

# Fragment Hotspot Mapping to Identify Selectivity-Determining Regions between Related Proteins

Mihaela D. Smilova,\* Peter R. Curran, Chris J. Radoux, Frank von Delft, Jason C. Cole, Anthony R. Bradley, and Brian D. Marsden\*



Cite This: *J. Chem. Inf. Model.* 2022, 62, 284–294



Read Online

ACCESS |



Metrics & More

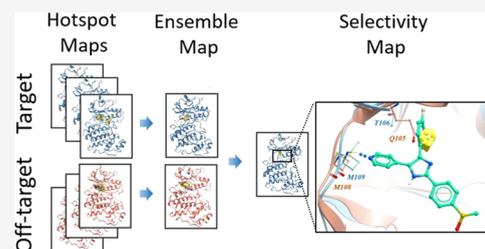


Article Recommendations



Supporting Information

**ABSTRACT:** Selectivity is a crucial property in small molecule development. Binding site comparisons within a protein family are a key piece of information when aiming to modulate the selectivity profile of a compound. Binding site differences can be exploited to confer selectivity for a specific target, while shared areas can provide insights into polypharmacology. As the quantity of structural data grows, automated methods are needed to process, summarize, and present these data to users. We present a computational method that provides quantitative and data-driven summaries of the available binding site information from an ensemble of structures of the same protein. The resulting ensemble maps identify the key interactions important for ligand binding in the ensemble. The comparison of ensemble maps of related proteins enables the identification of selectivity-determining regions within a protein family. We applied the method to three examples from the well-researched human bromodomain and kinase families, demonstrating that the method is able to identify selectivity-determining regions that have been used to introduce selectivity in past drug discovery campaigns. We then illustrate how the resulting maps can be used to automate comparisons across a target protein family.



## INTRODUCTION

The past decade has seen an explosion in the availability of genomic and structural data for a great number of biomolecular disease targets.<sup>1,2</sup> Rational drug discovery aims to use this knowledge to design chemical and biological agents that modulate target activity.<sup>2</sup> Despite recent technological advances, the development of such agents is exceedingly expensive, is not routine, and carries the risk of failure of compound development program after many years. There is a great need for methods that can streamline and automate the process of developing novel drugs and small molecule probes—the highly selective chemical agents used to investigate the underlying biology of disease.<sup>2</sup> Fragment-based drug discovery (FBDD) has established itself as a powerful tool to identify compounds as the starting point for probe and drug development.<sup>3,4</sup> Fragment hits form a small number of high-quality interactions with the protein, and compared to larger compounds traditionally used in high-throughput screens, they have been shown to provide greater hit rates resulting in more efficient exploration of relevant chemical spaces.<sup>5</sup> A critical aspect of FBDD is the elaboration of the initial, low-affinity fragment hits in a stepwise and rationally guided manner. In the past two decades, FBDD approaches have contributed four approved drugs,<sup>6–9</sup> and tens of FBDD-derived compounds are currently in clinical trials.<sup>10</sup> FBDD approaches have also influenced and enhanced drug discovery efforts that did not start from an experimental fragment screen. Recently, the potency of an existing

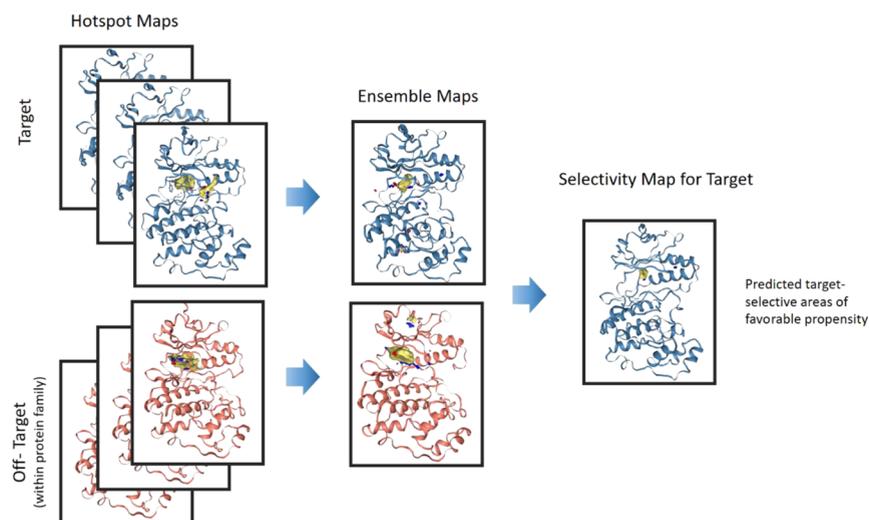
transition-state analogue lead was enhanced by incorporating insights from FBDD, demonstrating the utility of using the latter synergistically to add value to existing projects.<sup>11</sup> Such examples of “fragment-assisted” drug discovery are becoming increasingly common.<sup>10,12</sup>

The use of methods to predict fragment hotspots, regions within the protein’s binding site that make a disproportionately large contribution to binding affinity,<sup>13,14</sup> has been reported in the literature<sup>15–18</sup> and used both to determine potentially tractable pockets and subpockets on the protein surface and to guide the rational design of inhibitors.<sup>19</sup> In 2016, Radoux et al. introduced a method for hotspot mapping based on the wealth of structural interaction data in the Cambridge Structural Database (CSD),<sup>20</sup> which takes the molecular context of fragment binding into account. This method is a promising way for guiding the rational design of inhibitors as the maps provide an intuitive visual guide to favorable interactions within the binding site and can indicate suboptimal interactions within the original hit. The maps also give an objective numerical understanding of the features important

Received: July 9, 2021

Published: January 12, 2022





**Figure 1.** Workflow for generating the ensemble and selectivity maps. Protein structures in the ensemble are aligned in the region of the binding site using the CSD Python API, and ligands, metals, and waters are removed. Hotspot maps for each structure are then calculated and combined into an ensemble map. The ensemble maps for on and off-targets can then be compared, highlighting areas predicting target-selective interactions. The color coding is red for the hydrogen bond acceptor maps, blue for hydrogen bond donor maps, and yellow for apolar maps. This color coding is consistent throughout the manuscript.

for binding, which allows them to serve as the basis of automated approaches for hit prioritization and progression.

Tuning compound selectivity, or introducing polypharmacology, is a challenge frequently encountered in the progression of hits and their subsequent development into drugs and chemical probes. A popular way to identify potential protein off-targets for a compound of interest is to use a method for comparing binding sites. Algorithms for binding site comparison have been documented in the literature, as reviewed by Ehrt et al.<sup>21</sup> and more recently by Konc.<sup>22</sup> Binding site features can be encoded as fingerprints, graph models, or grids.<sup>21</sup> Fingerprints are highly computationally efficient and so can be used for comparisons between large numbers of structures, such as when identifying distantly related off-target effects. Grid-based methods are computationally more intensive to generate but can encode nuanced information on the interactions of the binding sites, which is beneficial in the comparison of closely related proteins. Examples of grid-based methods for binding site mapping include GRID,<sup>23</sup> FLAP,<sup>24</sup> APF,<sup>25</sup> DoGSiteScorer,<sup>26</sup> SiteMap,<sup>27</sup> and PLImap.<sup>28</sup> As the importance of accounting for protein flexibility in binding site comparisons has become increasingly apparent, methods that calculate binding pocket signatures from an ensemble of conformations have been developed and employed successfully in medicinal chemistry programs.<sup>21</sup> Examples include the study by Österberg et al.<sup>29</sup> and more recently by Volkamer et al.<sup>30</sup> The former combined ensembles of AutoDock<sup>31</sup> interaction energy grids into a single grid by using Boltzmann-weighted averages for the values at each point in space. These grids showed improved performance in docking compared to those derived by taking the mean or minimum values of the interaction energy. Volkamer and co-workers used grids generated by DogSiteScorer,<sup>26</sup> a method for predicting pocket druggability, and then compared the frequencies at the points observed in the target and off-target ensembles. Building on this study, in 2018, Turk et al. used AutoGrid atom-based energy grids using polar and apolar atom probes as a key part of a computational pipeline to guide the automated selectivity conversion of an Aurora kinase inhibitor

for the TrkA kinase.<sup>32</sup> In terms of extracting information from an ensemble of grids, Schmalhorst and Bergner have developed a method based on SiteMap to identify structures with unique design opportunities within the ensemble,<sup>27</sup> providing a further example of the utility and opportunities that can be explored by combining information from grid-based binding site representations.

