Published in final edited form as:

Chem Biol Drug Des. 2016 January ; 87(1): 5–20. doi:10.1111/cbdd.12631.

Considerations of Protein Subpockets in Fragment-Based Drug Design

Matthew Bartolowits and V. Jo Davisson*

Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, Purdue University, 575 Stadium Mall Dr., West Lafayette, IN 47907, USA

Abstract

While the fragment-based drug design approach continues to gain importance, gaps in the tools and methods available in the identification and accurate utilization of protein subpockets have limited the scope. The importance of these features of small molecule–protein recognition is highlighted with several examples. A generalized solution for the identification of subpockets and corresponding chemical fragments remains elusive, but there are numerous advancements in methods that can be used in combination to address subpockets. Finally, additional examples of approaches that consider the relative importance of small-molecule co-dependence of protein conformations are highlighted to emphasize an increased significance of subpockets, especially at protein interfaces.

Keywords

computational screening; fragment; fragment-based drug design; subpocket

Fragment-based drug design (FBDD) is an important strategy in both industry and academia for the discovery of novel ligands and aids the progression toward lead compounds (1). It is based on the idea that through the use of low molecular weight chemical fragments, which typically only bind weakly to their intended target, higher affinity lead ligands can be obtained by combining or ‘growing’ these small compounds into larger drug-like molecules. Due to the high levels of diversity between biological targets, incorporating FBDD as a high-throughput screening tool can have significant advantages over traditional higher molecular weight chemical libraries (2). The approach considers factors such as compound availability, ease of synthesis, large chemical space, and limits on steric ‘bulki-ness,’ which may otherwise preclude many higher molecular weight ligands from recognizing non-covalent enthalpically driven affinity factors (e.g. hydrogen bonding, etc.) at a target-binding site. Indeed, FBDD can prove to be robust for rational fragment identification in the absence of 3D structural data (3). However, FBDD still faces several challenges, such as the general lack of accountability for ligand specificity or selectivity (4,5), and the fact that key interactions and geometry of an original fragment hit may need to be changed when

*Corresponding author: V. Jo Davisson, davisson@purdue.edu.

^aMEDIT-SA [Internet]. Available from: <http://medit-pharma.-com/> [cited October 1 2014].

incorporated into a lead compound (6). Furthermore, the role of ligand-dependent receptor conformations has been largely untested.

A number of reviews have discussed recent advances in fragment-based drug design and how these tools can be used to improve the lead design process (4,7–14). Yet, relatively few evaluations have appeared which address the importance of small localized environments in a protein-binding site, and how microenvironments ultimately drive ligand binding and observed effects such as drug non-specificity. The field of drug discovery has long struggled with the accurate prediction of a drug's cross-pharmacologic profile (15–18) and side-effects. However, recent advances offer avenues toward understanding the significance of protein 'subpockets'—the physical, chemical, and geometric properties surrounding an individual residue. Traditional methods of computational analysis to find binding sites have sought to identify the similarity between proteins based on inherent sequence conservation or overall structural similarity. However, the localized chemical environments to be sampled by chemical fragment screens are potentially vast. Further development of generalized methods for the discovery of similar protein-/ligand-binding sites and predicting the interaction profile of molecular fragments remains of high interest (19).

Recently, multiple methods have been developed that compare proteins based on localized environments within binding sites, or the chemical environment around protein-bound ligands derived from PDB database crystal structures. These methods take one of two general approaches: (i) proteins are compared and binding sites organized based on their structural or chemical similarities, or (ii) data from known ligand–protein interactions are used to identify potentially similar sites in other proteins that could bind equivalent chemical fragments. This review aims to provide an overview of the background and emerging computational approaches that can define, and take into account, the significance of protein subpockets in the design of new chemical probes and pharmaceuticals. Many of these same tools offer methods which can also enhance understanding of drug activity. The relevance of new methodological insights into fragment-based drug discovery and the implications for lead development when considering factors such as non-specificity and side-effects are also considered.

Protein Fragment Recognition in Drug Design

Concept of shared subpockets

In many modern cases of structure-based drug discovery, structural information for a target in question drives the ligand development process (20,21). The RCSB protein data bank (PDB) (22) currently contains more than 89 000 structures solved by X-ray crystallography, with more than 10 000 others solved through other means (NMR, electron microscopy, etc.). This knowledge database, which continues to expand, is a vital tool for understanding the general structures of target proteins and the topology of target ligand-binding sites. Among 'sibling' members of a protein subfamily, binding sites may have a moderate-to-high level of overall similarity, such as the ATP-binding sites in the large family of protein kinases (23). Although, even in cases where two arbitrarily selected proteins have overall dissimilar ligand-binding sites, they may still bind the same chemical fragments if they share similar or equivalent topo-logical features at the subpocket level (see Figure 1 for an example). This

explains why two binding sites with substantially different sequences can bind identical chemical fragments, such as the case with the trifluoromethyl- and sulfonyl-binding subpockets between isoenzymes of the carbonic anhydrase family and cyclooxygenase-2 (24). Convergent evolution in nature is commonly observed to reveal similar enzyme active sites between proteins (25). Likewise, divergent evolution can result in shared ligand-binding sites among proteins that have highly dissimilar overall structures (26). Between proteins even of distant origin, similar ligand-binding motifs containing conserved consecutive residues can be found, indicating that it is quite efficient for nature to reuse localized features to bind similar molecules, even when the functional relationships may not be clear.

The principle of protein subpockets requires definition at the level of individual amino acids within a protein. Each residue is surrounded in three dimensions by other amino acid residues, water molecules, and/or metal ions. The resultant microenvironments create the recognition features for a particular chemical fragment, albeit with likely moderate ranges of affinities. These localized environments created by the protein generally define what are termed 'subpockets' within a binding cavity. Even within the limits of 20 different standard amino acids and the amide backbone, a substantial number of distinct subpockets with unique 'fingerprints' are possible. These subpockets differ by the relative proximity of specific amino acids to each other and distances between chemical functional groups. The amount of chemical space covered by chemical fragments is vast, but not all of them comprise preferable features for protein binding (27). Therefore, any given chemical fragment is not guaranteed to bind to a protein subpocket across the entire proteome.

Fragment diversity and protein interactions

For a chemical fragment to bind with favorable free energy, it must have the appropriate overall shape, proper spacing of its chemical functional groups, and generally compatible hydrophobic features to establish meaningful lifetimes within a subpocket. How these characteristics overlap with the physiochemical properties associated with the characteristics of drug-like molecules warrants consideration. In principle, chemical fragments can represent high diversity in shape and chemical features, but only a subset of compounds that could be classified as 'fragments' are useful for the purposes of screening against biological targets. A study by Zuegg and Cooper (28) analyzed more than eight million unique compounds from more than 100 chemical vendors and found that only 400 000 of these passed the fragment-like filter for the analysis of drug- and lead-likeness based upon 'the rule of three' (molecular weight < 300 Da, ClogP < 3, the number of hydrogen bond donors and acceptors < 3, and the number of rotatable bonds < 3) (29). A large number of fragments are represented in this subgroup, but there is an argument for ignoring a majority of commercially available compounds and enhancing practicality in a fragment-based screening platform. Of course, rules often have exceptions, and there is evidence that simply adhering to 'the rule of three' for the design of a fragment library may preclude compounds that would otherwise show up as hits (30).

To further facilitate the fragment library design process, there are a number of studies that have examined the binding preferences of proteins for chemical fragments (Table 1). By

analyzing fragments from lead-like compounds in the PDB (22), Chan *et al.* (31) found that particular types of fragments are more likely to interact with specific amino acid side chains (Asp, Glu, Arg and His) and engage in hydrogen bonding interactions. Separately, through the use of a developed algorithm, LigFrag-RPM, Wang *et al.* (32) mapped the interaction profiles of 315 unique fragments—derived from 71 798 different PDB ligands—against 20 naturally occurring amino acids, also identifying the preferences of fragment types for particular amino acids. This map could be used to determine whether a given fragment is in a favorable or unfavorable environment, and potentially guide the lead chemistry process.

While this may serve as a quick, efficient way to direct the design of ligand topology, the orientations of individual amino acids have to be taken into account. Even when comparing the same binding site between two crystal structures of the same protein, the microenvironments in that site may differ significantly based on residue orientations, and this can substantially bias docking studies. As an example, Cox *et al.* (33) noted that although co-crystal structures had been obtained of the G-protein-coupled receptor (GPCR), CXC-motif chemokine receptor 4 (CXCR4), bound to the anti-HIV small-molecule IT1t and peptide CVX15 (34), docking of IT1t back into the CXCR4: IT1t crystal structure resulted in large RMSD values compared to experimentally solved structures. Similarly, docking other small-molecule antagonists into the crystal structure gave poses that lacked critical interactions identified by site-directed mutagenesis (35,36). It was observed that the CXCR4:CVX15 (CVX15: cyclic peptide antagonist) crystal structure produced a result that showed agreement between the computational binding pose of another small-molecule CXCR4 inhibitor, AMD11070, and independent site-directed mutagenesis data. When analyzed in more detail, the orientations of residues Gln200 and His203 in the interior of the binding pocket differed substantially between the two co-crystal structures. Although the same amino acids are present at the binding site, changes in side chain orientation can substantially affect predicted binding modes and docking scores in virtual-compound screens, and may result in an unacceptably large false-negative rate, particularly when rigid protein docking approaches are used. Therefore, the most accurate representations of protein subpockets must take into account multiple possible rotameric states of amino acid side chains if binding site microenvironments are to be generalized as chemical spaces represented by single ‘fingerprints’. This presents a potential problem for many approaches to binding site comparison that make use of pre-existing ligand-bound data, as the amino acids that make up the microenvironment surrounding a ligand fragment are already conformationally biased in a given crystal structure.

