Published in final edited form as:

Trends Pharmacol Sci. 2015 November ; 36(11): 724–736. doi:10.1016/j.tips.2015.08.003.

Lessons from hot spot analysis for fragment-based drug discovery

David R. Hall¹, Dima Kozakov², Adrian Whitty³, and Sandor Vajda^{2,3}

¹Acpharis Inc, Boston University, Boston, MA, 02215, USA

²Department of Biomedical Engineering, Boston University, Boston, MA, 02215, USA

³Department of Chemistry, Boston University, Boston, MA, 02215, USA

Abstract

Analysis of binding energy hot spots at protein surfaces can provide crucial insights into the prospects for successful application of fragment-based drug discovery (FBDD), and whether a fragment hit can be advanced into a high affinity, druglike ligand. The key factor is the strength of the top ranking hot spot, and how well a given fragment complements it. We show that published data are sufficient to provide a sophisticated and quantitative understanding of how hot spots derive from protein three-dimensional structure, and how their strength, number and spatial arrangement govern the potential for a surface site to bind to fragment-sized and larger ligands. This improved understanding provides important guidance for the effective application of FBDD in drug discovery.

Fragment-based drug discovery

In fragment-based drug discovery (FBDD), a target of interest is screened against a library of low molecular weight, weak-binding compounds called “fragments,” with the best fragment hits subsequently being elaborated into larger, higher affinity ligands [1–3]. Fragment libraries generally comprise compounds of molecular weight (MW) 150–250 Da, substantially smaller than compounds used in traditional high-throughput screening (HTS). Since the number of theoretical compounds increases exponentially with MW, screening smaller compounds enables a more efficient exploration of chemical space [4]. Thus, compared to conventional HTS, fragment screens require testing of many fewer compounds to establish the binding potential of the target and to identify initial hits. In addition, it is easier to find a small molecule that complements a particular subsite within a binding site than a larger molecule that is complementary to the entire site; thus, FBDD usually yields higher hit rates than HTS. The key trade-off inherent in fragment screening is that, due to

Correspondence to: Dima Kozakov; Adrian Whitty.

Conflict of Interest

D.R.H. is a full-time employee of Acpharis, Inc. The company offers software similar to the FTMap program that was used in this paper. D.K. and S.V. own Acpharis stock. However, the FTMap software and server (<http://ftmap.bu.edu/>) are free for use.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

the small size of the compounds, even a fragment that is optimally complementary will interact with the target protein over a limited contact area, and so fragment hits are generally weak, with binding affinities of 1 mM being typical. A variety of approaches can be used to detect the binding of these weak ligands, including protein-ligand NMR [5, 6] and X-ray crystallography [3, 7, 8]. A number of companies have acquired substantial experience with these approaches, and have developed efficient technologies for screening libraries of fragment size compounds. Over the size range of typical fragments and drugs the interaction energy of appropriately complementary compounds with a protein target grows approximately linearly with compound MW [9]. Affinities of different-sized fragments and related compounds are therefore usually compared in terms of their ligand efficiency (LE), defined as the binding free energy per heavy (i.e. nonhydrogen) atom in the ligand [10]. The goal of an initial fragment screen is thus to identify compounds with high LE values, generally 0.3 kcal/mol per heavy atom. The next step in FBDD is to evolve such core fragments into larger, higher affinity leads, which generally involves extension of the fragment into neighboring pockets, as well as sometimes requiring optimization of the initial fragment hits to achieve better complementarity to its binding subsite [2, 11, 12].

In this review we show that knowledge of binding energy hot spots, defined as regions of the binding site that are major contributors to the binding free energy [13, 14], provides useful information for both stages in FBDD. As will be discussed, if the structure of a target protein is known, the hot spots can be identified by experiments or, very easily, by computation. In particular, we show that hot spots help to answer a number of important questions. First, is a particular target suitable for FBDD? Second, is a given fragment library suitable for the target? Third, if a fragment hit is identified, is its binding mode sufficiently robust for the hit to be used as the core of a higher affinity ligand? Fourth, if several fragment hits are found, which can be the better starting point? Fifth, what is the chance that a given fragment hit can be extended into a potent, rule-of-five compliant ligand? To answer these questions we will discuss how hot spots can be identified, and show how the hot spot structure of a protein affects the binding of fragments and of larger ligands.