Fragment hotspot mapping differs from other methods as it provides a quantitative estimate of the propensity of a fragment to exploit particular interactions within the binding site. In addition, the visualization of the maps grants the user an intuitive and visual understanding of the binding pocket. The combination of these two features makes this approach particularly attractive for progressing hits in a rational and data-driven way while also allowing for expert insight and intuition. This adds value to drug discovery campaigns by automating subjective decision making, leading the process to become more objective, reliable, and scalable.

As more structural data become available during the early stages of drug discovery campaigns, researchers face the challenge of processing information from up to hundreds of individual protein–ligand structures and distilling it into testable hypotheses. To address the challenge of usefully combining information for structures of the same protein target, we have developed an “ensemble” hotspot map approach. By comparing two ensemble maps, a hotspot selectivity map can be derived. This highlights the structural differences that determine the selectivity of a compound for one protein over another member of the same protein domain family. The ensemble and selectivity maps were parameterized using retrospective examples of compounds showing selectivity between proteins in the same family. The resulting maps can be visualized within PyMOL<sup>33</sup> through a script automatically generated by the Hotspots API<sup>34</sup> or in any molecular viewer that supports visualizing .ccp4 or .grd formats. The method has been adopted in-house at Exscientia within a number of drug discovery programs. The code supplied in the Github repository (<https://github.com/prcurran/hotspots>, <https://github.com/ccdc-opensource/hotspots/tree/master>, [285](https://</a></p></div><div data-bbox=)

github.com/CMD-Oxford/hotspotEnsembles) includes PyMOL scripts for visualization.

## METHODS

Figure 1 shows the full workflow for generating ensemble and selectivity hotspot maps. After curating the ensemble data, the selected structures are aligned in the region of the binding site and prepared for the hotspot map calculation. Ensemble and selectivity maps are then calculated as described below.

**Data Curation and Structure Preparation.** Human bromodomain structures were collected using the SIENA tool on the ProteinsPlus webserver.<sup>35</sup> The tool was queried through its RESTful API; the full parameters are provided in the Supporting Information, Table S1 and in Github repositories (<https://github.com/prcurran/hotspots>, <https://github.com/CMD-Oxford/hotspotEnsembles>, <https://github.com/ccdc-open-source/hotspots/tree/master>), which include the code for querying the API. The structures were protonated using the Protoss web server.<sup>36</sup> Only structures with completely modeled residues in the peptide binding site were included, and no mutations in the binding site were allowed. Only structures deposited after the year 2000 were used to ensure consistent model refinement and processing.

Human kinase structures were downloaded from the KLIFS database, accessed in September 2020.<sup>37</sup> Only structures with the DFG-in conformation, resolution better than 2.5 Å, and a KLIFS quality score above 7 were kept. No restrictions were placed on the position of the glycine-rich loop or  $\alpha$ C helix. All ligands were bound within the ATP-binding pocket. Structures in KLIFS were protonated with Protoss and have had alternative residue conformations removed. Structures with mutations in the binding site were discarded. Structures with an identical ligand (based on the ligand canonical SMILES) were removed, and the highest resolution structure was used in the ensemble calculation.

Ligands, water, metals, and ions were removed from all structures prior to calculating the hotspot maps. While metals and strongly bound waters can contribute toward selectivity, in this case we were interested in replicating scenarios where very little previous information is known about the target, and where such features have been previously identified, they can be retained through CSD Python API functionality, as described elsewhere.<sup>20</sup>

Finally, to mimic crystallographic fragment-screening scenarios, only structures with ligands (excluding solvents) with a molecular weight below 300 Da were included in both the kinase and bromodomain data sets. This molecular weight cutoff was chosen since restricting the filter to a strict “rule of three” results in missing key protein–fragment complexes that are historic starting points for successful fragment-to-lead optimizations, notably p38 $\alpha$  structures 1WBW and 1W84.<sup>38</sup> The full list of structures used in the case studies is provided in the Supporting Information, Table S3. No unliganded structures were included in the analysis.

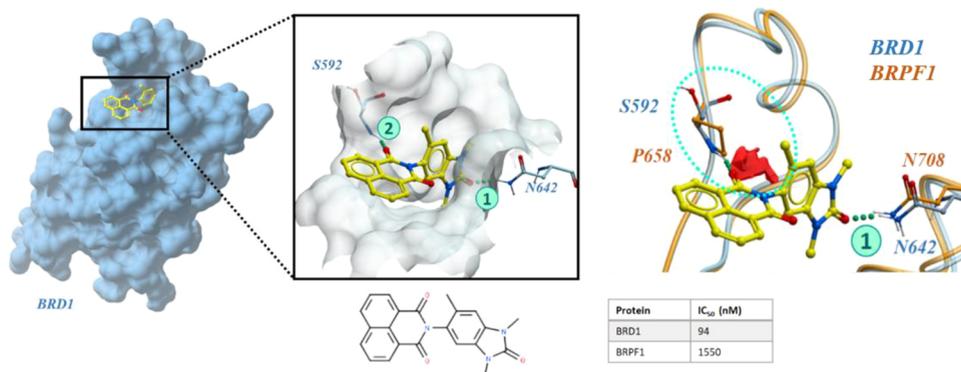
**ChEMBL Data Curation.** Compound activities were downloaded from the ChEMBL<sup>39,40</sup> database release 29 (July 2021), following the protocol described by Bosc et al. with modifications.<sup>41</sup> All activities recorded against human BRD1 (ChEMBL2176774), BRPF1 (ChEMBL3132741), BRD2 (ChEMBL1293289), BRD4 (ChEMBL1163125), BRD7 (ChEMBL3085622), and BRD9 (ChEMBL1163125) were retrieved. Only bioactivities with a standard relation of “=” and standard\_flag = True were considered. Mutant sequences,

potential duplicates, or data points with data validity comments were dropped. The assay type was restricted to “B” (binding assays) and data sources to src\_id = 1 (“scientific literature”). Only assays with standard units in nM were included, with standard\_type = “IC50” or “Kd.” Only entries with ChEMBL quality scores of 9 (human targets flagged as “SINGLE PROTEIN”) were included. Activities against the second bromodomains of BRD2 and BRD4 as well as against the BRPF1A isoform were removed using a keyword search in the “assay\_description” field. When multiple activity values were reported for a compound/target pair, the lowest (most potent) one was taken. Selectivity ratios were calculated by dividing the standard value for the off-target by that for the on-target. The selectivity ratios were calculated only for activities of the same standard type (“IC50” or “Kd.”) Only compounds with activity values reported against at least two of the six bromodomain targets were considered. In addition, selected compounds were required to have a crystal structure in complex with at least one of the targets for which activity had been reported. Crystal structures were retrieved from the Protein Data Bank (PDB)<sup>42</sup> by querying with the compound InChI.

**Structure Alignment.** For each structure, the binding site was defined by taking all residues within 5 Å of the binding site ligand using the CSD Python API's Protein.BindingSiteFromMolecule() function with the distance parameter set to 5.0 and “whole\_residues” set to true, meaning that any residue that places a heavy atom within 5 Å of the ligand is included in the binding site definition. The union of binding site residues from all protein structures within the ensemble then gave the ensemble binding site. The CSD Python API (version 3.0.4)<sup>20</sup> was used to align the ensemble structures based on the residues in the ensemble binding site using only their C $\alpha$  atoms.