While particular amino acids are seen to favor certain chemotypes, there exist some types of compounds that show relatively limited specificity and that can negatively affect fragment-based drug screening efforts. These molecules, being related to pan-assay interference compounds, are often widely promiscuous in the number and types of interactions with proteins and serve as artifacts, yielding false signals across many types of assays (5,37–41). There are varied reasons for the promiscuous nature of these compounds (42,43), but they represent a subclass of indirect observations of the potential ‘fuzziness’ in protein microenvironments. In essence, these compounds own a fragment chemical space that can recognize more general protein features that have recurring characteristics which contribute to the high level of promiscuity. When considered in this context, protein subpockets cannot

always be rigidly defined by a size and shape boundary as some chemical functionalities are seen to interact with less well-defined features.

Identifying Subpockets

In reality, defining what constitutes a ‘subpocket’ for fragment screening is not always a clear task. A significant number of topological features on a protein could potentially be classified as being part of a subpocket if one were to focus on the environment of a single amino acid. A common approach for defining pockets is to focus on the proximity of protein atoms or residues to a bound ligand. There have been many studies dealing with the analysis of binding sites, ranging from sequence comparisons, to overall binding site structural similarity and pharmacophore (‘fingerprint’) searching (8,44–54). Until recently, the majority of the work in the field has looked at the relationship between binding sites in the context of an entire defined pocket. Traditional computational methods have attempted to reduce the representation of protein surfaces in an effort to maximize the speed with which a search query can be performed [for examples, see Jordan *et al.* (55) and Xie & Bourne (56)]. It is likely that higher degrees of accuracy in predicting the similarity of binding sites can be obtained by taking into account more in-depth features such as hydrophobicity, aromaticity, and hydrogen-bonding capability at the level of single amino acids. As a result, attention in the field has turned to studying local areas of protein surfaces, with implications for traditional receptor-binding sites as well as protein–protein interaction interfaces. Table 2 highlights many of the currently available programs, methods, or databases used to screen binding sites, and while there is diversity between each of these, most can be generalized into one of two groups: protein topology- and ligand interaction-driven approaches.

Using protein topology to search for binding sites

A significant number of methods currently exist for comparing protein-binding sites (46). Some of these detect locally conserved residue patterns to define similar topological features between proteins and their probability of containing equivalent interaction sites. These approaches can function independently of known ligand-binding data, as the protein structure drives the query. A web server such as ProBiS (83) is useful for detecting either global or local similarities between proteins and identifying structurally conserved binding sites on proteins of interest. This tool takes into account geometric as well as physicochemical properties for aligning the local structure of two proteins; a user can decide to compare the entire surface of a protein, or only a selected surface patch, against a database of more than 35,000 non-redundant structures. The program COFACTOR (64) is able to identify functional sites on a protein as well as predict its biological function by comparing the local and global 3D structure of a protein against the large BioLiP database (95). Similarly, the MAAAGINE web server (76) can be used to define an arrangement of 3–8 amino acid residues that are converted into a search pattern to query the PDB and identify structural motifs with equivalent localized environments among the arrangement of amino acids within a defined tolerance distance (default of 1.5 Å). These examples demonstrate larger-scale (whole-protein) methods of comparison that identify conserved features of binding site topologies between proteins.

While these strategies can prove to be quite effective in certain circumstances, some of these ‘globalized’ methods may not be able to recognize distant similarities, particularly in cases where the dynamics of a protein allow a ligand to bind at sites that may be very structurally diverse. Recent tools have been developed to approach protein comparisons and binding site identification by taking a ‘subpocket-focused’ approach. Based on the 3D coordinates of a protein, DoGSiteScorer (66-68) is able to detect potential pockets on the surface and split each of them into subpockets. The program uses properties such as volume, depth, surface, ellipsoid main axes, and site-lining atoms and residues, as well as functional groups present to calculate the predicted pockets. This tool scores each of these pockets for their potential druggability, yielding accuracies of 88% when trained and tested against a dataset of 1069 different structures (67).

Among some of the other methods (81,96-98) used to compare binding sites are CavBase (59-63) and PrISE (55). CavBase derives 3D descriptors that characterize the surface properties of a binding cavity. The individual amino acids lining a cavity are analyzed to form descriptors of the localized chemical environment. Dummy atoms (‘pseu-docenters’) are placed on the surface, representing the overall chemical property expressed by the surrounding exposed atoms in that area. Therefore, cavities described by a series of pseudocenters are compared with a database [Relibase (99)] of cavities from PDB protein structures. Similarity is determined through the matching of 3D property descriptors (pseudocenters) between the different sites to provide a ‘pharmacophore-esque’ search that takes advantage of localized environments within a protein-binding cavity. Additionally, a new evaluation formalism for entries in Cavbase has recently been reported that improves efficiency of large-scale mining for similar protein-binding pockets (100). Separately, PrISE is able to predict interface residues of protein-protein interactions by defining structural elements consisting of a central amino acid residue and its surrounding residues on the protein surface. PrISE deconstructs the surface of a query protein into its structural elements and compares those to similar elements in a database that contains elucidated structural elements in PDB format. Elements are labeled as interface or non-interface based on the characteristics of the central residue. The labels are then used to predict whether the central residue of each structural element of the query protein is an interface residue. Ultimately, although different, CavBase and PrISE are able to use the chemical environment around a specific amino acid residue to predict the similarity between sites of interaction on different proteins.

For analyzing similarities between binding sites in terms of subcavities, the PocketFEATURE algorithm (80) takes a modified approach by analyzing ‘microenvironments’ to assess overall similarity between proteins for prediction of shared ligands. The method does not rely on the sequence or relative pattern of amino acids in a binding site. Instead, Liu & Altman define a ‘microenvironment’ as a local, spherical region within a protein structure that may encompass amino acid residues that are discontinuous in sequence and structure. A set of 80 physiochemical properties (101) is calculated over six concentric spherical shells, centered on a predefined functional center, with the total radius of the ‘microenvironment’ being 7.5 Å. Two sites are compared using an adjusted Tanimoto coefficient based on the presence of similar properties. Advantageously, this algorithm relies weakly on geometric requirements, instead using biophysical and biochemical

measurements to characterize a subregion of a binding pocket. This approach allows for dynamic conformational changes in both the ligand at a binding site and the pocket itself. In this way, the algorithm may most closely align itself with a true subpocket-type search when compared to the aforementioned computational methods.

Ligand interaction approaches to compare binding sites

As opposed to protein structure-driven approaches, ligand-driven approaches utilize known interaction data of ligands (small molecules and chemical fragments alike) with proteins to develop binding models that can be used to compare binding sites. Traditionally, many of these methods analyze the environment surrounding a bound ligand in a protein's binding site, encoding physiochemical and/or geometric features to form a general pharmacophore. These descriptors can then be screened against databases of known structures to detect similar sites. A program such as G-LoSA (74) uses this general strategy to predict or design a ligand for a target protein given known interaction data between ligands and predicted similar binding sites. The region surrounding a bound ligand is compared to localized regions in other proteins, and those containing a large set of aligned residues are identified. Similarly, GIRAF (72,73) utilizes a database of known ligand-protein structures to create an index of the geometric features of the surrounding atomic environment. Similar ligand-binding sites can then be identified and aligned with the query structure, independent of sequence homology, or protein fold. PROLIX (84) uses fingerprints of ligand-protein interaction patterns to rapidly mine large crystal structure databases for similar patterns. The PoSSuM database (82) enables the rapid exploration of similar binding sites between proteins based on physiochemical and geometric similarities, sorted using the neighbor search algorithm SketchSort (102). Each of these methods is useful for finding similar binding sites to a query protein in a small amount of time based on already-known interaction data, but these generally take whole-site approaches as opposed to screening small microenvironments of the binding site itself.