Binding energy hot spots

The concept of binding energy hot spots was originally introduced in the context of mutating interface residues to alanine at protein-protein or protein-peptide interfaces [15–18]. Using this method, a residue is considered to contribute to a hot spot if its mutation to alanine gives rise to a substantial drop in binding affinity. Early on it was proposed that hot spots identified by alanine scanning might correspond to regions with the potential to also interact strongly with small molecule ligands [15], and much subsequent work has established that this is indeed the case [13, 14, 19–26]. Of particular relevance to FBDD, it has been shown that individual hot spots are characterized by their ability to bind a variety of fragment-sized or even smaller organic molecules [5, 21, 27, 28]. In addition to its use in finding ligands, therefore, fragment screening also provides a direct method for identifying hot spots, when the screen is done using a structural technique such as X-ray crystallography [22, 27, 28] or nuclear magnetic resonance (NMR) [21] that can identify the locations on the protein surface where the probes tend to bind (Box 1). Determining binding hot spots by crystallographic or NMR-based fragment screening requires specialized expertise and

infrastructure, however, and its applicability can be limited by physical constraints of the protein of interest. Hence, a number of computational approaches have been developed to predict fragment binding and identify hot spots. These include the classical methods GRID [29] and Multiple Copy Simultaneous Search (MCSS) [30] (Box 1). In this paper we use the more recent FTMap algorithm [23, 31], which will be described in more detail below. All these methods are computational analogs of the X-ray and NMR based screening experiments, and use individual functional groups or small organic molecules to probe the protein surface and identify regions that are capable of binding multiple probe types.

Box 1

Methods for determining binding energy hot spots

In the multiple solvent crystal structures (MSCS) method, X-ray crystallography is used to determine the structure of the target protein soaked in aqueous solutions of 6–8 organic solvents used as probes. By superimposing the structures, regions that bind multiple different probes can be detected [27, 28]. While individual probes may bind at a number of locations, their clusters indicate binding hot spots. Similarly, in the structure-activity relationship (SAR) by NMR method, proteins are immersed in a series of small organic molecules, and perturbations in residue chemical shifts are used to identify residues that participate in small molecule binding [21]. It was shown that the small “probe” ligands cluster at hot spots, and that the fragment hit rate predicts the importance of the site, as determined by whether the site was subsequently shown to be able to support the binding of high affinity ligand [21, 27]. Computational methods for identifying hot spots include the classical methods GRID [29] and Multiple Copy Simultaneous Search (MCSS) [30], as well as the more recent FTMap algorithm [23, 31], which is described in detail in the main text. All these methods are computational analogs of the X-ray and NMR based screening experiments, and use individual functional groups or small organic molecules to probe the protein surface and identify regions that are capable of binding multiple probe types. More recently, mixed molecular dynamics (MD) has emerged as a promising approach to protein mapping. The method performs MD simulations of the target in an aqueous solution of probe molecules [86–89]. While the use of explicit water molecules potentially improves the accuracy of simulations, it also reduces the diffusivity of the probes, and hence it is frequently questionable whether equilibrium distribution can be achieved on reasonable time scales [90]. In addition, due to the need for long simulations, current mixed MD methods explore binding properties using only a few probe types, which is likely to limit the reliability of hot spot prediction. However, with the increase of computing power, mixed MD may become an important tool for determining hot spots of proteins.

The main shortcoming of GRID [29] and MCSS [30] is that the very small probes bind in many different pockets, resulting in a large number of false positive local minima [27, 32]. Using better solvation models, better sampling of the protein surface, and clustering the probe positions, the FTMap algorithm has overcome this problem, and has been shown to provide robust prediction of binding hot spots, with excellent agreement with experimental data [23, 31]. The standard probe set of FTMap includes 16 small organic molecules [31].

For each probe type, the interaction energy between the probe and protein is calculated on a dense rotational/translational grid around the protein, a number of low energy positions are retained and are further refined by using a higher resolution energy function, accounting for probe and protein flexibility. The resulting structures are clustered, and the clusters are ranked by average energy. The results obtained using each of the 16 different probes are then overlaid. At a few locations of the protein, multiple highly ranked probe clusters can be seen to overlap, forming consensus clusters (CCs) that identify the binding hot spots [23, 25, 31]. The consensus cluster containing the largest number of probe clusters, denoted as CC1, is considered the main or primary hot spot; all other consensus clusters are secondary hot spots. One advantage of using hot spots to evaluate the binding potential of a protein is that hot spot locations are less sensitive to conformational change than is the shape of the binding site itself, enabling the locations and energetic importance of binding sites to be identified by analysis of almost any available structure of a protein interest, including ones without a bound ligand [26, 33–35]. An additional advantage of FTMap is that the algorithm has been implemented in an easy to use server (<http://ftmap.bu.edu>) [31].

Hot spots and FBDD

There is a close relationship between FBDD and the hot spot, as the latter are essentially defined as sites that are capable of binding fragment-sized or even smaller probe molecules. This has two fairly straightforward consequences. First, if a protein does not have any strong hot spot, it will not produce a significant hit rate in an experimental fragment screen. It has been shown that fragment hit rate is reliable indicator of whether a protein is capable of binding a druglike ligand with high affinity [21]. The relationship between hot spots and druggability has since been explored by several computational methods. [23, 25, 36–38]. In particular, we have previously established that, when using 16 probes types for the mapping, the sites that are known to be druggable invariably contain a strong hot spot containing 16 or more probe clusters [24, 39]. Second, since the primary hot spot is the location with the highest binding potency for small molecules, we expect that fragment hits are most likely to bind at this site. As shown earlier [40] and demonstrated by further examples in the current paper, for the majority of proteins this is really the case.