**Fragment Hotspot Mapping.** Fragment hotspot maps were generated as previously described.<sup>14,34</sup> The default values for the fragment hotspot mapping method, release 1.0.5 were used. GHECOM<sup>43</sup> version 20200721 was used as a pocket detection and buriedness estimation method. A total of 3000 probe rotations was used for the fragment probe sampling step, as previously described.<sup>34</sup> We recommend this value as it offers sufficient thoroughness in sampling while still retaining a reasonable speed of calculation of the hotspot maps.<sup>34</sup> The default seven-atom fragment probes were used as described by Radoux et al.<sup>14</sup> Maps were then truncated to the region of the binding site to facilitate downstream analysis.

**Generating Ensemble Maps.** The fragment hotspot map algorithm outputs a set of three maps, one for each interaction probe type (donor, acceptor, and apolar). Highly scoring points in each map denote areas where a fragment is likely to form this type of interaction. The hotspot map for each probe for a single protein structure is a 3-dimensional grid with a spacing of 0.5 Å. An ensemble map is generated by calculating hotspot maps for multiple overlaid proteins in the same reference grid frame. This results in a set of hotspot maps where, for each probe type, grid points are assigned a set of values, one from each protein structure. For the apolar grid, values are aggregated by calculating the median, whereas for polar grids (donor and acceptor), values are aggregated by calculating the median of the nonzero points above a frequency cutoff. This difference is necessary as polar hotspots are intrinsically smaller than apolar hotspots (fewer than a hundred grid points in polar clusters compared to thousands



**Figure 2.** Selectivity in human bromodomains: BRD1 over BRPF1. The figure shows the selective precursor to the chemical probe BAY-299, which was cocrystallized with BRD1 (PDB ID: 5N49). It forms two hydrogen bonds with the protein in the binding pocket, shown as (1) and (2) in the pop-out. In BRPF1 (orange ribbon and sticks), this interaction (2) cannot occur because of the substitution of a proline (P658) at this location. The acceptor selectivity maps for BRD1 over BRPF1 identify this difference, shown by the area of acceptor propensity (red surface). The table shows the compound's IC<sub>50</sub> values for both proteins as reported by Bouché et al.<sup>49</sup>

of grid points in a cluster for the apolar maps) and more dependent on the orientation of the donor and acceptor groups. Including these zero values in the calculation of the median leads to dilution of information (see Figure 5, final column). Therefore, to avoid introducing consequential false negatives, only the nonzero values for grid points that have a hotspot value of at least 20% (of cases referred to as the “frequency” of a point in Figure 5) are considered when calculating the median. Apolar grid points are more likely to have associated values in most of the structures and individual zero values have less of an effect (see Figure 5). Thus, in the apolar case, the median map is calculated including zero values for all points.

**Generating Selectivity Maps.** For a given pair of proteins (on-target and off-target), the off-target ensemble maps are subtracted from the on-target ensemble maps to create difference maps. The difference maps are characteristically sparse grids with occasional clusters of nonzero values. To identify features within the difference map and enable downstream analysis, the density-based clustering algorithm HDBSCAN<sup>44</sup> is used. HDBSCAN does not require any initial estimate on the number of clusters to identify, merely requiring a single clustering parameter, which is the minimal number of points in a cluster. We use a value of 7 for polar maps as that is equivalent to the smallest spherical element in a voxelized grid with a radius comparable to that of the polar probe atoms (oxygen and nitrogen) and a value of 27 to approximate a methyl group as the minimal apolar feature.

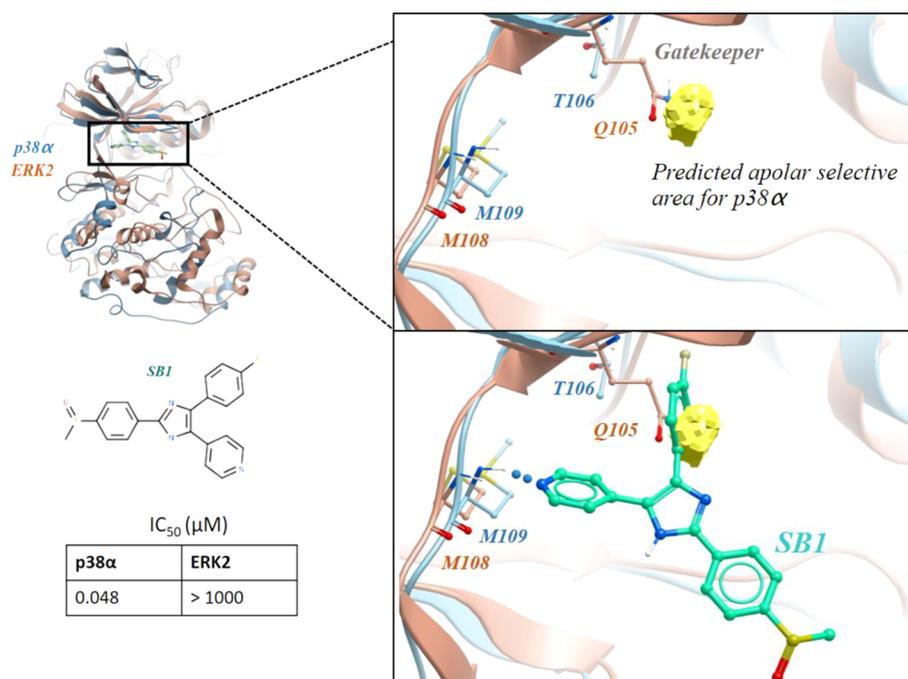
Clusters are annotated for the on-target (positive values) and off-target (negative values) regions of the difference map. Any overlapping clusters (centroids less than 1.5 Å for rigid binding sites and 3 Å for more flexible targets with flexible binding pockets) are disregarded. Each cluster is assigned a score corresponding to the median value of all the points in that cluster. Clusters with a median cluster score below 10 are not considered selective, reflecting the minimum hotspot value, which is considered to denote a favorable fragment interaction as previously reported.<sup>14</sup>

## RESULTS AND DISCUSSION

We envisage that the ensemble and selectivity maps can be used as a tool to guide the elaboration of fragment and lead-like compounds by both highlighting favorable unexploited interactions in the vicinity of the compound as well as

identifying unfavorable interactions that are currently made by a ligand. We demonstrate using three well-explored, therapeutically significant examples that the ensemble and selectivity map generation protocol can retrospectively rationalize observed compound selectivity between members of the same protein family. The example data sets include the bromodomain proteins BRD1 and BRPF1, the CAMK family kinases p38 $\alpha$  and ERK2, and the more distantly related kinases PIM1 and CK2 $\alpha$ . To mimic prospective crystallographic fragment-screening data, the example data sets consist solely of crystallographic structures in complex with fragment-sized bound ligands. While the methodology presented is applicable to a wider range of protein structural data (apo-structures, complexes with larger ligands, etc.) used in compound optimization, in this study, we have chosen to focus on fragments as there is a great need for methods in this space to drive the progression from hits to leads with increased potency and selectivity, and these methods need to be specifically developed and parameterized for use with fragment-sized ligands. Finally, we show how the ensemble and selectivity maps can enable automated analyses within the bromodomain protein family.

