | Hydrophobic and H-acceptor and H-donor

Figure 5. 3D representation and activity data for representative protein-ligand complexes identified as scaffold hops, namely (A) 2QD9_3OY1 (**11** with p38α[67] and **12** with JNK3[68]), (B) 2QU6_3D83 (**13** with VEGFR2[69] and **14** with p38α[70]), (C) 2BAK_3GCS (**15**[61] and **16**[71] with p38α). Potency difference, structure (ComboScore/2) and interaction similarity values are shown for each molecular pair. Panel D shows the number of interactions between the ligand and each of the 85 amino acids in the binding site numbered as defined in KLIFS.[27] Per residue, the total number of interacting bits is reported and color-coded according to (the combination of) interaction feature(s). Figure S4 shows the 2D depiction of these complexes (generated using PoseViewWeb 1.97.0[63, 64]) and the superposed 3D structures of these compounds extracted from the aligned crystal structure.

CONCLUSIONS

This paper discusses the applications and advantages of including molecular interaction fingerprints of protein-ligand complexes in activity landscape modelling. To this end, KLIFS[27] was divided in three different data sets, depending on the activity data available (pIC_{50}, pK_i, pK_d) in the MOAD database. The activity landscape of each data set was modelled using Structure-Activity-Interaction Similarity (SAIS) maps, which show the relationships between ligands 3D binding conformation, how they interact with the target, and the resulting biological activity. In general, the compounds in this study presented low structural and interaction similarity (ranging from 0.366 to 0.383 and from 0.384 to 0.425, respectively) accompanied by large potency differences (up to 7.5 log units of difference). It is noteworthy that only less than 6% of the data points of each data set presented high interaction similarity.

The use of IFPs did not only facilitate the structure-based interpretation of activity cliffs, but also allowed the identification of the “interaction cliffs” which are introduced in this work as

pairs of compounds that have a high (3D and/or 2D) structural and protein-ligand interaction similarity, but a large potency difference. On average, only 25% of the activity cliffs were also considered as interaction cliffs in the three datasets. Additionally, the information extracted from IFPs allowed the identification of activity cliff hot spots, where the hydrophobic contacts with the $K^{III.17}$ and $D^{xDFG.81}$ and the face-face interaction with $F^{hinge.47}$ seem to be involved in the formation of activity cliffs. Also, it was possible to identify scaffold hops with similar protein-ligand interactions and similar potency. Only less than 7% of compounds with similar potency but different molecular structure presented high interaction similarity.

Taken the results together, this paper shows that the added information given by the interaction fingerprints is very valuable to understand and rationalize activity cliffs from both the ligand and target point of view. However, the use of IFPs in activity landscape modelling is not restricted to the SAIS maps, and this opens up interesting perspectives and challenges. For example, the information encoded by IFPs can be incorporated in other activity landscape methods, either quantitative (e.g. SALI[72] and SARI[73]) or qualitative, or in the activity landscape modelling of other data sets with structural information as the one used by Desaphy *et al.*[21] In conclusion, IFPs represent a useful technique to extract valuable information from the ligand-target complex when used in the context of activity landscapes.

ASSOCIATED CONTENT

Supporting Information. Additional tables mentioned in the text and compounds structures. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*Phone: +44-1223-763-078. E-mail: om268@cam.ac.uk; Address: Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom. (O. Méndez-Lucio).

Phone: +(52)55-5622-5287; Fax: +(52)55-5622-5287. E-mail: medinajl@unam.mx; Address: Facultad de Química, Departamento de Farmacia, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, Mexico. (JL. Medina-Franco).

ACKNOWLEDGMENT

Authors thank OpenEye Scientific Software, Inc. for providing ROCS. O. M-L is very grateful to CONACyT (No. 217442/312933) and the Cambridge Overseas Trust for funding. AB thanks Unilever for funding and the European Research Council for a Starting Grant (ERC-2013-StG-336159 MIXTURE). This work was supported by a scholarship from the Secretariat of Public Education and the Mexican government.

ABBREVIATIONS

IFPs, interaction fingerprints; SAR, structure-activity relationship; SAS, structure-activity similarity; SAIS, structure-activity-interaction similarity.