The second stage of FBDD, in which an initial fragment hit is elaborated through medicinal chemistry into a larger and stronger-binding lead compound, relies on the premise that the interactions between a fragment hit and the protein are sufficiently strong and specific that the fragment binding mode will be robust and thus will be conserved as the fragment is grown. Put another way, it is assumed that a fragment will bind in the same orientation whether on its own or as a moiety of a larger ligand [41–43]. Successful case stories of FBDD, such as those reviewed below, demonstrate that this assumption often applies. However, studies in which a larger ligand has been deconstructed into its component fragments, and their binding characterized experimentally, indicate that conservation of binding mode is not always seen. In some such studies the fragment positions remain invariant when the rest of the ligand is removed. Examples include the conservation of dimethylguanylylurea in a series of dimethylguanylylurea-containing peptides from decomposition of the natural product argifin [44], deconstruction of inhibitors of the SH2 domain of $^{60}\text{S}^{\text{pp}}$ [45], and of inhibitors of the interaction between the von Hippel–Lindau

(VHL) complex and the hypoxia-inducible factor 1 α (HIF-1 α) [46]. However, other studies show that the binding mode of a component fragment is not necessarily conserved upon the deconstruction of a larger ligand. For example, Krimm and co-workers have shown that fragments resulting from the dissection of Bcl-xL inhibitors all shift to bind at a single site, irrespective of their bound positions in the full-sized inhibitors [47]. Similarly, Babaoglu and Shoichet considered three fragments of a β -lactamase inhibitor, and showed that all three changed location compared to their binding positions in the original inhibitor [48]. More recently, Shoichet and co-workers [49] deconstructed substrates of six enzymes from three different superfamilies into 41 overlapping fragments, and showed that even fragments containing the key reactive group had little activity, and that most fragments did not bind measurably until they captured most of the substrate features. Furthermore, deconstruction of some well-optimized ligands into fragments reveals that no individual fragment binds with any observable affinity. For example, small fragments of an MDM2/p53 inhibitor Nutlin do not bind to any detectable degree, and the smallest one that does has a molecular weight over 300 Da, exceeding the usual size-range for fragment libraries [50].

Recent work has shown that whether a target will be druggable (i.e. whether it will bind fragments at all), depends to a significant degree on the strength of the primary hot spot, CC1. Targets that lack a primary hot spot of sufficient strength are not suitable for FBDD, and have poor overall prospects for inhibitor development by any means [21, 27]. The FTMap analysis used to derive these results was based on mapping the ligand-free proteins, allowing druggability to be assessed for any target that has been crystallized even before investing time or resources into lead discovery [39]. Whether a bound fragment will have a robust binding mode also depends largely on CC1. Specifically, a fragment hit will retain its binding mode if it binds at CC1, provided it overlaps well enough with CC1 to efficiently exploit the potential of this site to generate noncovalent interaction energy [51]. In contrast, if a fragment binds at a site other than CC1, there is no guarantee that it will remain at that position upon further elaboration. To better define this condition, we formulated a quantitative measure of the degree of spatial overlap of a bound fragment with CC1, as described in Box 2. This condition implies that FBDD may be more difficult even for a druggable target if its main hot spot is too large and diffuse to fill with a fragment size compound. A large hot spot of this kind will facilitate the binding of various fragments, resulting in high hit rate, but the binding mode is unlikely to be conserved because any fragment occupies only a fraction of the major hot spot and so alternative binding modes within CC1 may be accessible. More generally, it appears that a sufficient number of interactions across a compound are needed to fix the binding pose. If such interactions are not available for a fragment, altering its structure may change its binding pose, which thus provides uncertain information for the design of larger ligands. While this does not imply that the target is unsuitable for FBDD, the second step of the method can be more difficult. To support this hypothesis, we refer to a study by the Shoichet group, who screened a fragment library for binding to AmpC β -lactamase, which has a large and diffuse primary hot spot [52]. Although they reported high hit rate, the bound fragment positions showed substantial variation, providing limited information for the design of higher affinity inhibitors and, in fact, no such inhibitors were found. The same condition can also be used to assess whether a fragment library is likely to be suitable for a particular target. Commercial

fragment libraries generally contain flat heterocyclic compounds with MW = 100–250 Da, and this size range may be too restrictive for a protein that has a large primary hot spot. Again, we emphasize that screening with the small planar fragments is likely to result in high hit rate, but expanding these hits into higher affinity leads may be more difficult, and hence using a library with fragments that are comparable in size and shape to the primary hot spot may be a better choice. However, since larger fragments explore smaller fraction of the chemical space [53] and generally reduce the hit rate, selecting the most efficient fragment library always requires a compromise.