**Bromodomains: Selectivity between BRD1 and BRPF1.** Bromodomain proteins are a family of epigenetic regulators acting as readers of histone tail modifications. They bind acetylated lysine residues on histone protein tails usually through a highly conserved asparagine residue in the bromodomain binding site.<sup>45</sup> Over the past decade, there has been significant pharmaceutical interest in the development of selective bromodomain inhibitors as epigenetic dysregulation underpins a number of human diseases, including many cancers.<sup>45–47</sup> In the human bromodomain BRPF subfamily, the substitution of a serine (S592) in BRD1 by a proline (P658) in the closely related BRPF1 (Figure 2) results in a notable difference proximal to the conserved asparagine residue within the binding sites of these two proteins. The serine backbone nitrogen can form a hydrogen bond with ligand acceptor atoms within the bromodomain binding pocket, while the corresponding proline backbone nitrogen in BRPF1 cannot. Although this substitution can be identified at the sequence level, visual inspection is needed to show that the serine backbone NH in BRD1 is accessible from the binding site. However, this is not sufficient information to conclude whether the presence or absence of this interaction



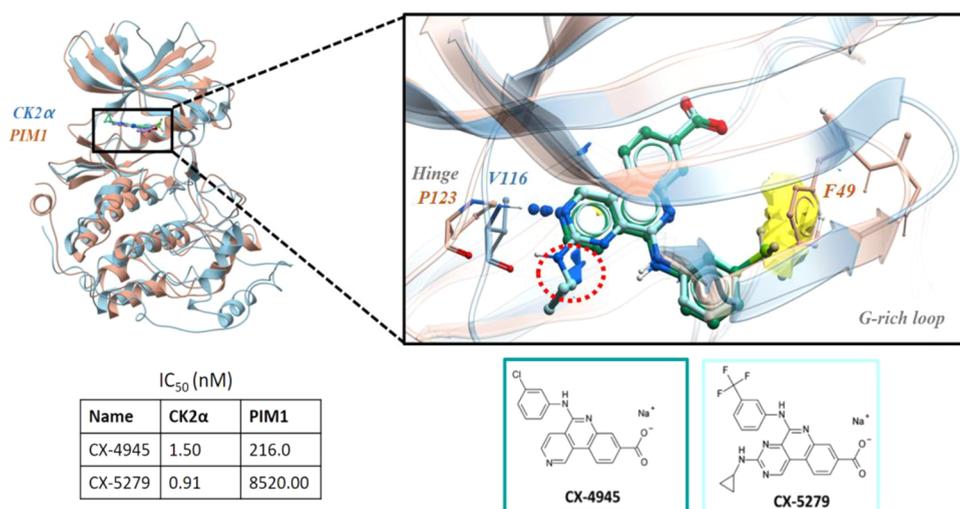
**Figure 3.** Kinase selectivity: identifying gatekeeper differences. The figure shows the overlaid structures of p38 $\alpha$  (PDB ID 1A9U, in steel blue) and ERK2 (PDB ID 4QP1, in peach). The close-up shows the binding mode of SB1 (light green). The ligand's fluorophenyl moiety is situated in a hydrophobic back pocket located behind the gatekeeper residue T106. The apolar selectivity map (yellow surface) for p38 $\alpha$  over ERK2 highlights this area as a favorable location to place a selective apolar group, such as fluorophenyl in SB1. The table shows the compound's IC<sub>50</sub> values for both proteins, as reported by Wang et al.<sup>50</sup>

would make a measurable difference in binding affinity and, therefore, whether it can be used to drive selectivity between the two targets. To explore whether this information can be automatically derived from our approach, ensemble maps were calculated using 23 BRD1 and 26 fragment-bound BRPF1 structures (the full list of PDB codes is provided in the Supporting Information, Table S3). Figure 2 shows that the maps were able to identify the selective binding feature, automatically capturing this information without the need for visual inspection of large numbers of protein structures. A review of the literature studies shows that this feature is exploited by the BRD1-selective probe BAY-299, published in 2017 by Bouché et al.<sup>48</sup> The original hit was found in a high-throughput screen and already exhibited nanomolar affinity for BRD1 while not exhibiting activity against BRPF1, BRPF3, and BRD4 up to 20  $\mu$ M. The addition of a methyl group at position 6 of the 1,3-dimethylbenzimidazolone core introduced a fourfold increase in affinity for BRD1, locking the tricyclic group in a bioactive conformation. The structure of this compound in complex with BRD1 was solved (PDB ID: 5N49), which showed that one of the carbonyl groups on naphthalimide makes the selective hydrogen bond to S592. The IC<sub>50</sub> values show that the compound has a little over 15-fold selectivity for BRD1 over BRPF1 (in a TR-FRET assay). As this compound had poor solubility, an alkyl alcohol tail was added at position 4 on naphthalimide, yielding the chemical probe BAY-299. The authors report the hydrogen bond with S592 as a key factor driving the selectivity of these compounds for BRD1 versus BRPF1. Generating selectivity maps from an ensemble was critical in this example as the selective feature is not present in all the individual maps from BRD1 structures (as exemplified by PDB ID SPOS, Supporting Information, Figure S1). This is due to the sensitivity of hydrogen bonds to

the orientation of the donor and acceptor groups, where small twisting motions in the backbone can mean that the feature is not detected in a minority of conformations.

**Designing Selectivity between Closely Related Kinases: p38 $\alpha$  and ERK2.** In the family of human protein kinases, an ATP-binding pocket residue known as the gatekeeper is an important determinant of selectivity.<sup>30</sup> This is exploited by the p38 $\alpha$ -selective inhibitor SB1 (SB203580) as shown in Figure 3, which possesses selectivity over related MAPK kinases, notably ERK2.<sup>49,50</sup> The apolar selectivity maps for an ensemble of five fragment-bound p38 $\alpha$  structures against 17 ERK2 fragment-bound structures (Supporting Information Table S3) can identify and highlight the selective hydrophobic pocket that the inhibitor binds in, as illustrated in Figure 3. As with the previous example, all structures in this case study had fragment-sized bound ligands in order to recreate a prospective fragment-screening scenario.

The fluorophenyl group of SB1 occupies the selective hydrophobic back pocket of p38 $\alpha$  and clashes with the glutamine gatekeeper of ERK2. The hotspot maps were able to identify this feature using only fragment-bound structures as an input for the ensemble maps. In fact, this back pocket is often explored by fragment-sized hits (e.g., PDB ID 1W7H). In the case of fragment screening against p38 $\alpha$ , the selectivity maps could be used to indicate which fragments (and chemical groups within the fragments) might be selective for p38 $\alpha$  over ERK2, providing suggestions for achieving selectivity at a very early stage in the compound design process. Despite using only five structures, all the bound ligands in the p38 $\alpha$  ensemble have unique Murcko scaffolds,<sup>51</sup> and they all explore the selective pocket. This includes the minimal pharmacophore chlorophenol (PDB ID 1WBO) that preferentially explores the selective area behind the gatekeeper compared to the canonical



**Figure 4.** Tuning the selectivity of CK2 $\alpha$  inhibitors toward PIM1. The binding mode of inhibitors in the ATP-binding pocket of CK2 $\alpha$  is shown on the left. The close-up shows the overlay of the CK2 $\alpha$ -bound structures of CX-5279 (PDB ID 3ROT, ligand only, shown in aqua) and CX-4945 (PDB ID 3PE1, protein shown in light blue and ligand in dark teal), showing the selective hotspot regions for CK2 $\alpha$ . The apolar and donor maps are shown as yellow and blue surfaces, respectively. PIM1 (PDB ID 2C3I) is shown in peach for comparison. Residues giving rise to the selective features are shown as sticks and labeled. The red circle indicates the donor nitrogen in the cyclopropylamine group of the selective compound and the underlying donor density (blue surface) in the selectivity map. The IC<sub>50</sub> values shown are as reported by Battistutta et al.<sup>54</sup>

kinase hinge hotspot. The ERK2 ensemble contains more structures and higher ligand diversity (19 structures and 11 unique scaffolds), and while structures such as PDB ID 3ERK (not included in the ensemble as it is above the 300 Da limit) indicate that it is possible to reach that pocket in ERK2, the majority of the ERK2 fragments do not. Despite a relatively low number of structures in the p38 $\alpha$  ensemble, and a significant imbalance in the number of structures between the p38 $\alpha$  and ERK2 ensembles, our approach is still able to help identify selective features.

**CK2 $\alpha$  and PIM1: Distantly Related Kinases that Bind the Same Ligand.** The above examples demonstrate that the selectivity maps are able to retrospectively rationalize polar and apolar selectivity features within protein subfamilies. In this final case study, we explore a retrospective example of selectivity between kinases from different subfamilies: human CK2 $\alpha$  (CK2 subfamily) and human PIM1 (CAMK subfamily).