REFERENCES

(1) Wassermann, A. M.; Wawer, M.; Bajorath, J. r., Activity Landscape Representations for Structure−Activity Relationship Analysis. *J. Med. Chem.* **2010**, 53, 8209-8223.
(2) Bajorath, J.; Peltason, L.; Wawer, M.; Guha, R.; Lajiness, M. S.; Van Drie, J. H., Navigating structure-activity landscapes. *Drug Discovery Today* **2009**, 14, 698-705.
(3) Shanmugasundaram, V.; Maggiora, G. M. Characterizing property and activity landscapes using an information-theoretic approach. In 222nd ACS National Meeting, Chicago, IL, USA, 2001; American Chemical Society, Washington, DC: Chicago, IL, USA, 2001.
(4) Medina-Franco, J. L., Scanning Structure–Activity Relationships with Structure–Activity Similarity and Related Maps: From Consensus Activity Cliffs to Selectivity Switches. *J. Chem. Inf. Model.* **2012**, 52, 2485-2493.
(5) Waddell, J.; Medina-Franco, J. L., Bioactivity landscape modeling: Chemoinformatic characterization of structure–activity relationships of compounds tested across multiple targets. *Bioorg. Med. Chem.* **2012**, 20, 5443-5452.
(6) Pérez-Villanueva, J.; Santos, R.; Hernandez-Campos, A.; Giulianotti, M. A.; Castillo, R.; Medina-Franco, J. L., Structure-activity relationships of benzimidazole derivatives as antiparasitic agents: Dual activity-difference (DAD) maps. *Med. Chem. Commun.* **2010**, 2, 44-49.
(7) Méndez-Lucio, O.; Pérez-Villanueva, J.; Castillo, R.; Medina-Franco, J. L., Activity landscape modeling of PPAR ligands with dual-activity difference maps. *Bioorg. Med. Chem.* **2012**, 20, 3523-3532.
(8) Wawer, M.; Peltason, L.; Weskamp, N.; Teckentrup, A.; Bajorath, J. r., Structure−Activity Relationship Anatomy by Network-like Similarity Graphs and Local Structure−Activity Relationship Indices. *J. Med. Chem.* **2008**, 51, 6075-6084.
(9) Guha, R., Exploring structure–activity data using the landscape paradigm. *WIREs Comput. Mol. Sci.* **2012**, 2, 829-841.
(10) Maggiora, G. M., On outliers and activity cliffs – Why QSAR often disappoints. *J. Chem. Inf. Model.* **2006**, 46, 1535.
(11) Medina-Franco, J. L., Activity Cliffs: Facts or Artifacts? *Chem. Biol. Drug Des.* **2013**, 81, 553-556.
(12) Pérez-Villanueva, J.; Medina-Franco, J. L.; Méndez-Lucio, O.; Yoo, J.; Soria-Arteche, O.; Izquierdo, T.; Lozada, M. C.; Castillo, R., CASE Plots for the Chemotype-Based Activity and Selectivity Analysis: A CASE Study of Cyclooxygenase Inhibitors. *Chem. Biol. Drug Des.* **2012**, 80, 752-762.
(13) Hernández-Vázquez, E.; Méndez-Lucio, O.; Hernández-Luis, F., Activity landscape analysis, CoMFA and CoMSIA studies of pyrazole CB1 antagonists. *Med Chem Res* **2013**, 22, 4133-4145.
(14) Pérez-Villanueva, J.; Santos, R.; Hernández-Campos, A.; Giulianotti, M. A.; Castillo, R.; Medina-Franco, J. L., Towards a systematic characterization of the antiprotozoal activity landscape of benzimidazole derivatives. *Bioorg. Med. Chem.* **2010**, 18, 7380-7391.
(15) Medina-Franco, J. L.; Yongye, A. B.; Pérez-Villanueva, J.; Houghten, R. A.; Martínez-Mayorga, K., Multitarget structure–activity relationships characterized by activity-difference maps and consensus similarity measure. *J. Chem. Inf. Model.* **2011**, 51, 2427-2439.
(16) de la Vega de León, A.; Bajorath, J., Design of a Three-Dimensional Multitarget Activity Landscape. *J. Chem. Inf. Model.* **2012**, 52, 2876-2883.