Box 2

Conditions for conservation of fragment binding mode

The primary hot spot CC1, determined the the consensus cluster (CC) that contains the highest number of probe clusters, provides a large fraction of free energy toward binding any ligand at the particular site. In agreement with the key role of this hot spot, it was shown that a fragment hit retains its binding mode if it overlaps with a large enough fraction of CC1, and thus can efficiently exploit the potential interaction energy that this site provides [51]. If the fragment binds to a location outside CC1, or within CC1 but without achieving sufficiently high overlap, there is no guarantee that it will remain at that position upon expansion of the fragment into a larger ligand. Similarly, a moiety of a larger ligand must satisfy this condition of overlap with CC1 to ensure that it will retain its position when tested as a separate molecular fragment. To make the condition more quantitative, fractional overlap, denoted as *FO*, was proposed as a quantitative measure of the degree of spatial overlap between a fragment and CC1. *FO* is defined by $FO = N_F/N_T$, where N_T denotes the total number of non-hydrogen atoms of all probe molecules in CC1, and N_F is the number of such atoms that are within 2 Å from any non-hydrogen atom of the fragment in its bound position. Thus, *FO* measures the fraction of CC1 occupied by the fragment, weighted according to the energetic importance of each region within the hot spot as measured by the local density of probe atoms [51]. It was shown that if a fragment binds at the primary hot spot, has fractional overlap over 0.8, and has acceptable ligand efficiency (LE >0.3) [40], then its binding mode is very likely conserved upon extension. If a number of fragment hits are found, the one with higher fractional overlap is likely to be the better candidate for expansion, unless there is substantial difference in LE values.

In discussing the use of hot spot information in FBDD, we have so far focused on CC1, which is clearly important for determining both the druggability of the target and the stability of the binding mode of the core fragment, and thus for deciding whether it can be the starting point for expansion. The importance of secondary hot spots is that they provide information for fragment expansion. As shown below, CC1 together with the secondary hot spots in its vicinity serve as anchor points for the binding of any ligand. While ligands can extend to regions of the protein that do not have a hot spot, the gain in terms of binding affinity is small, and hence FBDD should utilize the available hot spots as much as possible. It is also clear that interactions with regions between hot spots provide limited benefit, and hence hot spots close to CC1 are generally more important for extending the core. Based on

the analysis of hot spots for a number of drug targets we observed that to bind a rule-of-five compliant compound with high affinity, at least one secondary hot spot should be closer than 8 Å to CC1. Bridging a larger distance will require compounds of molecular weights in excess of 500 Da, typically achievable only using compounds that are not conventionally druglike, such as macrocycles or peptidomimetic foldamers. Mapping may show need for such larger ligands even for targets that have secondary hot spots within 8 Å of CC1 if these are all relatively weak, and hence achieving acceptable binding affinity requires ligands that also reach hot spots located farther from CC1. This occurs in a substantial number of protein-protein interaction targets that frequently have three or four relatively weak hot spots at the interface [24].

As already mentioned, hot spots are generally less sensitive to conformational changes than binding sites are, and hence the druggability of most targets can be assessed by mapping ligand-free structures [39]. However, ligand binding may affect the location and relative strength of hot spots [34, 39, 54], and hence it is advisable to also map ligand-bound structures when they become available in the process of drug discovery. For example, the ligand-free structure of Bcl-xL has three strong hot spots close to each other in the pocket that binds L578 of the BAK peptide [47], and it was shown that fragments of known inhibitors all bind at this single site. However, the hot spots move much farther apart upon binding the BAK peptide or the inhibitor ABT737 [55], requiring the design of inhibitors with MW exceeding 600 g/mol [24, 56, 57].

Case studies

We demonstrate the usefulness of hot spot information by analyzing the application of FBDD to six targets recently reviewed by Murray *et al.* [41]. The novelty is that our discussion is based on the analysis of hot spots. For each protein we downloaded a structure (ligand-free, where available), listed in Table 1, from the Protein Data Bank (PDB) [58], and determined the binding hot spots using FTMap. As mentioned, each hot spot is identified by a consensus cluster (CC) of probe clusters [31]. In the following figures, each probe cluster in a CC is represented by a single structure to avoid overcrowding. All representative probes in a CC have the same color. All ligands and crystallographic water molecules were removed from the structures prior to mapping, and the ligands shown in the figures were superimposed on the mapping results for reference only. A detailed examination of interactions between the protein and a fragment hit can provide important information on the chemotypes that are recommended to use [41]. However, here we focus exclusively on what can be learned from analysis of hot spots, and so do not provide a detailed analysis of protein-fragment interactions. As will be shown, in the first four examples the initial fragment hit overlaps with the primary hot spot very well (i.e., the fractional overlap $FO > 0.8$), and the fragment is completely conserved in the lead compounds, at least in the early stages of optimization. In contrast, $FO \approx 0.8$ for the fragment hits obtained for the last two targets (BACE-1 and CDK2), and the fragment structures were used only for determining key interactions rather than fully retained within the leads. For these targets the binding modes became conserved only when the compounds were expanded to reach several hot spots.