CX-4945 was originally developed as an ATP-competitive, orally available CK2 $\alpha$  inhibitor with nanomolar affinity for its target but which also displayed off-target activity for the PIM1 kinase.<sup>52,53</sup> In 2011, Battistutta et al. developed a series of CK2 $\alpha$  inhibitors among which CX-5279 (Figure 4) retained affinity for CK2 $\alpha$  while achieving selectivity against PIM1.<sup>53</sup>

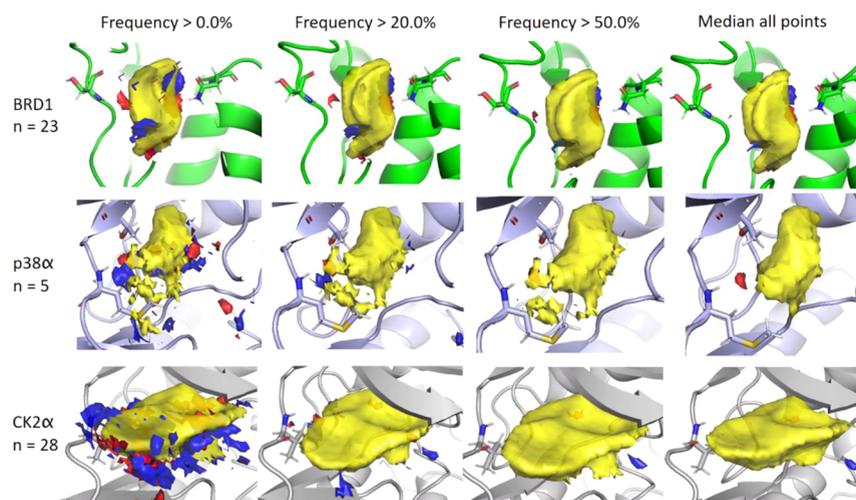
Ensemble and selectivity maps were calculated for 28 CK2 $\alpha$  and 32 PIM1 fragment-bound structures (Supporting Information, Table S3). The apolar selectivity map for CK2 $\alpha$  over PIM1 reveals an area of apolar propensity in CK2 $\alpha$  that is inaccessible in PIM1 due to the conformation of residue F49 in the PIM1 structures (Figure 4). This difference stems from the dominant conformation of this residue in the majority of ligand-bound PIM1 structures and so it would be nontrivial to predict from the sequence alone and without visual inspection of a large number of off-target structures. The ensemble and selectivity maps automatically highlight this feature without the need for human inspection and subjective judgment. The starting inhibitor, CX-4945, has a chlorophenyl group in this location. The selectivity maps suggest that placing a larger, highly lipophilic substituent on the phenyl ring in that location

may improve the selectivity for CK2 $\alpha$  over PIM1. In the closely related inhibitor CX-5279, trifluoromethyl was substituted for chlorine, which resulted in improved selectivity for CK2 $\alpha$ , despite the relatively small increase in molecular volume between the substituents (Figure 3). The authors reported that increasing the size of the apolar substituent at that position leads to a decrease in potency for PIM1, providing a link between that substituent and the selectivity of the compounds against PIM1. The authors suggested that this is due to a clash with the inward-facing F49 in the PIM1 kinase structures.

Another difference between CK2 $\alpha$  and PIM1 lies in the hinge region and is identifiable by the selectivity maps. PIM1 contains a proline (P123) at the position of the hinge valine (V116) in CK2 $\alpha$ , meaning it is unable to form one of the hydrogen bonds that CX-4945 makes with CK2 $\alpha$  (Figure 4). The PIM1 hinge also has an insertion of two residues. This, coupled with the valine to proline substitution, prevents the donor nitrogen in the cyclopropylamine group from forming a hydrogen bond with the backbone acceptor. The donor selectivity map highlights this feature (Figure 4). In the study reported in ref 54 in 2011, the SAR for the compound series that includes CX-4945 showed no increase in affinity for CK2 $\alpha$  when cyclopropylamine was added.<sup>53</sup> Furthermore, the influence of this moiety on selectivity against PIM1 is not explicitly shown. However, the resulting ligand is less lipophilic and makes an additional hydrogen bond, which is desirable in the design process; the selectivity maps suggest this modification. In the case of the apolar selective feature, a visual inspection of the ensemble structures reveals that F49 consistently adopts the inward-facing conformation that clashes with the selective ligand, increasing confidence in the selectivity of the feature.

#### Exploring and Adjusting the Method Parameters.

The quality of the ensemble maps is highly dependent on the quality and quantity of the input ensemble data. A key property of the ensemble is the number of structures included. While crystallographic fragment-screening experiments can provide



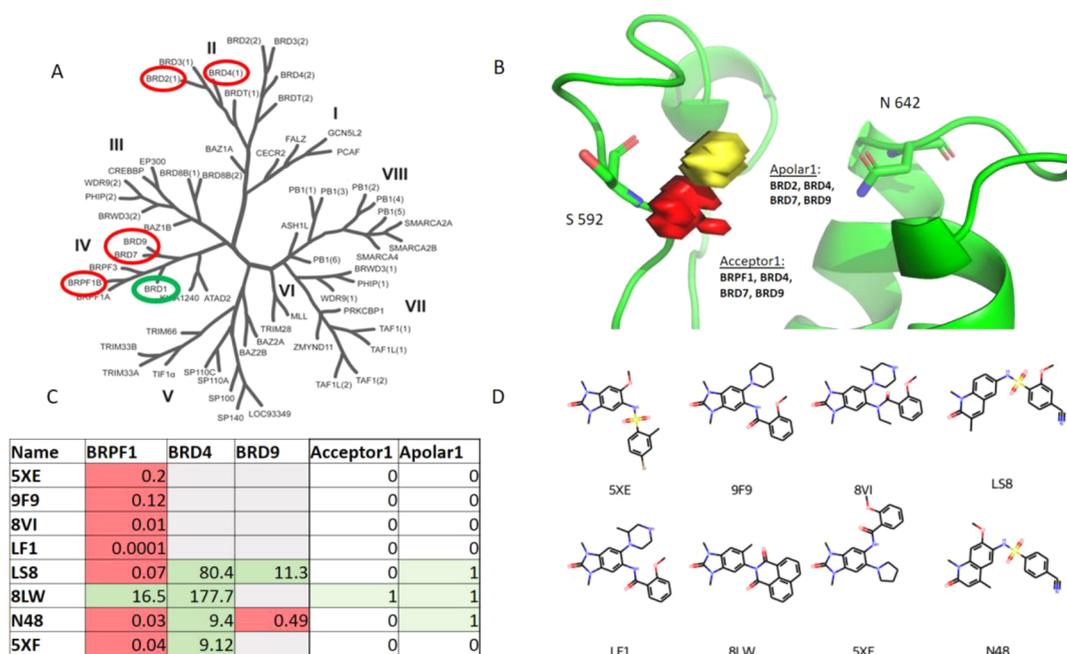
**Figure 5.** Setting the frequency threshold parameter for the ensemble maps. Ensemble maps are compiled by taking the median values of samples at a grid point for all points that have nonzero values in at least the threshold fraction of structures in the ensemble. The first three columns demonstrate the effect of setting this parameter to 0 (all points sampled at least once), 20, and 50% for the respective protein ensembles. The last column shows the median maps for the ensemble when zero values are also included.