As with protein topology-driven approaches for binding site analysis, recent methods utilizing ligand-bound data are advancing the analysis of subpockets. The web tool Patch-Surfer (78) represents a binding pocket as a set of small, localized surface patches. Each of these patches is further characterized using 3D Zernike descriptors (103), enabling the identification of corresponding regions in pockets on other proteins, even if the overall pocket shapes are different. This method is able to compare a queried pocket to known ligand-binding pockets and predicts binding ligands for the query. In its most recent version, Patch-Surfer was tested against a large dataset of more than 6000 non-redundant pockets, with 2707 different ligands, where it displayed better predictive performance than many other currently available methods to predict protein pockets. Another program, CrystalDock (65), is able to take a ligand-receptor complex, break the ligand into its constituent molecular parts, use the microenvironments surrounding the ligand fragments, and perform a geometric comparison to identify similar microenvironments in ligand-bound PDB structures. This information can be used to predict chemical fragments that would bind a site of interest. FragFEA-TURE (71) has a similar function in that it also compares the structural environments within a target protein to those in databases to find statistically preferred fragments at a binding site. KRIPO (48) is able to use microenvironment pharmacophore

fingerprints to identify similar binding sites between proteins, and potential bioisosteric replacements for queried molecular fragments. Combining a subcavity comparison search with pharmacophoric analysis, Wallach and Lilien developed a method (94) to cluster similar binding site subcavities to predict patterns of binding between proteins that do not share any structural similarity with known systems.

Taking a modified approach to ligand-driven subpocket analysis, Kalliokoski et al. (90) developed SubCav, a tool for comparing and aligning protein subpockets. In this method, a modified version of the 3D pharmacophore fingerprint descriptor F-SPE-FP-PH3 (69) was used to define pharmacophoric features of all protein atoms within 4.5 Å of a bound ligand. These atoms were described according to eight different chemical features. Fingerprints composed of 7680 elements incorporating pharmacophoric features and the three dimensional triangular distances between them were generated for a grouping of subpockets within a binding site. Normalized Tanimoto scores were then used to assess the similarity between two fingerprints. Subpockets were aligned using the methodology described by Kabsch (104) implemented in BioPython (105). An advantage of this method is that it can identify similar fragment-binding sites independent of protein structural or sequence similarity. This focus on local equivalent microenvironments enables more comprehensive predictions of the pharmacological selectivity for on- and off-target binding events for fragments in ligands.

Shared Subpockets in ATP-Binding Sites and Relevance to Ligand Selectivity

A current challenge for small-molecule lead development is represented by target site similarities within the super-family of human protein kinases. Major efforts for drug discovery targeting these kinases have focused on their ATP-binding sites. Although these proteins have evolved distinct cellular functions and hence substrates, there exists a significant level of similarity at the ATP-binding sites. One of the challenges in developing inhibitors to these proteins is the optimization of selectivity for one or more related kinases (106). A study by Anastassiadis *et al.* (107) showed that a large proportion of commercially available kinase inhibitors display significant levels of cross-reactivity, with some having the ability to inhibit the catalytic activity by more than 50% in nearly 30% of all tested kinases. The clinical implications for these observations are not always clear, but offers a platform for understanding the role of subpockets in rendering predictable inhibitor kinase cross-reactivity.

Within kinase ATP-binding sites, there exists a conserved phenylalanine residue of the Asp-Phe-Gly (DFG) motif that is buried in a hydrophobic pocket, which is positioned in a groove between the two 'lobes' of the kinase. This motif is targeted by type I kinase inhibitors (108,109), being generally more promiscuous than type II and III inhibitors, which typically take advantage of an allosteric binding site that becomes available upon a structural change of the DFG motif (108–110). The idea of a protein subpocket finds significant application here given the large similarities in the binding pockets for type I inhibitors. An example can be seen with the JNK3 human kinase, where the residues Glu147, Met149, and Val196 form a microenvironment (Figure 2C) that recognizes shared chemical features between a natural

ligand, adenosine monophosphate (AMP) (111), and a dihydroanthrapyrazole-based inhibitor (112), though their structures, overall, are quite different.

These binding site subpockets are used advantageously when designing mimics of natural ligands as inhibitors, but the similarities in these microenvironments between kinases can also give rise to promiscuous chemical landscapes. Staurosporine is a classical ATP-competitive inhibitor that is notoriously promiscuous (113), with there already being 48 crystal structures available in the PDB that show it bound to a variety of human kinases. With roughly 32% sequence similarity, the TAO2 and serine–threonine kinase 16 (STK16) kinases both bind staurosporine with good affinity. The sequence of residues at the ligand-binding site differs between the two proteins, but TAO2 (114) and STK16 (115) both fold in a way that causes chemically similar amino acids to overlap, forming shared subpockets that result in a favorable binding site for the competitive inhibitor (Figure 2A).

There are additional proteins in this family that display high levels of promiscuity as well. The RET tyrosine kinase has been observed to be significantly inhibited by a variety of competitive inhibitors (107). While many of these ligands have structurally distinct features and vary in size, one conserved binding orientation (116) makes use of an overall hydrophobic subpocket surrounding Val804, as can be seen in Figure 2B. Even though the inhibitors shown in this example are not vastly dissimilar, they do demonstrate that when designing ligands, consideration of the fact that some proteins recognize more generalized chemical features. This insight is especially important when considering kinases as biological targets, but importantly the example is instructive to apply to other protein families as well. Furthermore, documented ligand-kinase associations offer many cases to benchmark tests of methodologies to identify and utilize fragment subpockets for the desired selectivity or cross-reactivity.

Consideration of Protein Interfaces and Cooperative Interactions

In contrast to the types of ligand interaction sites considered so far, protein-protein interaction (PPI) interfaces give fewer clues on a residue-by-residue basis for what is required for a specific binding event. Protein surfaces most often are relatively flat, and the interface at which two proteins engage typically makes up a large surface area (1000-2000 Å²) with multiple contact points (117). Drug discovery efforts, particularly in the pharmaceutical industry, traditionally avoided exploring PPIs as drug targets, considering them ‘undruggable’ until recent challenges of this classification (118). Within the last decade, research into targeting PPIs as a therapeutic strategy for multiple diseases has continued to accelerate. Previous work has shown that not all of the residues at a PPI interface are critical for the interaction. In the best studied cases to date, there exist regions of ‘hot spots’ that confer most of the binding energy (119). These ‘hot spots’ are adept at binding protein or peptide-esque molecules and have been suitable targets for peptides and a number of mimetic compounds (120). However, the practical utility of the hot spot theory requires additional consideration with respect to the design of specific protein–ligand interactions.

Previous studies have concluded that amino acids such as Ile, Leu, Met, Phe, Trp, Tyr, and Val appear more frequently at PPI interfaces relative to other amino acids, and their average

distribution throughout the genome (121,122). Some residues at interface hot spots form ‘anchor sites,’ where they serve as critical recognition features that drive the binding process (123). In this way, these sites are analogous to the previously discussed concept of a subpocket in that the localized environment and relative orientation of particular amino acids in anchoring sites are important for the recognition of specific chemo-types. The feasibility of targeting anchor sites was demonstrated when developing small-molecule PPI inhibitors (122). The surface of the E3 ubiquitin–protein ligase MDM2 forms a natural binding groove that the Phe19, Trp23, and Leu25 residues of p53 favorably bind (Figure 3A). Even when visualizing the unbound form of MDM2, the pocket where Trp23 of p53 ultimately binds is visible, suggesting that this site is important for the recognition and binding of p53. Phe19 and Leu25, upon binding, appear to induce the formation of hydrophobic pockets on MDM2 that result in favorable binding environments for the two residues. In this case, the recognition features of a subpocket are not necessarily prearranged for a meaningful binding event to occur. Rather, a molecular fragment may induce the formation of a new subpocket with optimal chemical environment for that fragment to bind, when it otherwise would not if considered a static system.

In a similar fashion, proliferating cell nuclear antigen (PCNA) contains a region on its surface that recognizes specific chemical features. PCNA acts as a homotrimeric scaffold protein that binds many different proteins associated with DNA replication and damage repair (124). Many of these proteins contain a conserved eight-member sequence known as the PCNA-interacting protein (PIP box) motif (125). While four of the eight amino acids in this sequence are highly variable, the remaining four residues represent conserved chemical functionalities that are critical for proteins to bind at the PIP box-binding site on PCNA (Figure 3B). The conserved aromatic residues (tyrosine or phenylalanine) in the seventh and eighth positions of the PIP box are part of a short 3_{10} helix that binds in a highly flexible (126) hydrophobic pocket on the surface of PCNA. This site generally recognizes hydrophobic molecular fragments exemplified by the discovery of a small molecule, T2AA, that binds in that pocket and inhibits the interaction of PCNA with other PIP box-containing proteins (127). Upon comparison between the structures of PCNA bound to p21 (PDB ID: 1AXC) (128) and T3 (an analog of T2AA; PDB ID: 3VKX) (127), a subpocket formed by Ile128, Tyr133, Tyr250, Pro234, and Val236 binds an aromatic fragment of both T3 and p21’s PIP box tyrosine (Figure 3C). Furthermore, the inhibitor’s core iodine atoms induce a structural widening of the binding pocket, exposing a larger hydrophobic surface, resulting in a stably bound planar orientation of the molecule. The ligand dependence of the subpocket within the PIP box-binding region is another example of how the plasticity of protein interfaces must be taken into consideration.