Heat shock protein 90 (HSP90)

Heat shock protein 90 (HSP90) has been targeted using FBDD by several companies [41–43, 59, 60]. Astex screened 1600 fragments using NMR ligand-observed gradient spectroscopy, and 125 were further subjected to X-ray crystallography [43]. Several different chemotypes were identified, including 2-methyl-4-diethylamide-phenol (Figure 1a). The fragment overlaps very well with CC1 ($FO=0.91$; see Box 2 for definition), and also extends into CC2 and CC5 that overlap with CC1, predicting good conservation of the fragment binding mode. Indeed, replacing of OMe by a hydrophobic moiety, incorporating a second hydroxy group onto the phenol, and adding a phenyl ring at the diethylamide, did not cause any change in the binding mode of the core (Figure 1b). The additional moieties further improved the overlap with CC1 to $FO=0.98$, and dramatically improved the binding affinity, improving K_d from 790,000 nM to 0.54 nM. Further addition of a methylpiperazine group improved physicochemical and PK properties, but the added moiety did not interact with any strong hot spot and hence did not improve the affinity ($K_d = 0.71$ nM) [61], and the binding mode of the core fragment remained invariant (Figure 1c). Figure 1d shows a novel benzolactam inhibitor that was designed to optimize selectivity versus the HSP90 endoplasmic reticulum and mitochondrial isoforms (GRP94 and TRAP1, respectively) [62]. Although this inhibitor has a different chemotype, the position of the phenyl ring in CC1 remained essentially invariant. The benzolactam inhibitor reaches CC3 but overlaps less well with CC5 than the other three inhibitors discussed above. Since CC3 is stronger than CC5, we might expect that the affinity will further improve. However, the K_d value is only 5 nM, perhaps because trying to reach a hot spot far from CC1 requires bridging regions of the surface that have no hot spots, and hence do not contribute much to the binding free energy.

B-Raf

B-Raf is the target of the FDA approved drug vemurafenib (Zelboraf®, PLX4032) [63–65]. Figure 1e shows N-phenyl-1H-pyrrolo[2,3-b]pyridin-3-amine that was found from a biochemical screen of a kinase focused library [65]. Although this is a fragment sized compound and was used for the design of larger inhibitors, the Plexxikon approach is more chemogenomic in nature - i.e. they had developed target focused libraries of skeletons (MW 150–350) rather than a true fragment screen. From the initial screen, 238 compounds were selected for co-crystal analysis: the compounds were mixed with recombinant kinase domains from various kinases that the compounds bound in buffers, resulting in 100 structures. In particular, the N-phenyl-1H-pyrrolo[2,3-b]pyridin-3-amine ($IC_{50} = 100,000$ nM) was co-crystallized with the serine/threonine kinase Pim-1. However, the compounds in Figures 1f through 1h are shown bound to BRAF, and the hot spots were also determined by mapping a BRAF structure. As shown in Figure 1e, the 1H-Pyrrolo[2,3-b]pyridine moiety overlaps well with CC1 ($FO = 0.99$), assuring stable binding. Indeed, the position of this moiety is only slightly different in the compound PLX4720 (Figure 1f), although the latter was co-crystallized with BRAF rather than Pim-1. The overlap with CC3 is substantially improved, and PLX4720 reaches into both CC2 and CC4, resulting in much higher affinity ($IC_{50} = 13–160$ nM). The excellent overlap with CC1 and CC3 assures perfect conservation of the core during further optimization resulting in PLX3203 and Vemurafenib (PLX4032), shown in Figures 1g and 1h, respectively.

Protein kinase B and A (PKB/PKA)

Protein kinase B and A (PKB/PKA) inhibitors have also been developed by Astex Therapeutics using FBDD. CC1 of PKB has two well-defined ring positions. Figures 2a and 2b, respectively, show the core fragment and a highly potent derivative [66]. The fragment overlaps fully with CC1 ($FO=1.0$), and binding mode conservation is correspondingly excellent. Note that extension of the fragment into CC3 improves IC_{50} from 80 to 34 nM. Figures 2c and 2d show a pair of PKA inhibitors of different chemotypes, though one ring and part of another in CC1 are well conserved [67, 68]. Both inhibitors extend into CC3, and are potent, with IC_{50} values of 5 nM and 2.1 nM, respectively.