up to tens of structures of ligands in complex with the target protein, such data are not always available. The ensemble of p38 $\alpha$  structures above shows that even a small ensemble with a diverse selection of ligands can be used to identify potentially selective features. With smaller ensemble sizes, individual structures contribute proportionally more to the ensemble map. Visual inspection, or plots such as those shown in the Supporting Information Figure S1, can be used to establish which structures contribute to a particular hotspot cluster. In the case of p38 $\alpha$ , all the maps contribute to the selective feature discussed. By reducing the frequency threshold of the ensemble maps, the contribution of individual structures to the maps can be amplified in larger ensembles as well. A complementary computational method can then be used to assess the significance of rare or unique conformations. For instance, molecular dynamics-based methods such as dynamic undocking<sup>54</sup> can be used to assess the stability of the ligand–protein complex and guide the decision whether that would be a desirable structure to include in the ensemble. For an ensemble with tens of structures, even at a 20–30% frequency threshold, signals arising from rare conformations may be lost in noise from adjacent clusters. In such cases, the question of the minimum representative ensemble arises. In the examples above, we employed a ligand-based measure for ensemble diversity—the number of unique Murcko scaffolds (Supporting Information Table S4)—and considered its value to guide the minimum number of structures in the ensemble. Protein or protein-and-ligand-based methods, such as protein–ligand interaction fingerprints<sup>55,56</sup> or backbone clustering methods,<sup>57</sup> may be used to select a diverse ensemble or even to generate ensemble maps for a group of structures. For instance, in a lead optimization campaign, a subset of structures in complex with ligands from the same series might be used to identify unexploited hotspot interactions in this very specific context. A general recommendation for ensembles with over 10 structures is to build ensemble maps with a frequency threshold of 20% to identify the most frequent hotspots throughout the ensemble. If a comparison of these maps against an off-target generates no selective clusters, switching to a lower threshold can reveal rarer features. Conversely, if the maps appear noisy

(see Figure 5, first column), switching to a higher frequency threshold can reveal the most prominent hotspots.

Another key point to consider when assembling a protein ensemble is the presence of mutually exclusive protein conformational states. In the kinase examples discussed above, only structures in the DFG-in conformation were included. In the case where a protein can adopt mutually exclusive states, we have found that compiling separate ensembles for each state can provide more detailed information for designing compounds against that particular conformation. Class imbalances within the ensembles can result in artifacts when calculating selectivity maps downstream. An example of this is shown in Supporting Information Figure S2, where a feature originating from the DFG-out conformation of p38 $\alpha$  passes the frequency cutoff in that ensemble (35 structures were in the DFG-in conformation, 17 in DFG-out, and 3 were classified as “out-like” in the KLIFS database. The full list of PDB codes is provided in the Supporting Information, Table S5). The corresponding ERK2 ensemble is exclusively DFG-in (69 structures). When selectivity maps of p38 $\alpha$  over ERK2 were calculated, a false positive region of density corresponding to the DFG-out conformation confounds the interpretation of the maps.

In the case of selectivity maps, two key parameters to be considered are the minimal distance between the centers of selective clusters in the target and off-target selectivity maps as well as the minimum hotspot score a cluster must pass in order to be considered selective. In principle, the ideal selective feature would both score highly and be located at a reasonable distance from other features that possess binding propensity for the off-target. The lower limit for the first parameter is twice the step size of the grid: 1 Å between features. The minimum distance is dependent on other factors such as the volume and flexibility of the binding site. For the small and rigid bromodomain binding sites, we found that using a cutoff of 1.5–2.0 Å was sufficient to isolate the polar selective features, which is consistent with values published previously.<sup>58</sup> For kinases, a threshold of 3 Å was chosen as their binding sites exhibit conformational diversity, so a value twice the minimal distance was used as a precaution against false



**Figure 6.** Selectivity maps identify selectivity-determining regions across subsets of targets in the same protein family. (A) Human bromodomain phylogenetic tree (adapted from <https://www.thesgc.org/chemical-probes>) showing the target (BRD1) circled in green and off-targets circled in red. (B) Summary maps of the BRD1 selectivity maps against the off-targets. The acceptor feature (Acceptor1) has contributions from the BRPF1, BRD4, BRD7, and BRD9 selectivity maps and the apolar feature (Apolar1) has contributions from BRD2, BRD4, BRD7, and BRD9. The apo-protein structure shown is of BRD1, PDB ID 3CRW. (C) Selectivity ratios for BRD1 versus off-targets for the eight compounds in the data set. Values above 1 indicate selectivity for BRD1 and are colored in green. Values below 1 are considered not selective and are colored in red, and combinations for which information is not available in the data set are colored in gray. The last two columns indicate whether the crystal structure of the compound places a heavy atom in the predicted selective region. (D) 2D structures of the compounds in the data set and their PDB molecule IDs. The corresponding ChEMBL IDs and the PDB codes of the protein-bound structures are provided in the Supporting Information Table S7.

positives. In cases where little information is available for the target and off-target proteins, prioritizing highly scoring clusters at a distance of 3 Å or more from the center of the closest off-target cluster would be a recommended starting point, lowering this parameter only if no features are detected, with the caveat that features will be identified with lower certainty.

### Selectivity Maps Identify Selectivity-Determining Regions across Subsets of Targets in the Same Protein Family.

After ascertaining that the selectivity maps are able to identify known selectivity features between pairs of proteins within the same family, we developed a procedure that would allow for automated and objective analyses across a target protein family. We again chose to focus on human bromodomains as there is a wealth of structural and activity data publicly available for validating the method predictions. In the example presented in Figure 6, BRD1 was chosen as the target protein and compared to both high-sequence identity (BRPF1, BRD7, and BRD9) and lower-sequence identity (BET bromodomains BRD2(1) and BRD4(1)) off-targets. Ensemble maps were calculated for all the proteins without applying the 300 Da cutoff on bound ligands, in order to take advantage of all the information available for the potential off-targets, as would be the case in a drug discovery project. Selectivity maps were then calculated for BRD1 against each of the off-target ensembles. The selectivity maps were combined into “summary” selectivity maps, using the methodology developed for compiling the ensemble maps. The frequency cutoff was set to zero, so a feature had to be present in at least one of the individual selectivity maps in order to be included. The summary selectivity maps contain information on the

specific off-targets against which this feature can be exploited. As in the previous case studies, a hotspot score cutoff of 10 was used for the selectivity maps. Two prominent features in the summary selectivity maps are shown below in Figure 6, panel B. The Acceptor1 feature adjacent to S592 includes contributions from the selectivity maps against BRPF1, BRD4, BRD7, and BRD9, while the apolar feature (Apolar1) is scored favorably against BRD7, BRD9, BRD2, and BRD4. We then identified compounds in the literature with published activities against BRD1 and at least one of the off-targets along with a published crystal structure in complex with at least one of the target or off-targets.

The procedure for querying activities and crystals identified eight compounds for which high-quality data were available for binding to BRD1 and for at least one of the off-targets, and which have also been crystallized in complex with one of the proteins of interest. We chose to limit our search to compounds with available crystal structures in order to avoid introducing noise that would necessarily be created with a docking procedure.

The structures of the eight compounds were rescored against the summary selectivity maps using the Hotspots API (as previously described<sup>14,34</sup>) in order to identify substituent groups that interact with the predicted selective areas. A hit was defined if at least one heavy atom in the compound structure scored favorably for a particular feature (last two columns in Figure 6, panel C). This analysis shows that compound 8LW, which is selective for BRD1 against both BRPF1 and the more distantly related targets, places appropriate atom types in both clusters. Compounds that hit only the apolar cluster do not exhibit selectivity over BRPF1, as

expected. However, compound N48 demonstrates that the selective apolar cluster on its own may not be sufficient to grant selectivity over BRD9. This is not unexpected as selectivity is a complex phenomenon achieved by the interplay of multiple structural features; hence, covering multiple features that are selective against different subsets of the protein family would result in a more selective ligand. This principle has been used extensively for well-researched protein families with known selectivity determinants such as kinases.<sup>59</sup> The selectivity maps have shown to be able to detect such features, and as scoring compound poses against the maps is computationally very fast (a few seconds per pose), these types of analyses can be used to objectively score large numbers of docked potential follow-up poses. The previously published Hotspots API<sup>34</sup> also allows for the extraction of pharmacophoric features from the maps, which can then be provided as an input to programs such as CrossMiner<sup>60</sup> and used for the growing and merging of compounds.