(17) Cruz-Monteagudo, M.; Medina-Franco, J. L.; Pérez-Castillo, Y.; Nicolotti, O.; Cordeiro, M. N. D. S.; Borges, F., Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? *Drug Discovery Today* **2014**, 19, 1069-80.
(18) Medina-Franco, J.; Méndez-Lucio, O.; Yoo, J., Rationalization of Activity Cliffs of a Sulfonamide Inhibitor of DNA Methyltransferases with Induced-Fit Docking. *Int. J. Mol. Sci.* **2014**, 15, 3253-3261.
(19) Méndez-Lucio, O.; Pérez-Villanueva, J.; Castillo, R.; Medina-Franco, J. L., Identifying Activity Cliff Generators of PPAR Ligands Using SAS Maps. *Mol. Inf.* **2012**, 31, 837-846.
(20) Seebeck, B.; Wagener, M.; Rarey, M., From Activity Cliffs to Target-Specific Scoring Models and Pharmacophore Hypotheses. *ChemMedChem* **2011**, 6, 1630-1639.
(21) Desaphy, J.; Raimbaud, E.; Ducrot, P.; Rognan, D., Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. *J. Chem. Inf. Model.* **2013**, 53, 623-637.
(22) Brewerton, S. C., The use of protein-ligand interaction fingerprints in docking. *Curr. Opin. Drug Discovery Dev.* **2008**, 11, 356-364.
(23) de Graaf, C.; Kooistra, A. J.; Vischer, H. F.; Katritch, V.; Kuijer, M.; Shiroishi, M.; Iwata, S.; Shimamura, T.; Stevens, R. C.; de Esch, I. J. P.; Leurs, R., Crystal Structure-Based Virtual Screening for Fragment-like Ligands of the Human Histamine H1 Receptor. *J. Med. Chem.* **2011**, 54, 8195-8206.
(24) de Graaf, C.; Rein, C.; Piwnica, D.; Giordanetto, F.; Rognan, D., Structure-Based Discovery of Allosteric Modulators of Two Related Class B G-Protein-Coupled Receptors. *ChemMedChem* **2011**, 6, 2159-2169.
(25) de Graaf, C.; Vischer, H. F.; de Kloe, G. E.; Kooistra, A. J.; Nijmeijer, S.; Kuijer, M.; Verheij, M. H. P.; England, P. J.; van Muijlwijk-Koezen, J. E.; Leurs, R.; de Esch, I. J. P., Small and colorful stones make beautiful mosaics: fragment-based chemogenomics. *Drug Discovery Today* **2013**, 18, 323-330.
(26) Kooistra, A. J.; Kuhne, S.; de Esch, I. J. P.; Leurs, R.; de Graaf, C., A structural chemogenomics analysis of aminergic GPCRs: lessons for histamine receptor ligand design. *Br. J. Pharmacol.* **2013**, 170, 101-126.
(27) van Linden, O. P.; Kooistra, A. J.; Leurs, R.; de Esch, I. J.; de Graaf, C., KLIFS: a knowledge-based structural database to navigate kinase-ligand interaction space. *J Med Chem* **2014**, 57, 249-77.
(28) Furtmann, N.; Hu, Y.; Bajorath, J., Comprehensive Analysis of Three-Dimensional Activity Cliffs Formed by Kinase Inhibitors with Different Binding Modes and Cliff Mapping of Structural Analogs. *J. Med. Chem.* **2014**, DOI: 10.1021/jm5009264.
(29) Benson, M. L.; Smith, R. D.; Khazanov, N. A.; Dimcheff, B.; Beaver, J.; Dresslar, P.; Nerothin, J.; Carlson, H. A., Binding MOAD, a high-quality protein–ligand database. *Nucleic Acids Res.* **2008**, 36, D674-D678.
(30) Liegi Hu, M. L. B., Richard D. Smith, Michael G. Lerner, Heather A. Carlson,, Binding MOAD (Mother Of All Databases). *Proteins: Struct., Funct., Bioinf.* **2005**, 60, 333-340.
(31) Sud, M. MayaChemTools URL: http://www.MayaChemTools.org (accessed Feb. 2013).
(32) Filimonov, D.; Poroikov, V.; Borodina, Y.; Gloriozova, T., Chemical similarity assessment through multilevel neighborhoods of atoms:□ definition and comparison with the other descriptors. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 666-670.
(33) Hall, L. H.; Kier, L. B., Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 1039-1045.