Urokinase-type plasminogen activator (uPA)

Urokinase-type plasminogen activator (uPA) is also a good target for FBDD (Figure 2e, Figure 2f) [69]. Although the overlap is not perfect, according to our measure the fragment in Figure 2e [69] covers a large fraction of the small CC1 ($FO=0.92$), which explains the good conservation of the binding mode as the fragment is extended into CC2 (Figure 2f), which improved IC_{50} from 1 mM to 72 nM [69]. Figures 2g and 2h show inhibitors with different chemotypes. Both inhibitors have moieties in CC1 that have even better overlap with the primary hot spot than the compounds discussed above, also extend into CC2, and have very high affinities, with $K_i = 11$ nM [70] and $IC_{50} = 25$ nM [71], respectively.

β -secretase (BACE-1)

β -secretase (BACE-1), an aspartic protease, plays an important role in Alzheimer's disease, and has been a challenging target [72]. HTS, virtual screening, and rational design of transition state analog inhibitors all led to compounds with ill-defined SAR and non-competitive behavior [72]. Inspection of this enzyme shows a large and complex binding pocket containing a number of sub-pockets, with five partially overlapping hot spots. Nevertheless, fragment screening resulted in some early hits [73], which were followed up by successful FBDD campaigns [74, 75]. For example, Figure 3a shows a thioamidine core fragment identified by Schering-Plough researchers [76]. In contrast to the other targets we have discussed so far, in BACE-1 this fragment does not overlap well with CC1 ($FO=0.6$), and hence conservation of the binding mode is not assured. In fact, this core fragment was used only for identifying interactions with the catalytic aspartate groups of the protease [77], and structure-based design was then used to generate compounds that could exploit similar interactions but in different compound scaffolds, e.g., Figure 3b ($IC_{50} = 605$ nM) [77]. Neither the chemotype nor the binding mode of the initial fragment was retained. However, this inhibitor has better overlap with both CC1 and CC3, which seems to stabilize the binding mode. Indeed, the recently developed inhibitors reported by Lilly [78] (Figure 3c) and by Amgen [79] (Figure 3d) retain the binding mode of the Schering-Plough compound, making similar interactions with CC1, CC3, CC7 (overlapping with CC1), and CC8.

Cyclin dependent kinase 2 (CDK2)

Cyclin dependent kinase 2 (CDK2) is the sixth target discussed by Murray *et al.* [41] as an example of successful FBDD application [80]. Figure 3e shows one of the fragment hits identified by Astex using X-ray crystallographic screening. The fractional overlap with CC1,

located in the binding site, is borderline ($FO=0.8$), which does not necessarily predict the conservation of the binding mode. In fact, the fragment was used only for determining some key interactions, and the system was simplified by removal of the fused benzene ring to afford the pyrazole [80]. Although the resulting inhibitor shown in Figure 3f is weak ($IC_{50} = 850$ nM), it has better overlap with CC1 and CC3, predicting conservation of the binding mode. Indeed, the inhibitor shown in Fig. 3g binds exactly the same way. The newly added difluorophenyl group reaches into CC5, and results in a substantial improvement of binding ($IC_{50} = 3$ nM) [80]. Finally, Figure 3h shows an inhibitor that, while retaining some of the interactions with the hot spots, has a different chemotype, and also has fairly high affinity ($IC_{50} = 17$ nM) [81].

Hot spots and ligand binding mode

To further illustrate the relationships between hot spots and ligand binding, in Table 1 we show FTMap results for 23 proteins. The first six are the FBDD targets [41] already discussed, and the others taken from the papers on druggability by Huang and co-workers [82, 83]. As before, for each protein we downloaded a (preferably ligand-free) structure, also listed in Table 1, from the Protein Data Bank (PDB) [58], and determined the binding hot spots using FTMap. We then extracted all structures, co-crystallized with small ligands, from the PDB. For each target protein, we show the number of hot spots in the functional site and list these hot spots, including their ranks, the number of probe clusters in the consensus cluster that defines each hot spot, and the number of ligands that come within 1 Å of any probe in the particular hot spot.

The first observation from the results in Table 1 is that, in agreement with the finding for the FBDD targets, the largest consensus cluster CC1 is in the functional site for all but one of the proteins. The exception is cyclooxygenase-2 (COX-2), where CC1 is located at the edge of the heme binding site, such that the strongest hot spot in the active site is CC2. The second observation is that CC1 includes at least 16 probe clusters, which was previously defined as a requirement for druggability [23, 24, 39]. This is not unexpected, since these proteins have highly potent ligands. The table contains one exception to the druggability rule, HMG CoA reductase, which has only 14 probe clusters in its strongest hot spot. HMG CoA reductase has potent statin inhibitors [84], and thus is obviously druggable. This contradiction is due to the fact that statins occupy only a fraction of the large binding site. While this is enough to block access of the substrate to the active site, the unoccupied region also includes a number of hot spots, which attracts probes away from the main binding site reducing the number of probe clusters overlapping with the inhibitor [84]. Since these additional hot spots are far from CC1, located in the statin binding pocket, they are not listed in Table 1.