## CONCLUSIONS

Fragment hotspot mapping has previously been shown to be a promising method for guiding the hit-to-lead phase of drug and probe discovery campaigns. Building upon this approach, we have introduced ensemble hotspot maps to summarize important fragment-binding interactions made by an ensemble of structures, extending the original hotspot method to work over multiple conformations of the same protein. Importantly, ensemble hotspot maps can then be used to highlight differences between related proteins in the same family. These maps can highlight nuanced differences between protein binding sites. We have shown three case studies from well-researched protein families in which the selectivity maps were able to identify binding site differences used to design selective inhibitors. The method will be applied to further protein families to obtain the recommended values for the various map parameters as well as to cases in which X-ray structures of defragmented bound ligands have been crystallized. Compiling ensemble maps from snapshots of molecular dynamics trajectories is also being investigated to further understand the behavior of binding site hotspots, especially in cases where few experimental structures are available. Currently, the fragment hotspot mapping method has been validated for apolar, hydrogen bond donor, and hydrogen bond acceptor probes. These interactions represent a large portion of protein–fragment intermolecular contacts, but the method could in the future be extended by adding charged or halogen fragment probes. The workflows for calculating ensemble and selectivity maps can also be applied to process grid-based representations from other binding site mapping methods although the values for the map-specific parameters will likely have to be adjusted. Overall, the ensemble and selectivity maps are a quick and scalable means to summarize and intuitively present structural information from closely related proteins and generate hypotheses on achieving selectivity.

**Data and Software Availability.** The code for the ensemble and hotspot maps along with the data from the examples presented above and tutorials on installing the Hotspots API and its dependencies can be found in the following Github repositories: <https://github.com/ccdc-opensource/hotspots>, <https://github.com/prcurran/hotspots> (latest version of the Hotspots API) and <https://github.com/CMD-Oxford/hotspotEnsembles> (this contains the data for the figures presented in this study and the scripts used to

generate it). The code is available for free but is based on the CSD Python API<sup>20,34</sup> and the CSD program SuperStar, which require a CSD license. The full data for the case studies, including the prepared input protein structures, are provided in the following link: <https://doi.org/10.5281/zenodo.5574443>.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.1c00823>.

Ensemble maps identify binding site interactions that could be missed by focusing on a single structure (Figure S1), rationale on not including mutually exclusive conformations in the ensemble maps (Figure S2), SIENA query parameters (Table S1), KLIFS query parameters (Table S2), structures used in the case study examples (Table S3), number of unique Murcko scaffolds (Table S4), KLIFS DFG motif annotation of the p38 $\alpha$  and ERK2 structures in the Supporting Information Figure S2 ensembles (Table S5), ensemble PDB codes for the maps shown in Figure 6 (Table S6), and structural data for the compounds shown in Figure 6 (Table S7) (PDF)

Names and SMILES strings of the compounds presented in the figures (XLSX)

## AUTHOR INFORMATION

### Corresponding Authors

**Mihaela D. Smilova** – Centre for Medicines Discovery, University of Oxford, Oxford OX3 7DQ, U.K.; [orcid.org/0000-0001-6255-7312](https://orcid.org/0000-0001-6255-7312); Email: [mihaela.smilova@cmd.ox.ac.uk](mailto:mihaela.smilova@cmd.ox.ac.uk)

**Brian D. Marsden** – Centre for Medicines Discovery, University of Oxford, Oxford OX3 7DQ, U.K.; Kennedy Institute of Rheumatology, NDORMS, University of Oxford, Oxford OX3 7DQ, U.K.; [orcid.org/0000-0002-1937-4091](https://orcid.org/0000-0002-1937-4091); Email: [brian.marsden@cmd.ox.ac.uk](mailto:brian.marsden@cmd.ox.ac.uk)

### Authors

**Peter R. Curran** – The Cambridge Crystallographic Data Centre (CCDC), Cambridge CB2 1EZ, U.K.; Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, U.K.

**Chris J. Radoux** – Exscientia Ltd., Oxford OX4 4GE, U.K.

**Frank von Delft** – Centre for Medicines Discovery, University of Oxford, Oxford OX3 7DQ, U.K.; Diamond Light Source Ltd., Didcot OX11 0DE, U.K.; Research Complex at Harwell, Didcot OX11 0FA, U.K.; Department of Biochemistry, University of Johannesburg, Auckland Park 2006, South Africa

**Jason C. Cole** – The Cambridge Crystallographic Data Centre (CCDC), Cambridge CB2 1EZ, U.K.; [orcid.org/0000-0002-0291-6317](https://orcid.org/0000-0002-0291-6317)

**Anthony R. Bradley** – Exscientia Ltd., Oxford OX4 4GE, U.K.

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jcim.1c00823>

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

## Funding

M.D.S. was supported by funding from the Engineering and Physical Sciences Research Council (EPSRC), the Medical Research Council (MRC) [grant number EP/L016044/1], Exscientia, and the CCDC. P.R.C. was funded by the Biotechnology and Biological Sciences Research Council and UCB (BB/P50466X/1). B.D.M. and F.v.D. were supported by funding from the Innovative Medicines Initiative (EU/EFPIA, ULTRA-DD grant no. 115766) and the Wellcome Trust (106169/ZZ14/Z). B.D.M. was further supported by The Kennedy Trust for Rheumatology Research. This work has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 875510. The JU receives support from the European Union's Horizon 2020 research and innovation program, EFPIA, and the Ontario Institute for Cancer Research, Royal Institution for the Advancement of Learning McGill University, Kungliga Tekniska Högskolan, Diamond Light Source Limited.

## Notes

The authors declare the following competing financial interest(s): J.C.C. works for the CCDC, the creators of the underlying software (the CSD System) that is used by this library to generate the maps.