(34) Rogers, D.; Hahn, M., Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, 50, 742-754.
(35) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G., Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **2002**, 42, 1273-1280.
(36) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R., Atom pairs as molecular features in structure-activity studies: definition and applications. *J. Chem. Inform. Comput. Sci.* **1985**, 25, 64-73.
(37) Nilakantan, R.; Bauman, N.; Dixon, J. S.; Venkataraghavan, R., Topological torsion: a new molecular descriptor for SAR applications. Comparison with other descriptors. *J. Chem. Inf. Comput. Sci.* **1987**, 27, 82-85.
(38) Renner, S.; Fechner, U.; Schneider, G. Alignment-free pharmacophore patterns – A correlation-vector approach. In *Pharmacophores and Pharmacophore Searches*; Wiley-VCH Verlag GmbH & Co. KGaA: 2006, pp 49-79.
(39) Bonachéra, F.; Parent, B.; Barbosa, F.; Froloff, N.; Horvath, D., Fuzzy tricentric pharmacophore fingerprints. 1. Topological fuzzy pharmacophore triplets and adapted molecular similarity scoring schemes. *J. Chem. Inf. Model.* **2006**, 46, 2457-2477.
(40) Jaccard, P., Etude Comparative de la Distribution Florale dans une Portion des Alpes et des Jura. *Bull. Soc. Vaudoise Sci. Nat.* **1901**, 37, 547-579.
(41) Willett, P.; Barnard, J. M.; Downs, G. M., Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 983-996.
(42) Bender, A.; Glen, R. C., Molecular similarity: a key technique in molecular informatics. *Org. Biomol. Chem.* **2004**, 2, 3204-3218.
(43) ROCS ROCS 3.2.0.4, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, http://www.eyesopen.com.
(44) Marcou, G.; Rognan, D., Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *J. Chem. Inf. Model.* **2006**, 47, 195-207.
(45) Medina-Franco, J. L.; Martínez-Mayorga, K.; Bender, A.; Marín, R. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A., Characterization of Activity Landscapes Using 2D and 3D Similarity Methods: Consensus Activity Cliffs. *J. Chem. Inf. Model.* **2009**, 49, 477-491.
(46) Yongye, A. B.; Byler, K.; Santos, R.; Martínez-Mayorga, K.; Maggiora, G. M.; Medina-Franco, J. L., Consensus models of activity landscapes with multiple chemical, conformer, and property representations. *J. Chem. Inf. Model.* **2011**, 51, 1259-1270.
(47) Sun, H. M.; Tawa, G.; Wallqvist, A., Classification of scaffold-hopping approaches. *Drug Discovery Today* **2012**, 17, 310-324.
(48) Pérez-Villanueva, J.; Méndez-Lucio, O.; Soria-Arteche, O.; Izquierdo, T.; Concepción Lozada, M.; Gloria-Greimel, W. A.; Medina-Franco, J. L., Cyclic Systems Distribution Along Similarity Measures: Insights for an Application to Activity Landscape Modeling. *Mol. Inf.* **2013**, 32, 179-190.
(49) Medina-Franco, J. L.; Yongye, A. B.; López-Vallejo, F. Consensus Models of Activity Landscapes. In *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*; Wiley-VCH Verlag GmbH & Co. KGaA: 2012, pp 307-326.
(50) Matthews, T. P.; Klair, S.; Burns, S.; Boxall, K.; Cherry, M.; Fisher, M.; Westwood, I. M.; Walton, M. I.; McHardy, T.; Cheung, K.-M. J.; Van Montfort, R.; Williams, D.; Aherne, G. W.; Garrett, M. D.; Reader, J.; Collins, I., Identification of Inhibitors of Checkpoint Kinase 1 through Template Screening. *J. Med. Chem.* **2009**, 52, 4810-4819.