Our third observation is that more than 95% of ligands interact with CC1 in all but one of the proteins. In this case the exception is acetylcholine esterase (AChE) that, in addition to the main acylation (A) site, is known to have a peripheral (P) site [85]. Five of the 45 AChE ligands in the PDB bind at this P site, blocking access to the A site without directly occupying it. Since the existence of two binding sites in AChE is well known, this does not impact the conclusion that CC1 always serves as the main anchor for the majority of ligands.

Fourth, all targets considered here have at least one secondary hot spot that also participates in ligand binding. In proteins with only two hot spots in the ligand binding site, essentially all inhibitors are anchored by these two hot spots (Table 1). Thus, while inhibitors of such targets may have several chemotypes, all must have similar binding modes. For targets with multiple secondary hot spots, the number of potentially different binding modes increases. For example, disregarding the small number of non-typical ligands that do not start from CC1, with two secondary hot spots the number of potential binding modes is three: two spanned by CC1 and one of the secondary hot spots, and one spanned by all three. Of course, it is possible that only the last provides high enough affinity. Similarly, with three secondary hot spots the number of different potential binding modes is seven: three spanned by two hot spots; three spanned by three; and one by all four. Thus, the number and arrangement of hot spots limit the potential binding modes of all ligands.

Concluding remarks

Over the past two decades, experimental fragment-based screening has become widely used for lead identification. We have described how the concept of binding energy hot spots, determined using computational analogs of these methods, can provide crucial insights into how the structure of a particular target protein dictates the likelihood of success in identifying fragment hits, and into the prospects that a particular fragment hit can be advanced into a high affinity, druglike ligand. This analysis additionally resolves longstanding questions concerning when a fragment, obtained by deconstructing a larger ligand, will retain its binding position, and into the nature of ligand binding sites and the assessment of protein druggability in general. The key lesson is that it is the strength of the top-ranked hot spot, CC1, that determines what fragment hit rate can be expected. A minimum strength for CC1 is also required for a site to be druggable with respect to larger ligands, as illustrated here by our analysis of a set of drug targets, which revealed the key role that CC1 plays in ligand binding. Furthermore, the degree of overlap with CC1 determines whether a moiety of a larger ligand will display a robust binding mode when experimentally tested as a separate fragment, and conversely whether a fragment hit has a binding mode that is robust enough to support expansion into a higher affinity ligand. The presence of secondary hot spots within a short distance of CC1 is additionally required to support binding of a full-sized ligand with high affinity for the target. Thus, the locations of these secondary hot spots can help guide the elaboration of a fragment hit, through medicinal chemistry, into a high affinity, druglike ligand, as illustrated by the analysis of several examples where FBDD was successful. Overall, we believe that published data are now sufficient to provide a sophisticated and quantitative understanding of how binding energy hot spots derive from protein three-dimensional structure, and how the strength, number and spatial arrangement of hot spots govern the potential for a surface site to bind to fragment-size and larger ligands. This improved understanding provides important guidance for the effective application of FBDD in drug discovery. Although we believe that determining the binding hot spots using the free FTMap server (<http://ftmap.bu.edu/>) provides useful information, we admit that a number of problems, listed in the Outstanding Questions Box, need further analysis.

Acknowledgments

This work on this review was supported by grants GM064700 to SV and GM094551 to AW and SV from the National Institute of General Medical Sciences.