Because drug-like compounds are often composed of several segmental fragments, any one substructure of a molecule could have affinity for a subpocket fingerprint shared between two or more proteins. Of course, the observable binding affinity of the drug molecule depends on all of its interacting components with a protein; a single interaction within a microenvironment may not be nearly sufficient enough to stabilize a specific bound conformation to elicit a pharmacological effect. However, protein systems do exist that have evolved to accept a variety of chemical structures, whether through flexible binding sites, or weak requirements for a binding event to produce a desired effect. The large family of

GPCRs that make up the mammalian olfactory system is one example. Many individual members of this group become activated upon interaction with a plethora of different odorants (129–132) (Figure 3D), implicating the existence of ‘fuzzy subcavities’ making up interaction sites. Receptor OR1G1, and others like it, thus exhibits flexible recognition of general chemo-types, rather than a specific orientation of atoms.

The aforementioned examples demonstrate that PPI interfaces containing subpockets are analogous to those found in traditional small ligand-binding sites. However, these characteristics indicate the transient aspects of these sub-pockets and articulate both opportunities and challenges for discovery of selective ligands with useful pharmacological effects. Many cases highlight limitations of structure-based virtual screening, and improved methods to achieve overall accuracy continue to emerge (133–138). Computational approaches to address the issues of ligand and receptor dynamics has been an area of continued advancement (10,133,134,136,139,140). The distinctions are likely that induced-fit pathways and conformational selection play more significant roles in the formation of subpockets and the resulting affinity of a molecule composed of multiple substructures (10,141–143). These mechanistic principles have recently been exemplified in the formation of transient binding sites for protein interface inhibitors where evidence for both local conformational selection and induced-fit pathways have been evaluated (141).

An understanding of how subpockets are influenced by other microenvironments in proximity is a general characteristic for consideration in FBDD. When a fragment of a small molecule binds in a given subpocket, the binding event can induce structural changes that modulate the features of surrounding regions. This means that in a fragment-based screen, suitable binding in a particular chemical microenvironment may not occur in the absence of nearby transient binding events of other fragments, resulting in false negatives. A recent case study addressed how current FBDD methods are limited in the detection of fragments that exhibit cooperative binding (144). However, the deployment of these methods to provide improved accuracy in fragment-based screens is less than complete.

Several computational approaches to meet these limitations are emerging and show significant promise for expanding successes. Mahasenan & Li were able to develop an approach that incorporates multiple protein conformers via ligand-directed modeling (145,146). Substantial improvements for sampling and free energy scores in docking studies can be achieved when taking into account multiple fragment ligands simultaneously (147,148). Novel inhibitors of STAT3 and the IL-6/GP130 protein interface were discovered by considering multiple subpockets, and simultaneous ligand screening (149–152). Type II kinase inhibitors were also developed for the MELK kinase by taking into account an induced conformational shift of the protein (153). These examples implicate the importance for ligand-dependent subpockets when using FBDD approaches; they may represent a general process that can allow for cooperative binding into subpockets. The full extent to which this impacts current drug discovery efforts and computational screening methods remains to be fully established, but opens avenues for improvement in the utilization of computational and experimental approaches to discover multiple fragment sites at protein interfaces.

Summary

There are an increasing number of computational tools available to complement experimental fragment-based screens and ligand design. Advantages to these approaches are that they enable a potential general solution to the challenges of optimal fragment selectivity and linkages, but these goals remain to be fulfilled. The prediction of protein structure microenvironments and dynamics requires consideration, and methodological advances are emerging. As previously highlighted, the similarities of protein microenvironments between evolutionarily related or convergent proteins could argue for increased chances of off-target effects. To the extent that this may be a concern at protein interfaces still remains largely unknown (118,119). However, the potentially unique combinations of transient subpockets that arise upon the binding of a ligand may ultimately result in higher degrees of selectivity.

The occurrence of subpockets in unrelated proteins is generally recognized as a potential contribution to drug effects; however, an account of these features when selecting specific fragments in drug development has been traditionally based upon empirical deduction. Experimental and computational approaches that can account for potential cooperative interactions are subjects of considerable importance for structure-based drug design. It is increasingly feasible using computational approaches to consider a classification process to stratify both risk and benefit of new molecular candidates. These tools can also be applied to molecular fragments and their associated libraries. As opposed to promiscuous scaffolds, fragments with high propensity for subpocket interactions can be argued to represent important tools for lead identification when considered in an appropriate knowledge framework. Whether fragment libraries can be qualified on a pro-teome-wide scale remains a future challenge, but the combinations of experimental and computational methods offers increasing promise.

Declaration of Interest

The authors declare no competing interests in the content of this paper. Support is recognized for a Purdue Research Foundation Fellowship (MDB), the Department of Defense Breast Cancer Research Program for Award W81XWH-10-0105 (VJD), and the Purdue Center for Cancer Research via an NIH NCI grant (P30CA023168).

References

1. Erlanson DA. Introduction to fragment-based drug discovery. *Top Curr Chem*. 2012; 317:1–32. [PubMed: 21695633]
2. Erlanson DA, McDowell RS, O'Brien T. Fragment-based drug discovery. *J Med Chem*. 2004; 47:3463–3482. [PubMed: 15214773]
3. Henen MA, Coudeville N, Geist L, Konrat R. Toward rational fragment-based lead design without 3D structures. *J Med Chem*. 2012; 55:7909–7919. [PubMed: 22889313]
4. Sheng C, Zhang W. Fragment informatics and computational fragment-based drug design: an overview and update. *Med Res Rev*. 2013; 33:554–598. [PubMed: 22430881]
5. Hu Y, Bajorath J. What is the likelihood of an active compound to be promiscuous? Systematic assessment of compound promiscuity on the basis of pubchem confirmatory bioassay data. *AAPS J*. 2013; 15:808–815. [PubMed: 23605807]
6. Babaoglu K, Shoichet BK. Deconstructing fragment-based inhibitor discovery. *Nat Chem Biol*. 2006; 2:720–723. [PubMed: 17072304]

7. Schneider G. Future de novo drug design. *Mol Inform.* 2014; 33:397–402. [PubMed: 27485977]
8. Konc J, Janezic D. Binding site comparison for function prediction and pharmaceutical discovery. *Curr Opin Struct Biol.* 2014; 25:34–39. [PubMed: 24878342]
9. Joseph-McCarthy D, Campbell AJ, Kern G, Mous-takas D. Fragment-based lead discovery and design. *J Chem Inf Model.* 2014; 54:693–704. [PubMed: 24490951]
10. Feixas F, Lindert S, Sinko W, McCammon JA. Exploring the role of receptor flexibility in structure-based drug discovery. *Biophys Chem.* 2014; 186:31–45. [PubMed: 24332165]
11. Rognan D. Towards the next generation of computational chemogenomics tools. *Mol Inform.* 2013; 32:1029–1034. [PubMed: 27481148]
12. Jalencas X, Mestres J. Identification of similar binding sites to detect distant polypharmacology. *Mol Inform.* 2013; 32:976–990. [PubMed: 27481143]
13. Scott DE, Coyne AG, Hudson SA, Abell C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry.* 2012; 51:4990–5003. [PubMed: 22697260]
14. Kumar A, Voet A, Zhang KYJ. Fragment based drug design: from experimental to computational approaches. *Curr Med Chem.* 2012; 19:5128–5147. [PubMed: 22934764]
15. Mestres J, Gregori-Puigjané E, Valverde S, Solé RV. The topology of drug-target interaction networks: implicit dependence on drug properties and target families. *Mol BioSyst.* 2009; 5:1051–1057. [PubMed: 19668871]
16. Vogt I, Mestres J. Drug-target networks. *Mol Inform.* 2010; 29:10–14. [PubMed: 27463845]
17. Briano F, Carrascosa MC, Oprea TI, Mestres J. Cross-pharmacology analysis of G protein-coupled receptors. *Curr Top Med Chem.* 2011; 11:1956–1963. [PubMed: 21851335]
18. Jalencas X, Mestres J. On the origins of drug polypharmacology. *MedChemComm.* 2012; 4:80–87.
19. Schneider G. Virtual screening: an endless staircase? *Nat Rev Drug Discov.* 2010; 9:273–276. [PubMed: 20357802]
20. Eder J, Sedrani R, Wiesmann C. The discovery of first-in-class drugs: origins and evolution. *Nat Rev Drug Discov.* 2014; 13:577–587. [PubMed: 25033734]
21. Swinney DC, Anthony J. How were new medicines discovered? *Nat Rev Drug Discov.* 2011; 10:507–519. [PubMed: 21701501]
22. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res.* 2000; 28:235–242. [PubMed: 10592235]
23. Kinnings SL, Jackson RM. Binding site similarity analysis for the functional classification of the protein kinase family. *J Chem Inf Model.* 2009; 49:318–329. [PubMed: 19434833]
24. Weber A, Casini A, Heine A, Kuhn D, Supuran CT, Scozzafava A, Klebe G. Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. *J Med Chem.* 2004; 47:550–557. [PubMed: 14736236]
25. Gherardini PF, Wass MN, Helmer-Citterich M, Sternberg MJE. Convergent evolution of enzyme active sites is not a rare phenomenon. *J Mol Biol.* 2007; 372:817–845. [PubMed: 17681532]
26. Xie L, Bourne PE. Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments. *Proc Natl Acad Sci USA.* 2008; 105:5441–5446. [PubMed: 18385384]
27. Baskin I, Varnek A. Building a chemical space based on fragment descriptors. *Comb Chem High Throughput Screen.* 2008; 11:661–668. [PubMed: 18795885]
28. Zuegg J, Cooper MA. Drug-likeness and increased hydrophobicity of commercially available compound libraries for drug screening. *Curr Top Med Chem.* 2012; 12:1500–1513. [PubMed: 22827520]
29. Congreve M, Carr R, Murray C, Jhoti H. A “rule of three” for fragment-based lead discovery? *Drug Discov Today.* 2003; 8:876–877.
30. Koöster H, Craan T, Brass S, Herhaus C, Zentgraf M, Neumann L, Heine A, Klebe G. A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J Med Chem.* 2011; 54:7784–7796. [PubMed: 21972967]