(51) Oza, V.; Ashwell, S.; Brassil, P.; Breed, J.; Deng, C.; Ezhuthachan, J.; Haye, H.; Horn, C.; Janetka, J.; Lyne, P.; Newcombe, N.; Otterbien, L.; Pass, M.; Read, J.; Roswell, S.; Su, M.; Toader, D.; Yu, D.; Yu, Y.; Valentine, A.; Webborn, P.; White, A.; Zabludoff, S.; Zheng, X., Discovery of a novel class of triazolones as Checkpoint Kinase inhibitors—Hit to lead exploration. *Bioorg. Med. Chem. Lett.* **2010**, 20, 5133-5138.
(52) Cowan-Jacob, S. W.; Fendrich, G.; Floersheimer, A.; Furet, P.; Liebetanz, J.; Rummel, G.; Rheinberger, P.; Centeleghe, M.; Fabbro, D.; Manley, P. W., Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, 63, 80-93.
(53) Namboodiri, H. V.; Bukhtiyarova, M.; Ramcharan, J.; Karpusas, M.; Lee, Y.; Springman, E. B., Analysis of Imatinib and Sorafenib Binding to p38α Compared with c-Abl and b-Raf Provides Structural Insights for Understanding the Selectivity of Inhibitors Targeting the DFG-Out Form of Protein Kinases. *Biochemistry* **2010**, 49, 3611-3618.
(54) Richardson, C. M.; Williamson, D. S.; Parratt, M. J.; Borgognoni, J.; Cansfield, A. D.; Dokurno, P.; Francis, G. L.; Howes, R.; Moore, J. D.; Murray, J. B.; Robertson, A.; Surgenor, A. E.; Torrance, C. J., Triazolo[1,5-a]pyrimidines as novel CDK2 inhibitors: Protein structure-guided design and SAR. *Bioorg. Med. Chem. Lett.* **2006**, 16, 1353-1357.
(55) Fischmann, T. O.; Hruza, A.; Duca, J. S.; Ramanathan, L.; Mayhood, T.; Windsor, W. T.; Le, H. V.; Guzi, T. J.; Dwyer, M. P.; Paruch, K.; Doll, R. J.; Lees, E.; Parry, D.; Seghezzi, W.; Madison, V., Structure-guided discovery of cyclin-dependent kinase inhibitors. *Biopolymers* **2008**, 89, 372-379.
(56) Zhao, B.; Smallwood, A.; Yang, J.; Koretke, K.; Nurse, K.; Calamari, A.; Kirkpatrick, R. B.; Lai, Z., Modulation of kinase-inhibitor interactions by auxiliary protein binding: Crystallography studies on Aurora A interactions with VX-680 and with TPX2. *Protein Sci.* **2008**, 17, 1791-1797.
(57) Pflug, A.; de Oliveira, T. M.; Bossemeyer, D.; Engh, R. A., Mutants of protein kinase A that mimic the ATP-binding site of Aurora kinase. *Biochem. J.* **2011**, 440, 85-93.
(58) Angell, R. M.; Bamborough, P.; Cleasby, A.; Cockerill, S. G.; Jones, K. L.; Mooney, C. J.; Somers, D. O.; Walker, A. L., Biphenyl amide p38 kinase inhibitors 1: Discovery and binding mode. *Bioorg. Med. Chem. Lett.* **2008**, 18, 318-323.
(59) Angell, R.; Aston, N. M.; Bamborough, P.; Buckton, J. B.; Cockerill, S.; deBoeck, S. J.; Edwards, C. D.; Holmes, D. S.; Jones, K. L.; Laine, D. I.; Patel, S.; Smee, P. A.; Smith, K. J.; Somers, D. O.; Walker, A. L., Biphenyl amide p38 kinase inhibitors 3: Improvement of cellular and in vivo activity. *Bioorg. Med. Chem. Lett.* **2008**, 18, 4428-4432.
(60) Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J., Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Mol. Biol.* **2002**, 9, 268-272.
(61) Sullivan, J. E.; Holdgate, G. A.; Campbell, D.; Timms, D.; Gerhardt, S.; Breed, J.; Breeze, A. L.; Bermingham, A.; Pauptit, R. A.; Norman, R. A.; Embrey, K. J.; Read, J.; VanScyoc, W. S.; Ward, W. H. J., Prevention of MKK6-Dependent Activation by Binding to p38α MAP Kinase‡. *Biochemistry* **2005**, 44, 16475-16490.
(62) Liao, J. J.-L., Molecular Recognition of Protein Kinase Binding Pockets for Design of Potent and Selective Kinase Inhibitors. *J. Med. Chem.* **2007**, 50, 409-424.
(63) Stierand, K.; Maaß, P. C.; Rarey, M., Molecular complexes at a glance: automated generation of two-dimensional complex diagrams. *Bioinformatics* **2006**, 22, 1710-1716.
(64) PoseViewWeb 1.97.0 URL: http://poseview.zbh.uni-hamburg.de (accessed Feb. 2014).