## REFERENCES

- (1) Bradley, A. R.; et al. The SGC beyond structural genomics: redefining the role of 3D structures by coupling genomic stratification with fragment-based discovery. *Essays Biochem.* **2017**, *61*, 495–503.
- (2) van Montfort, R. L. M.; Workman, P. Structure-based drug design: aiming for a perfect fit. *Essays Biochem.* **2017**, *61*, 431–437.
- (3) Jacquemard, C.; Kellenberger, E. A bright future for fragment-based drug discovery: what does it hold? *Expert Opin. Drug Discovery* **2019**, *14*, 413–416.
- (4) Doak, B. C.; Norton, R. S.; Scanlon, M. J. The ways and means of fragment-based drug design. *Pharmacol. Ther.* **2016**, *167*, 28–37.
- (5) Patel, D.; Bauman, J. D.; Arnold, E. Advantages of crystallographic fragment screening: Functional and mechanistic insights from a powerful platform for efficient drug discovery. *Prog. Biophys. Mol. Biol.* **2014**, *116*, 92–100.
- (6) Markham, A. Erdafitinib: First Global Approval. *Drugs* **2019**, *79*, 1017–1021.
- (7) Souers, A. J.; et al. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* **2013**, *19*, 202–208.
- (8) Bollag, G.; et al. Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat. Rev. Drug Discovery* **2012**, *11*, 873–886.
- (9) Monestime, S.; Lazaridis, D. Pexidartinib (TURALIO™): The First FDA-Indicated Systemic Treatment for Tenosynovial Giant Cell Tumor. *Drugs R. D.* **2020**, *20*, 189–195.
- (10) Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discovery* **2016**, *15*, 605–619.
- (11) Zhang, D.; et al. Transition State Analogues Enhanced by Fragment-Based Structural Analysis: Bacterial Methylthioadenosine Nucleosidases. *Biochemistry* **2020**, *59*, 831–835.
- (12) Taylor, S. J.; et al. Discovery of Potent, Selective Chymase Inhibitors via Fragment Linking Strategies. *J. Med. Chem.* **2013**, *56*, 4465–4481.
- (13) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **2005**, *48*, 2518–2525.
- (14) Radoux, C. J.; Olsson, T. S. G.; Pitt, W. R.; Groom, C. R.; Blundell, T. L. Identifying Interactions that Determine Fragment Binding at Protein Hotspots. *J. Med. Chem.* **2016**, *59*, 4314–4325.
- (15) Brenke, R.; et al. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621–627.
- (16) Graham, S. E.; Leja, N.; Carlson, H. A. MixMD Probeview: Robust Binding Site Prediction from Cosolvent Simulations. *J. Chem. Inf. Model.* **2018**, *58*, 1426–1433.
- (17) Alvarez-Garcia, D.; Barril, X. Molecular simulations with solvent competition quantify water displaceability and provide accurate interaction maps of protein binding sites. *J. Med. Chem.* **2014**, *57*, 8530–8539.
- (18) Cafilisch, A.; Miranker, A.; Karplus, M. Multiple Copy Simultaneous Search and Construction of Ligands in Binding Sites: Application to Inhibitors of HIV-1 Aspartic Proteinase. *J. Med. Chem.* **1993**, *36*, 2142–2167.
- (19) Straniero, V.; et al. Benzodioxane-Benzamides as Antibacterial Agents: Computational and SAR Studies to Evaluate the Influence of the 7-Substitution in FtsZ Interaction. *ChemMedChem* **2020**, *15*, 195–209.
- (20) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge structural database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 171–179.
- (21) Ehrst, C.; Brinkjost, T.; Koch, O. Impact of Binding Site Comparisons on Medicinal Chemistry and Rational Molecular Design. *J. Med. Chem.* **2016**, *59*, 4121–4151.
- (22) Konc, J. Binding site comparisons for target-centered drug discovery. *Expert Opin. Drug Discovery* **2019**, *14*, 445–454.
- (23) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- (24) Cross, S.; Baroni, M.; Carosati, E.; Benedetti, P.; Clementi, S. GRID Molecular Interaction Fields in Virtual Screening. Validation using the DUD Data Set. *J. Chem. Inf. Model.* **2010**, *50*, 1442–1450.
- (25) Totrov, M. Ligand binding site superposition and comparison based on Atomic Property Fields: identification of distant homologues, convergent evolution and PDB-wide clustering of binding sites. *BMC Bioinf.* **2011**, *12*, S35–S35.
- (26) Volkamer, A.; Kuhn, D.; Rippmann, F.; Rarey, M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **2012**, *28*, 2074–2075.
- (27) Schmalhorst, P. S.; Bergner, A. A Grid Map Based Approach to Identify Nonobvious Ligand Design Opportunities in 3D Protein Structure Ensembles. *J. Chem. Inf. Model.* **2020**, *60*, 2178–2188.
- (28) Rathi, P. C.; et al. Predicting 'Hot' and 'Warm' Spots for Fragment Binding. *J. Med. Chem.* **2017**, *60*, 4036–4046.
- (29) Österberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins: Struct., Funct., Bioinf.* **2002**, *46*, 34–40.
- (30) Volkamer, A.; Eid, S.; Turk, S.; Rippmann, F.; Fulle, S. Identification and Visualization of Kinase-Specific Subpockets. *J. Chem. Inf. Model.* **2016**, *56*, 335–346.
- (31) Morris, G. M.; et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- (32) Turk, S.; Merget, B.; Eid, S.; Fulle, S. From Cancer to Pain Target by Automated Selectivity Inversion of a Clinical Candidate. *J. Med. Chem.* **2018**, *61*, 4851–4859.
- (33) DeLano, W. L. Pymol: An open-source molecular graphics tool, CCP4 Newsletter on Protein Crystallography; 2002; Vol. 40.
- (34) Curran, P. R.; et al. Hotspots API: A Python Package for the Detection of Small Molecule Binding Hotspots and Application to Structure-Based Drug Design. *J. Chem. Inf. Model.* **2020**, *1911*.
- (35) Bietz, S.; Rarey, M. SIENA: Efficient Compilation of Selective Protein Binding Site Ensembles. *J. Chem. Inf. Model.* **2016**, *56*, 248–259.
- (36) Bietz, S.; Urbaczek, S.; Schulz, B.; Rarey, M. Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. *Aust. J. Chem.* **2014**, *6*, 12.

- (37) Kooistra, A. J.; et al. KLIFS: A structural kinase-ligand interaction database. *Nucleic Acids Res.* **2016**, *44*, D365–D371.
- (38) Gill, A. L.; et al. Identification of Novel p38 $\alpha$  MAP Kinase Inhibitors Using Fragment-Based Lead Generation. *J. Med. Chem.* **2005**, *48*, 414–426.
- (39) Davies, M.; et al. ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res.* **2015**, *43*, W612–W620.
- (40) Gaulton, A.; et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954.
- (41) Bosc, N.; et al. Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery. *Aust. J. Chem.* **2019**, *11*, 4.
- (42) Berman, H. M.; et al. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (43) Kawabata, T. Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1195–1211.
- (44) McInnes, L.; Healy, J.; Astels, S. hdbSCAN: Hierarchical density based clustering. *J. Open Source Softw.* **2017**, *2*, 205.
- (45) Muller, S.; Filippakopoulos, P.; Knapp, S. Bromodomains as therapeutic targets. *Expert Rev. Mol. Med.* **2011**, *13*, No. e29.
- (46) Preston, A.; et al. Design and Synthesis of a Highly Selective and in Vivo-Capable Inhibitor of the Second Bromodomain of the Bromodomain and Extra Terminal Domain Family of Proteins. *J. Med. Chem.* **2020**, *63*, 9070–9092.
- (47) Bhattacharya, S.; Piya, S.; Borthakur, G. Bromodomain inhibitors: What does the future hold? *Clin. Adv. Hematol. Oncol.* **2018**, *16*, 504–515.
- (48) Bouché, L.; et al. Benzoisoquinolinediones as Potent and Selective Inhibitors of BRPF2 and TAF1/TAF1L Bromodomains. *J. Med. Chem.* **2017**, *60*, 4002–4022.
- (49) Wang, Z.; et al. Structural basis of inhibitor selectivity in MAP kinases. *Structure* **1998**, *6*, 1117–1128.
- (50) Gallagher, T. F.; et al. Regulation of stress-induced cytokine production by pyridinylimidazoles inhibition of CSBP kinase. *Bioorganic Med. Chem.* **1997**, *5*, 49–64.
- (51) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.
- (52) Pierre, F.; et al. Discovery and SAR of 5-(3-Chlorophenylamino)benzo[c][2,6]naphthyridine-8- carboxylic Acid (CX-4945), the first clinical stage inhibitor of protein kinase CK2 for the Treatment of Cancer. *J. Med. Chem.* **2011**, *54*, 635–654.
- (53) Battistutta, R.; et al. Unprecedented Selectivity and Structural Determinants of a New Class of Protein Kinase CK2 Inhibitors in Clinical Trials for the Treatment of Cancer. *Biochemistry* **2011**, *50*, 8478–8488.
- (54) Ruiz-Carmona, S.; et al. Dynamic undocking and the quasi-bound state as tools for drug discovery. *Nat. Chem.* **2016**, *9*, 201.
- (55) Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. PLIP: Fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* **2015**, *43*, W443–W447.
- (56) Deng, Z.; Chuaqui, C.; Singh, J. Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein–Ligand Binding Interactions. *J. Med. Chem.* **2004**, *47*, 337–344.
- (57) da Silva Neto, A. M.; Silva, S. R.; Vendruscolo, M.; Camilloni, C.; Montalvão, R. W. A superposition free method for protein conformational ensemble analyses and local clustering based on a differential geometry representation of backbone. *Proteins Struct. Funct. Bioinforma.* **2019**, *87*, 302–312.
- (58) Bradley, A. R.; et al. WONKA: objective novel complex analysis for ensembles of protein-ligand structures. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 963–973.
- (59) Subramanian, V.; Prusis, P.; Pietilä, L.-O.; Xhaard, H.; Wohlfahrt, G. Visually Interpretable Models of Kinase Selectivity Related Features Derived from Field-Based Proteochemometrics. *J. Chem. Inf. Model.* **2013**, *53*, 3021–3030.
- (60) Korb, O.; et al. Interactive and Versatile Navigation of Structural Databases. *J. Med. Chem.* **2016**, *59*, 4257–4266.