31. Chan AWE, Laskowski RA, Selwood DL. Chemical fragments that hydrogen bond to Asp, Glu, Arg, and his side chains in protein binding sites. *J Med Chem*. 2010; 53:3086–3094. [PubMed: 20230000]
32. Wang L, Xie Z, Wipf P, Xie XQ. Residue preference mapping of ligand fragments in the protein data bank. *J Chem Inf Model*. 2011; 51:807–815. [PubMed: 21417260]
33. Cox BD, Prosser AR, Katzman BM, Alcaraz AA, Liotta DC, Wilson LJ, Snyder JP. Anti-HIV small-molecule binding in the peptide subpocket of the CXCR4:CVX15 crystal structure. *Chembiochem Eur J Chem Biol*. 2014; 15:1614–1620.
34. Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC. Structures of the CXCR4 chemokine receptor in complex with small molecule and cyclic peptide antagonists. *Science*. 2010; 330:1066–1071. [PubMed: 20929726]
35. Kawatkar SP, Yan M, Gevariya H, Lim MY, Eisold S, Zhu X, Huang Z, An J. Computational analysis of the structural mechanism of inhibition of chemokine receptor CXCR4 by small molecule antagonists. *Exp Biol Med*. 2011; 236:844–850.
36. Zhang C, Hou T, Feng Z, Li Y. Structure-based development of antagonists for chemokine receptor CXCR4. *Curr Comput Aided Drug Des*. 2013; 9:60–75. [PubMed: 22734712]
37. Baell J, Walters MA. Chemistry: chemical con artists foil drug discovery. *Nature*. 2014; 513:481–483. [PubMed: 25254460]
38. Pohjala L, Tammela P. Aggregating behavior of phenolic compounds — A source of false bioassay results? *Molecules*. 2012; 17:10774–10790. [PubMed: 22960870]
39. Mendgen T, Steuer C, Klein CD. Privileged scaffolds or promiscuous binders: a comparative study on rhodanines and related heterocycles in medicinal chemistry. *J Med Chem*. 2012; 55:743–753. [PubMed: 22077389]
40. Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem*. 2010; 53:2719–2740. [PubMed: 20131845]
41. Baell JB, Ferrins L, Falk H, Nikolakopoulos G. PAINS: relevance to tool compound discovery and fragment-based screening. *Aust J Chem*. 2013; 66:1483–1494.
42. Priyadarsini KI. Chemical and structural features influencing the biological activity of curcumin. *Curr Pharm Des*. 2013; 19:2093–2100. [PubMed: 23116315]
43. Ingolfsson HI, Thakur P, Herold KF, Hobart EA, Ramsey NB, Periole X, de Jong DH. Phytochemicals perturb membranes and promiscuously alter protein function. *ACS Chem Biol*. 2014; 9:1788–1798. [PubMed: 24901212]
44. Kellenberger E, Schalon C, Rognan D. How to measure the similarity between protein ligand-binding sites? *Curr Comput - Aided Drug Des*. 2008; 4:209–220.
45. Henrich S, Salo-Ahen OMH, Huang B, Rippmann FF, Cruciani G, Wade RC. Computational approaches to identifying and characterizing protein binding sites for ligand design. *J Mol Recognit JMR*. 2010; 23:209–219. [PubMed: 19746440]
46. Nisius B, Sha F, Gohlke H. Structure-based computational analysis of protein binding sites for function and druggability prediction. *J Biotechnol*. 2012; 159:123–134. [PubMed: 22197384]
47. Vulpetti A, Kalliokoski T, Milletti F. Chemogenomics in drug discovery: computational methods based on the comparison of binding sites. *Future Med Chem*. 2012; 4:1971–1979. [PubMed: 23088277]
48. Wood DJ, de Vlieg J, Wagener M, Ritschel T. Pharmacophore fingerprint-based approach to binding site subpocket similarity and its application to bioisostere replacement. *J Chem Inf Model*. 2012; 52:2031–2043. [PubMed: 22830492]
49. Walter P, Metzger J, Thiel C, Helms V. Predicting where small molecules bind at protein-protein interfaces. *PLoS ONE*. 2013; 8:e58583. [PubMed: 23505538]
50. Wirth M, Volkamer A, Zoete V, Rippmann F, Michielin O, Rarey M, Sauer WHB. Protein pocket and ligand shape comparison and its application in virtual screening. *J Comput Aided Mol Des*. 2013; 27:511–524. [PubMed: 23807262]
51. Volkamer A, Rarey M. Exploiting structural information for drug-target assessment. *Future Med Chem*. 2014; 6:319–331. [PubMed: 24575967]

52. Zheng X, Gan L, Wang E, Wang J. Pocket-based drug design: exploring pocket space. *AAPS J*. 2013; 15:228–241. [PubMed: 23180158]
53. Salentin S, Haupt VJ, Daminelli S, Schroeder M. Polypharmacology rescored: protein-ligand interaction profiles for remote binding site similarity assessment. *Prog Biophys Mol Biol*. 2014; 116:174–186. [PubMed: 24923864]
54. Haupt VJ, Schroeder M. Old friends in new guise: repositioning of known drugs with structural bioinformatics. *Brief Bioinform*. 2011; 12:312–326. bbr011. [PubMed: 21441562]
55. Jordan RA, EL-Manzalawy Y, Dobbs D, Honavar V. Predicting protein-protein interface residues using local surface structural similarity. *BMC Bioinformatics*. 2012; 13:41. [PubMed: 22424103]
56. Xie L, Bourne PE. A robust and efficient algorithm for the shape description of protein structures and its application in predicting ligand binding sites. *BMC Bioinformatics*. 2007; 8(Suppl. 4):S9.
57. Nilmeier JP, Kirshner DA, Wong SE, Lightstone FC. Rapid catalytic template searching as an enzyme function prediction procedure. *PLoS ONE*. 2013; 8:e62535. [PubMed: 23675414]
58. Kirshner DA, Nilmeier JP, Lightstone FC. Catalytic site identification—a web server to identify catalytic site structural matches throughout PDB. *Nucleic Acids Res*. 2013; 41(W1):W256–W265. [PubMed: 23680785]
59. Hendlich M, Rippmann F, Barnickel G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J Mol Graph Model*. 1997; 15:359–363. 389. [PubMed: 9704298]
60. Schmitt S, Hendlich M, Klebe G. From structure to function: a new approach to detect functional similarity among proteins independent from sequence and fold homology. *Angew Chem Int Ed*. 2001; 40:3141–3144.
61. Schmitt S, Kuhn D, Klebe G. A new method to detect related function among proteins independent of sequence and fold homology. *J Mol Biol*. 2002; 323:387–406. [PubMed: 12381328]
62. Bergner, A., Günther, J. Structural aspects of binding site similarity: a 3D upgrade for chemogenomics. In: Kubinyi, H., Müller, G., editors. *Chemogenomics in Drug Discovery: A Medicinal Chemistry Perspective*. Weinheim: Wiley-VCH Verlag GmbH & CoKGaA; 2004. p. 97-135.
63. Kuhn D, Weskamp N, Hüllermeier E, Klebe G. Functional classification of protein kinase binding sites using Cavbase. *ChemMedChem*. 2007; 2, 1432–1447.
64. Roy A, Yang J, Zhang Y. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res*. 2012; 40(Web Server issue):W471–W477. [PubMed: 22570420]
65. Durrant JD, Friedman AJ, McCammon JA. CrystalDock: a novel approach to fragment-based drug design. *J Chem Inf Model*. 2011; 51:2573–2580. [PubMed: 21910501]
66. Volkamer A, Griewel A, Grombacher T, Rarey M. Analyzing the topology of active sites: on the prediction of pockets and subpockets. *J Chem Inf Model*. 2010; 50:2041–2052. [PubMed: 20945875]
67. Schmidtke P, Barril X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J Med Chem*. 2010; 53:5858–5867. [PubMed: 20684613]
68. Volkamer A, Kuhn D, Grombacher T, Rippmann F, Rarey M. Combining global and local measures for structure-based druggability predictions. *J Chem Inf Model*. 2012; 52:360–372. [PubMed: 22148551]
69. Vulpetti A, Schiering N, Dalvit C. Combined use of computational chemistry, NMR screening, and X-ray crystallography for identification and characterization of fluorophilic protein environments. *Proteins Struct Funct Bioinforma*. 2010; 78:3281–3291.
70. Brylinski M, Skolnick J. FINDSITE_LHM: a threading-based approach to ligand homology modeling. *PLoS Comput Biol*. 2009; 5:e1000405. [PubMed: 19503616]
71. Tang GW, Altman RB. Knowledge-based fragment binding prediction. *PLoS Comput Biol*. 2014; 10:e1003589. [PubMed: 24762971]
72. Kinjo AR, Nakamura H. Similarity search for local protein structures at atomic resolution by exploiting a database management system. *Biophysics*. 2007; 3:75–84. [PubMed: 27857569]
73. Kinjo AR, Nakamura H. Comprehensive structural classification of ligand-binding motifs in proteins. *Structure*. 2009; 17:234–246. [PubMed: 19217394]