(65) Liao, J. J.-L.; Andrews, R. C., Targeting Protein Multiple Conformations: A Structure-Based Strategy for Kinase Drug Design. *Curr. Top. Med. Chem.* **2007**, 7, 1394-1407.
(66) Blanc, J.; Geney, R.; Menet, C., Type II kinase inhibitors: an opportunity in cancer for rational design. *Anticancer Agents Med. Chem.* **2013**, 13, 731-47.
(67) Murali Dhar, T. G.; Wrobleski, S. T.; Lin, S.; Furch, J. A.; Nirschl, D. S.; Fan, Y.; Todderud, G.; Pitt, S.; Doweyko, A. M.; Sack, J. S.; Mathur, A.; McKinnon, M.; Barrish, J. C.; Dodd, J. H.; Schieven, G. L.; Leftheris, K., Synthesis and SAR of p38α MAP kinase inhibitors based on heterobicyclic scaffolds. *Bioorg. Med. Chem. Lett.* **2007**, 17, 5019-5024.
(68) Probst, G. D.; Bowers, S.; Sealy, J. M.; Truong, A. P.; Hom, R. K.; Galemmo Jr, R. A.; Konradi, A. W.; Sham, H. L.; Quincy, D. A.; Pan, H.; Yao, N.; Lin, M.; Tóth, G.; Artis, D. R.; Zmolek, W.; Wong, K.; Qin, A.; Lorentzen, C.; Nakamura, D. F.; Quinn, K. P.; Sauer, J.-M.; Powell, K.; Ruslim, L.; Wright, S.; Chereau, D.; Ren, Z.; Anderson, J. P.; Bard, F.; Yednock, T. A.; Griswold-Prenner, I., Highly selective c-Jun N-terminal kinase (JNK) 2 and 3 inhibitors with in vitro CNS-like pharmacokinetic properties prevent neurodegeneration. *Bioorg. Med. Chem. Lett.* **2011**, 21, 315-319.
(69) Potashman, M. H.; Bready, J.; Coxon, A.; DeMelfi, T. M.; DiPietro, L.; Doerr, N.; Elbaum, D.; Estrada, J.; Gallant, P.; Germain, J.; Gu, Y.; Harmange, J.-C.; Kaufman, S. A.; Kendall, R.; Kim, J. L.; Kumar, G. N.; Long, A. M.; Neervannan, S.; Patel, V. F.; Polverino, A.; Rose, P.; van der Plas, S.; Whittington, D.; Zanon, R.; Zhao, H., Design, Synthesis, and Evaluation of Orally Active Benzimidazoles and Benzoxazoles as Vascular Endothelial Growth Factor-2 Receptor Tyrosine Kinase Inhibitors. *J. Med. Chem.* **2007**, 50, 4351-4373.
(70) Angell, R. M.; Angell, T. D.; Bamborough, P.; Bamford, M. J.; Chung, C.-w.; Cockerill, S. G.; Flack, S. S.; Jones, K. L.; Laine, D. I.; Longstaff, T.; Ludbrook, S.; Pearson, R.; Smith, K. J.; Smee, P. A.; Somers, D. O.; Walker, A. L., Biphenyl amide p38 kinase inhibitors 4: DFG-in and DFG-out binding modes. *Bioorg. Med. Chem. Lett.* **2008**, 18, 4433-4437.
(71) Simard, J. R.; Getlik, M.; Grütter, C.; Pawar, V.; Wulfert, S.; Rabiller, M.; Rauh, D., Development of a Fluorescent-Tagged Kinase Assay System for the Detection and Characterization of Allosteric Kinase Inhibitors. *J. Am. Chem. Soc.* **2009**, 131, 13286-13296.
(72) Guha, R.; Van Drie, J. H., Structure−Activity Landscape Index:□ Identifying and Quantifying Activity Cliffs. *J. Chem. Inf. Model.* **2008**, 48, 646-658.
(73) Peltason, L.; Bajorath, J., SAR Index:□ Quantifying the Nature of Structure−Activity Relationships. *J. Med. Chem.* **2007**, 50, 5571-5578.

For Table of Contents use only

Analysing Multitarget Activity Landscapes using Protein-Ligand Interaction Fingerprints: *Interaction cliffs*

Oscar Méndez-Lucio,[1,*] Albert J. Kooistra,[2] Chris de Graaf,[2] Andreas Bender[1] and José L. Medina-Franco[3,*]