74. Lee HS, Im W. Identification of ligand templates using local structure alignment for structure-based drug design. *J Chem Inf Model*. 2012; 52:2784–2795. [PubMed: 22978550]
75. Davis FP. Proteome-wide prediction of overlapping small molecule and protein binding sites using structure. *Mol BioSyst*. 2011; 7:545. [PubMed: 21103609]
76. Nadzirin N, Willett P, Artymiuk PJ, Firdaus-Raih M. IMAAAGINE: a webserver for searching hypothetical 3D amino acid side chain arrangements in the Protein Data Bank. *Nucleic Acids Res*. 2013; 41(Web Server issue):W432–W440. [PubMed: 23716645]
77. Parca L, Ferre F, Ausiello G, Helmer-Citterich M. Nucleos: a web server for the identification of nucleotide-binding sites in protein structures. *Nucleic Acids Res*. 2013; 41(Web Server issue):W281–W285. [PubMed: 23703207]
78. Zhu X, Xiong Y, Kihara D. Large-scale binding ligand prediction by improved patch-based method Patch-Surfer2.0. *Bioinformatics*. 2015; 31:707–713. [PubMed: 25359888]
79. Parca L, Mangone I, Gherardini PF, Ausiello G, Helmer-Citterich M. Phosfinder: a web server for the identification of phosphate-binding sites on protein structures. *Nucleic Acids Res*. 2011; 39(Web Server issue):W278–W282. [PubMed: 21622655]
80. Liu T, Altman RB. Using multiple microenvironments to find similar ligand-binding sites: application to kinase inhibitor binding. *PLoS Comput Biol*. 2011; 7:e1002326. [PubMed: 22219723]
81. Yeturu K, Chandra N. PocketAlign a novel algorithm for aligning binding sites in protein structures. *J Chem Inf Model*. 2011; 51:1725–1736. [PubMed: 21662242]
82. Ito JI, Tabei Y, Shimizu K, Tsuda K, Tomii K. PoSSuM: a database of similar protein-ligand binding and putative pockets. *Nucleic Acids Res*. 2012; 40(D1):D541–D548. [PubMed: 22135290]
83. Konc J, Janezic D. ProBiS-2012: web server and web services for detection of structurally similar binding sites in proteins. *Nucleic Acids Res*. 2012; 40(Web Server issue):W214–W221. [PubMed: 22600737]
84. Weisel M, Bitter HM, Diederich F, So WV, Kondru R. PROLIX: rapid mining of protein-ligand interactions in large crystal structure databases. *J Chem Inf Model*. 2012; 52:1450–1461. [PubMed: 22582806]
85. Regad L, Martin J, Nuel G, Camproux AC. Mining protein loops using a structural alphabet and statistical exceptionality. *BMC Bioinformatics*. 2010; 11:75. [PubMed: 20132552]
86. Nuel G, Regad L, Martin J, Camproux AC. Exact distribution of a pattern in a set of random sequences generated by a Markov source: applications to biological data. *Algorithms Mol Biol*. 2010; 5:15. [PubMed: 20205909]
87. Teyra J, Doms A, Schroeder M, Pisabarro MT. SCOWLP: a web-based database for detailed characterization and visualization of protein interfaces. *BMC Bioinformatics*. 2006; 7:104. [PubMed: 16512892]
88. Sehnal D, Vareková RS, Huber HJ, Geidl S, Ionescu CM, Wimmerová M, Ko a J. SiteBinder: an improved approach for comparing multiple protein structural motifs. *J Chem Inf Model*. 2012; 52:343–359. [PubMed: 22296449]
89. Lin Y, Yoo S, Sanchez R. SiteComp: a server for ligand binding site analysis in protein structures. *Bioinformatics*. 2012; 28:1172–1173. [PubMed: 22368247]
90. Kalliokoski T, Olsson TSG, Vulpetti A. Sub-pocket analysis method for fragment-based drug discovery. *J Chem Inf Model*. 2013; 53:131–141. [PubMed: 23327721]
91. Wang L, Ma C, Wipf P, Liu H, Su W, Xie XQ. TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. *AAPS J*. 2013; 15:395–406. [PubMed: 23292636]
92. Kokh DB, Richter S, Henrich S, Czodrowski P, Rippmann F, Wade RC. TRAPP: a tool for analysis of transient binding pockets in proteins. *J Chem Inf Model*. 2013; 53:1235–1252. [PubMed: 23621586]
93. Von Behren MM, Volkamer A, Henzler AM, Schomburg KT, Urbaczek S, Rarey M. Fast protein binding site comparison via an index-based screening technology. *J Chem Inf Model*. 2013; 53:411–422. [PubMed: 23390978]

94. Wallach I, Lilien RH. Prediction of sub-cavity binding preferences using an adaptive physicochemical structure representation. *Bioinformatics*. 2009; 25:i296–i304. [PubMed: 19478002]
95. Yang J, Roy A, Zhang Y. BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res*. 2013; 41(Database issue):D1096–D1103. [PubMed: 23087378]
96. Feldman HJ, Labute P. Pocket similarity: are alpha carbons enough? *J Chem Inf Model*. 2010; 50:1466–1475. [PubMed: 20690656]
97. Yeturu K, Chandra N. PocketMatch: a new algorithm to compare binding sites in protein structures. *BMC Bioinformatics*. 2008; 9:543. [PubMed: 19091072]
98. Villoutreix BO, Lagorce D, Labbé CM, Sperandio O, Miteva MA. One hundred thousand mouse clicks down the road: selected online resources supporting drug discovery collected over a decade. *Drug Discov Today*. 2013; 18:1081–1089. [PubMed: 23831439]
99. Bergner A, Gunther J, Hendlich M, Klebe G, Verdonk M. Use of Relibase for retrieving complex three-dimensional interaction patterns including crys-tallographic packing effects. *Biopolymers*. 2001; 61:99–110. [PubMed: 11987159]
100. Krotzky T, Grunwald C, Egerland U, Klebe G. Large-scale mining for similar protein binding pockets: with RAPMAD retrieval on the fly becomes real. *J Chem Inf Model*. 2014; 55:165–179. [PubMed: 25474400]
101. Wei L, Altman RB, Chang JT. Using the radial distributions of physical features to compare amino acid environments and align amino acid sequences. *Pac Symp Biocomput*. 1997:465–476. [PubMed: 9390315]
102. Tabei, Y., Uno, T., Sugiyama, M., Tsuda, K. Single versus multiple sorting for all pairs similarity search. Tokyo, Japan. The 2nd Asian Conference on Machine Learning (ACML2010) JMLR: Workshop and Conference Proceedings; 2010. p. 145-160.
103. Chikhi R, Sael L, Kihara D. Real-time ligand binding pocket database search using local surface descriptors. *Proteins*. 2010; 78:2007–2028. [PubMed: 20455259]
104. Kabsch W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr Sect A*. 1976; 32:922–923.
105. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009; 25:1422–1423. [PubMed: 19304878]
106. Traxler P, Furet P. Strategies toward the design of novel and selective protein tyrosine kinase inhibitors. *Pharmacol Ther*. 1999; 82:195–206. [PubMed: 10454197]
107. Anastassiadis T, Deacon SW, Devarajan K, Ma H, Peterson JR. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol*. 2011; 29:1039–1045. [PubMed: 22037377]
108. Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, Grob PM, Hickey ER, Moss N, Pav S, Regan J. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat Struct Biol*. 2002; 9:268–272. [PubMed: 11896401]
109. Liu Y, Gray NS. Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol*. 2006; 2:358–364. [PubMed: 16783341]
110. Bogoyevitch MA, Fairlie DP. A new paradigm for protein kinase inhibition: blocking phosphorylation without directly targeting ATP binding. *Drug Discov Today*. 2007; 12:622–633. [PubMed: 17706543]
111. Han, BG., Shim, MB., Ahn, HC. The crystal structure of AMP-bound JNK3 [Internet]. RCSB PDB Protein Data Bank. 2013. Available from: <http://www.pdb.org/pdb/explore/explore.do?structureId=4KKE> [cited 10 November 2014]
112. Scapin G, Patel SB, Lisnock J, Becker JW, LoGrasso PV. The structure of JNK3 in complex with small molecule inhibitors: structural basis for potency and selectivity. *Chem Biol*. 2003; 10:705–712. [PubMed: 12954329]

113. Tanramluk D, Schreyer A, Pitt WR, Blundell TL. On the origins of enzyme inhibitor selectivity and promiscuity: a case study of protein kinase binding to staurosporine. *Chem Biol Drug Des*. 2009; 74:16–24. [PubMed: 19519740]
114. Tian-Jun Zhou LGS. Crystal structure of the MAP3K TAO2 kinase domain bound by an inhibitor staurosporine. *Acta Biochim Biophys Sin*. 2006; 38:385–392. [PubMed: 16761096]
115. Debreczeni, JE., Eswaran, J., Bullock, A., Filip-pakopoulos, P., Kavanagh, K., Amos, A., Fedorov, O., Sobott, F., Ball, LJ., Von Delft, F., Arrowsmith, C., Sund-strom, M., Edwards, A., Knapp, S. Crystal structure of the human serine-threonine kinase 16 in complex with staurosporine [Internet]. RCSB PDB Protein Data Bank. 2005. Available from: <http://www.pdb.org/pdb/explore/explore.do?structureId=2buj>
116. Knowles PP, Murray-Rust J, Kjær S, Scott RP, Hanrahan S, Santoro M, Ibáñez CF, McDonald NQ. Structure and chemical inhibition of the RET tyrosine kinase domain. *J Biol Chem*. 2006; 281:33577–33587. [PubMed: 16928683]
117. Hwang H, Vreven T, Janin J, Weng Z. Protein-protein docking benchmark version 4.0. *Proteins*. 2010; 78:3111–3114. [PubMed: 20806234]
118. Arkin MR, Tang Y, Wells JA. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem Biol*. 2014; 21:1102–1114. [PubMed: 25237857]
119. Arkin MR, Wells JA. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov*. 2004; 3:301–317. [PubMed: 15060526]
120. Nero TL, Morton CJ, Holien JK, Wielens J, Parker MW. Oncogenic protein interfaces: small molecules, big challenges. *Nat Rev Cancer*. 2014; 14:248–262. [PubMed: 24622521]
121. Jones S, Thornton JM. Principles of protein-protein interactions. *Proc Natl Acad Sci USA*. 1996; 93:13–20. [PubMed: 8552589]
122. Koes D, Khoury K, Huang Y, Wang W, Bista M, Popowicz GM, Wolf S, Holak TA, Dömling A, Camacho CJ. Enabling large-scale design, synthesis and validation of small molecule protein-protein antagonists. *PLoS ONE*. 2012; 7:e32839. [PubMed: 22427896]
123. Rajamani D, Thiel S, Vajda S, Camacho CJ. Anchor residues in protein-protein interactions. *Proc Natl Acad Sci USA*. 2004; 101:11287–11292. [PubMed: 15269345]
124. Maga G, Hubscher U. Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci*. 2003; 116(Pt 15):3051–3060. [PubMed: 12829735]
125. Moldovan GL, Pfander B, Jentsch S. PCNA, the maestro of the replication fork. *Cell*. 2007; 129:665–679. [PubMed: 17512402]
126. Pedley AM, Lill MA, Davisson VJ. Flexibility of PCNA-protein interface accommodates differential binding partners. *PLoS ONE*. 2014; 9:e102481. [PubMed: 25036435]
127. Punchihewa C, Inoue A, Hishiki A, Fujikawa Y, Connelly M, Evison B, Shao Y, Heath R, Kuraoka I, Rodrigues P, Hashimoto H, Kawanishi M, Sato M, Yagi T, Fujii N. Identification of small molecule proliferating cell nuclear antigen (PCNA) inhibitor that disrupts interactions with PIP-box proteins and inhibits DNA replication. *J Biol Chem*. 2012; 287:14289–14300. [PubMed: 22383522]
128. Gulbis JM, Kelman Z, Hurwitz J, O'Donnell M, Kuriyan J. Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell*. 1996; 87:297–306. [PubMed: 8861913]
129. Buck LB. Unraveling the sense of smell (Nobel Lecture). *Angew Chem Int Ed*. 2005; 44:6128–6140.
130. Sanz G, Schlegel C, Pernollet JC, Briand L. Comparison of odorant specificity of two human olfactory receptors from different phylogenetic classes and evidence for antagonism. *Chem Senses*. 2005; 30:69–80. [PubMed: 15647465]
131. DeMaria S, Ngai J. The cell biology of smell. *J Cell Biol*. 2010; 191:443–452. [PubMed: 21041441]
132. Figueroa XA, Cooksey GA, Votaw SV, Horowitz LF, Folch A. Large-scale investigation of the olfactory receptor space using a microfluidic microwell array. *Lab Chip*. 2010; 10:1120–1127. [PubMed: 20390129]
133. Hsieh JH, Yin S, Wang XS, Liu S, Dokholyan NV, Tropsha A. Cheminformatics meets molecular mechanics: a combined application of knowledge-based pose scoring and physical force field-

based hit scoring functions improves the accuracy of structure-based virtual screening. *J Chem Inf Model.* 2012; 52:16–28. [PubMed: 22017385]

134. Nichols SE, Baron R, McCammon JA. On the use of molecular dynamics receptor conformations for virtual screening. *Methods Mol Biol.* 2012;819, 93–103.
135. Wereszczynski J, McCammon JA. Accelerated molecular dynamics in computational drug design. *Methods Mol Biol.* 2012; 819:515–524. [PubMed: 22183555]
136. Baron R, McCammon JA. Molecular recognition and ligand association. *Annu Rev Phys Chem.* 2013; 64:151–175. [PubMed: 23473376]
137. Lindert S, Meiler J, McCammon JA. Iterative molecular dynamics-rosetta protein structure refinement protocol to improve model quality. *J Chem Theory Comput.* 2013; 9:3843–3847. [PubMed: 23956701]
138. Lindert S, Maslennikov I, Chiu EJC, Pierce LC, McCammon JA, Choe S. Drug screening strategy for human membrane proteins: from NMR protein backbone structure to in silica- and NMR-screened hits. *Biochem Biophys Res Commun.* 2014; 445:724–733. [PubMed: 24525125]
139. Sinko W, Lindert S, McCammon JA. Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design. *Chem Biol Drug Des.* 2013; 81:41–49. [PubMed: 23253130]
140. Ivetac A, McCammon JA. Molecular recognition in the case of flexible targets. *Curr Pharm Des.* 2011; 17:1663–1671. [PubMed: 21619526]
141. Bohnuud T, Kozakov D, Vajda S. Evidence of conformational selection driving the formation of ligand binding sites in protein-protein interfaces. *PLoS Comput Biol.* 2014; 10:e1003872. [PubMed: 25275445]
142. Basom EJ, Spearman JW, Thielges MC. Conformational landscape and the selectivity of cytochrome 450cam. *J Phys Chem B.* 2015; 119:6620–6627. [PubMed: 25955684]
143. Johnson DK, Karanicolas J. Selectivity by small-molecule inhibitors of protein interactions can be driven by protein surface fluctuations. *PLoS Comput Biol.* 2015; 11:e1004081.
144. Mahasenani KV, Li C. Novel inhibitor discovery through virtual screening against multiple protein conformations generated via ligand-directed modeling: a maternal embryonic leucine zipper kinase example. *J Chem Inf Model.* 2012; 52:1345–1355. [PubMed: 22540736]
145. Nair PC, Malde AK, Drinkwater N, Mark AE. Missing fragments: detecting cooperative binding in fragment-based drug design. *ACS Med Chem Lett.* 2012; 3:322–326. [PubMed: 24900472]
146. Kuhn B, Fuchs JE, Reutlinger M, Stahl M, Taylor NR. Rationalizing tight ligand binding through cooperative interaction networks. *J Chem Inf Model.* 2011; 51:3180–3198. [PubMed: 22087588]
147. Li H, Li C. Multiple ligand simultaneous docking: orchestrated dancing of ligands in binding sites of protein. *J Comput Chem.* 2010; 31:2014–2022. [PubMed: 20166125]
148. Vass M, Tarcsay A, Keseru GM. Multiple ligand docking by Glide: implications for virtual second-site screening. *J Comput Aided Mol Des.* 2012; 26:821–834. [PubMed: 22639078]
149. Yu W, Xiao H, Lin J, Li C. Discovery of novel STAT3 small molecule inhibitors via in silico site-directed fragment-based drug design. *J Med Chem.* 2013; 56:4402–4412. [PubMed: 23651330]
150. Li H, Xiao H, Lin L, Jou D, Kumari V, Lin V, Li J, Li C. Drug design targeting protein-protein interactions (PPIs) using multiple ligand simultaneous docking (MLSD) and drug repositioning: discovery of raloxifene and bazedoxifene as novel inhibitors of IL-6/GP130 interface. *J Med Chem.* 2014; 57:632–641. [PubMed: 24456369]
151. Li H, Liu A, Zhao Z, Xu Y, Lin J, Jou D, Li C. Fragment-based drug design and drug repositioning using multiple ligand simultaneous docking (MLSD): identifying celecoxib and template compounds as novel inhibitors of signal transducer and activator of transcription 3 (STAT3). *J Med Chem.* 2011; 54:5592–5596. [PubMed: 21678971]
152. Reed S, Li H, Li C, Lin J. Celecoxib inhibits STAT3 phosphorylation and suppresses cell migration and colony forming ability in rhabdomyosarcoma cells. *Biochem Biophys Res Commun.* 2011; 407:450–455. [PubMed: 21397587]
153. Johnson CN, Adelinet C, Berdini V, Beke L, Bonnet P, Brehmer D, Calo F. Structure-based design of type II inhibitors applied to maternal embryonic leucine zipper kinase. *ACS Med Chem Lett.* 2014; 6:31–36. [PubMed: 25589926]

Abbreviations

FBDD	fragment-based drug design
GPCR	G-protein-coupled receptor
PCNA	proliferating cell nuclear antigen
PDB	RCSB protein data bank
PIP	PCNA-interacting protein sequence motif
RMSD	root-mean-square deviation

Figure 1.

The conformation of the adenine portion of ACP is conserved between two structurally and sequentially diverse proteins, demonstrating a shared subpocket at the binding site (HSP90 N-terminal domain, yellow, PDB: 3t10; chemotaxis protein CheA, blue, PDB: 1i5a).

Figure 2.

Conserved subpockets in protein kinases contribute to their inherent promiscuity. (A) Staurosporine (shown as partially transparent) is known to have inhibitory activity against a plethora of protein kinases from different families (107). Conserved subpockets between the TAO2 MAP3-level kinase (green loop) and serine-threonine kinase 16 (STK16) (blue loop) allow staurosporine to bind with comparable affinities to both proteins (TAO2, PDB ID: 2GCD; STK16, PDB ID: 2BUJ). (B) The promiscuous RET tyrosine kinase (orange surface) contains a binding site subpocket that allows for the binding of overall structurally distinct inhibitors. The red highlighted portion of each inhibitor overlaps in the subpocket, shown with the red circle on the right (green, PDB ID: 2IVU; blue, PDB ID: 2IVV). (C) Within the binding pocket of JNK3 MAP kinase, residues Glu147, Met149, and Val196 form a microenvironment that binds comparable chemical features between a natural ligand, adenosine monophosphate (AMP) (PDB ID: 4KKE), and inhibitors such as the dihydroanthrapyrazole-based antagonist shown on the right (PDB ID: 1PMV).

Figure 3.

Protein interface hotspots contain inducible microenvironments that bind conserved fragments between molecule types. (A) Three residues of p53 (F19, W23 and L25; green sticks) become buried in the surface of MDM2 (blue surface), inducing the formation of hydrophobic subcavities. Residue W23, in particular, likely acts as an anchoring residue, substantially contributing to the binding affinity (PDB ID: 1YCR). (B) Four key residue positions (green sticks) in the highly conserved PIP box sequence are essential for the binding of PIP box-containing proteins (p21 shown as green loop) to PCNA (orange surface). The fifth, seventh, and eighth residues in the sequence bind at a surface pocket on PCNA made up of several hydrophobic microenvironments (PDB ID: 1AXC). (C) Within the hydrophobic PIP box-binding site on PCNA (orange surface), the subpocket defined by the relative orientation of I128, Y133, Y250, P234, and V236 to one another has affinity for an aromatic ring moiety, with a tyrosine residue (green sticks) of p21's PIP box sequence (green loop) anchored in the same location as the tyrosine-analogous fragment of the small-molecule inhibitor T3 (127) (PDB ID: 3VKX, p21 peptide from 1AXC). (D) Some receptors exist that themselves accept a variety of structurally diverse substrates, the best example being the large family of GPCR olfactory receptor proteins. OR1G1, a member of this family, becomes activated upon exposure to numerous diversified odorants (130).

Table 1
Preferential amino acid fragments

Residue	Examples of fragments
Arginine
Aspartic acid
Cysteine
Glutamic acid
Histidine
Methionine
Phenylalanine
Threonine
Tryptophan
Tyrosine

Examples of chemical fragments that select amino acids interact with (31,32).

Table 2
Methods or tools for analysis of binding site similarity and pharmaceutical discovery

Method or database	Publication or software URL	Description
CatSld (57,58)	http://catsid.llnl.gov	Searches for matches between catalytic sites and proteins
CavBase (Relibase) (59–63)	http://relibase.ccdc.cam.ac.uk	Uses 3D property descriptors to detect similarities between protein cavities that share little or no sequence homology
COFACTOR (64)	http://zhanglab.ccmb.med.umich.edu/COFACTOR	Structure-based method for biological protein function annotation
CrystalDock (65)	http://nbcrc.ucsd.edu/data/sw/hosted/crystaldock/	Finds suitable fragments that match protein pocket-lining residues
DoGSiteScorer (66–68)	http://dogsite.zbh.uni-hamburg.de	Detects protein subpockets and predicts site druggability
F-SPE-FP-PH3 (69)	http://www.ncbi.nlm.nih.gov/pubmed/20886466	3D fingerprint descriptors for environments surrounding fluorinated ligands in the PDB
FINDSITE ^{LHM} (70)	http://cssb.biology.gatech.edu/findsitelhm	Homology modeling approach to flexible ligand docking
FragFEATURE (71)	http://www.ncbi.nlm.nih.gov/pubmed/24762971	Predicts small fragments preferred by a target protein structure
GIRAF (72,73)	http://pdj.org/giraf/source/distr.tar.gz	Searches and flexibly aligns protein ligand-binding interfaces
G-LoSA (74)	http://pubs.acs.org/doi/abs/10.1021/ci300178e	Uses known protein-/ligand-binding site data to predict/design a ligand for the target protein
HOMOLOBIND (75)	http://piibase.janelia.org/homolobind/	Identifies residues that are similar to structurally characterized binding sites
IMAAAGINE (76)	http://mfrlab.org/grafss/imaaagine	Searches 3D arrangements of amino acids in PDB structures
KRIPO (48)	http://www.ncbi.nlm.nih.gov/pubmed/22830492	Quantifies subpocket similarities to identify bioisosteres
Med-SuMo ^a	http://medit-pharma.com/index.php?page=med-sumo	Locates similar regions on protein surfaces that are linked to certain chemical function
Nucleos (77)	http://nucleos.bio.uniroma2.it/nucleos	Identifies nucleotide-binding sites based on nucleotide modularity
Patch-Surfer (78)	http://kiharalab.org/patchsurfer2.0/	Web tool for predicting binding ligands for a protein
Phosfinder (79)	http://phosfinder.bio.uniroma2.it	Web server for the prediction of phosphate-binding sites
PocketFEATURE (80)	https://simtk.org/home/pocketfeature	Uses protein microenvironments to identify similar binding sites; method is able to recognize distant protein relationships
PocketAlign (81)	http://proline.physics.iisc.ernet.in/pocketalign/	Generates structural superpositions of binding sites
PoSSuM (82)	http://possum.cbrc.jp/PoSSuM	Enables rapid exploration of similar binding sites between structurally diverse proteins
PrISE (55)	http://prise.cs.iastate.edu/index.py	Predicts interface residues using local surface structural similarity
ProBiS (83)	http://probis.cmm.ki.si	Enables structural comparison and local alignment of proteins
PROLIX (84)	http://www.ncbi.nlm.nih.gov/pubmed/22582806	Mines protein–ligand interactions in large structure databases

RCSB PDB (22)	http://www.rcsb.org/pdb/home/home.do	Worldwide repository of 3D structural information of large biological molecules
SA-Mot (85,86)	http://sa-mot.mti.univ-paris-diderot.fr	Extracts structural motifs of interest from protein loop structures
SCOWLP (87)	http://www.scowlp.org	Enables individual and comparative analysis of protein interactions
SiteBinder (88)	http://ncbr.muni.cz/SiteBinder	Allows the superimposition of large sets of protein structural motifs
SiteComp (89)	http://sitecomp.sanchezlab.org	Compares binding sites, evaluates residue contribution to binding, and identifies subsites with distinct molecular interaction properties
SubCav (90)	http://www.ncbi.nlm.nih.gov/pubmed/23327721	Similarity searching and alignment of protein subpockets
TargetHunter (91)	http://www.cbligand.org/TargetHunter	Web-based target prediction tool
TRAPP (92)	http://www.mcm.h-its.org/trapp/	Analyzes binding pocket variation along protein motion trajectory and assesses the druggability of a target
TrixP (93)	http://www.ncbi.nlm.nih.gov/pubmed/23390978	Fast protein-binding site comparison and function prediction
Wallach & Lilien algorithm (94)	http://bioinformatics.oxfordjournals.org/content/25/12/i296.full	Uses pharmacophoric features to predict subcavity binding preferences

A selected set of programs, databases or general methods for assessing protein similarity and analyzing binding sites for the purposes of predicting new fragments for drug discovery/lead development.