References

1. Rees DC, et al. Fragment-based lead discovery. *Nat Rev Drug Discov*. 2004; 3:660–672. [PubMed: 15286733]
2. Erlanson DA, et al. Fragment-based drug discovery. *J Med Chem*. 2004; 47:3463–3482. [PubMed: 15214773]
3. Hartshorn MJ, et al. Fragment-based lead discovery using X-ray crystallography. *J Med Chem*. 2005; 48:403–413. [PubMed: 15658854]
4. Leach AR, Hann MM. Molecular complexity and fragment-based drug discovery: Ten years on. *Curr Opin Chem Biol*. 2011; 15:489–496. [PubMed: 21665521]
5. Hajduk PJ, et al. NMR-based screening in drug discovery. *Q Rev Biophys*. 1999; 32:211–240. [PubMed: 11194565]
6. Hajduk PJ, et al. Identification of novel inhibitors of urokinase via NMR-based screening. *J Med Chem*. 2000; 43:3862–3866. [PubMed: 11052791]
7. Blundell TL, et al. High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Disc*. 2002; 1:45–54.
8. Winter A, et al. Biophysical and computational fragment-based approaches to targeting protein-protein interactions: Applications in structure-guided drug discovery. *Q Rev Biophys*. 2012; 45:383–426. [PubMed: 22971516]
9. Hajduk PJ. Fragment-based drug design: How big is too big? *J Med Chem*. 2006; 49:6972–6976. [PubMed: 17125250]
10. Hopkins AL, et al. Ligand efficiency: A useful metric for lead selection. *Drug Discov Today*. 2004; 9:430–431. [PubMed: 15109945]
11. Foloppe N. The benefits of constructing leads from fragment hits. *Future Med Chem*. 2011; 3:1111–1115. [PubMed: 21806375]
12. Congreve M, et al. Recent developments in fragment-based drug discovery. *J Med Chem*. 2008; 51:3661–3680. [PubMed: 18457385]
13. DeLano WL. Unraveling hot spots in binding interfaces: Progress and challenges. *Curr Opin Struct Biol*. 2002; 12:14–20. [PubMed: 11839484]
14. DeLano WL, et al. Convergent solutions to binding at a protein-protein interface. *Science*. 2000; 287:1279–1283. [PubMed: 10678837]
15. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. *Science*. 1995; 267:383–386. [PubMed: 7529940]
16. Keskin O, et al. Hot regions in protein-protein interactions: The organization and contribution of structurally conserved hot spot residues. *J Mol Biol*. 2005; 345:1281–1294. [PubMed: 15644221]
17. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *J Mol Biol*. 1998; 280:1–9. [PubMed: 9653027]
18. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci U S A*. 2002; 99:14116–14121. [PubMed: 12381794]
19. Koes DR, Camacho CJ. Small-molecule inhibitor starting points learned from protein-protein interaction inhibitor structure. *Bioinformatics*. 2012; 28:784–791. [PubMed: 22210869]
20. Zerbe BS, et al. Relationship between hot spot residues and ligand binding hot spots in protein-protein interfaces. *J Chem Inf Model*. 2012; 52:2236–2244. [PubMed: 22770357]
21. Hajduk PJ, et al. Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem*. 2005; 48:2518–2525. [PubMed: 15801841]
22. Ciulli A, et al. Probing hot spots at protein-ligand binding sites: A fragment-based approach using biophysical methods. *J Med Chem*. 2006; 49:4992–5000. [PubMed: 16884311]

23. Brenke R, et al. Fragment-based identification of druggable ‘hot spots’ of proteins using Fourier domain correlation techniques. *Bioinformatics*. 2009; 25:621–627. [PubMed: 19176554]
24. Kozakov D, et al. Structural conservation of druggable hot spots in protein-protein interfaces. *Proc Natl Acad Sci U S A*. 2011; 108:13528–13533. [PubMed: 21808046]
25. Landon MR, et al. Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *J Med Chem*. 2007; 50:1231–1240. [PubMed: 17305325]
26. Landon MR, et al. Detection of ligand binding hot spots on protein surfaces via fragment-based methods: Application to DJ-1 and glucocerebrosidase. *J Comput Aid Mol Des*. 2009; 23:491–500.
27. Mattos C, Ringe D. Locating and characterizing binding sites on proteins. *Nat Biotechnol*. 1996; 14:595–599. [PubMed: 9630949]
28. Allen KN, et al. An experimental approach to mapping the binding surfaces of crystalline proteins. *J Phys Chem-US*. 1996; 100:2605–2611.
29. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem*. 1985; 28:849–857. [PubMed: 3892003]
30. Miranker A, Karplus M. Functionality maps of binding-sites - a multiple copy simultaneous search method. *Proteins*. 1991; 11:29–34. [PubMed: 1961699]
31. Kozakov D, et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc*. 2015; 10:733–755. [PubMed: 25855957]
32. English AC, et al. Locating interaction sites on proteins: The crystal structure of thermolysin soaked in 2% to 100% isopropanol. *Proteins*. 1999; 37:628–640. [PubMed: 10651278]
33. Dennis S, et al. Computational mapping identifies the binding sites of organic solvents on proteins. *Proc Natl Acad Sci U S A*. 2002; 99:4290–4295. [PubMed: 11904374]
34. Silberstein M, et al. Identification of substrate binding sites in enzymes by computational solvent mapping. *J Mol Biol*. 2003; 332:1095–1113. [PubMed: 14499612]
35. Kuttner YY, Engel S. Protein hot spots: The islands of stability. *J Mol Biol*. 2012; 415:419–428. [PubMed: 22100447]
36. Vajda S, Guarnieri F. Characterization of protein-ligand interaction sites using experimental and computational methods. *Curr Opin Drug Discov Devel*. 2006; 9:354–362.
37. Chuang GY, et al. Binding hot spots and amantadine orientation in the influenza A virus M2 proton channel. *Biophys J*. 2009; 97:2846–2853. [PubMed: 19917240]
38. Landon MR, et al. Novel druggable hot spots in avian influenza neuraminidase H5N1 revealed by computational solvent mapping of a reduced and representative receptor ensemble. *Chem Biol Drug Des*. 2008; 71:106–116. [PubMed: 18205727]
39. Kozakov D, et al. New frontiers in druggability. *J Med Chem*. 2015 Jul 31. Epub ahead of print.
40. Hall DR, et al. Hot spot analysis for driving the development of hits into leads in fragment-based drug discovery. *J Chem Inf Model*. 2012; 52:199–209. [PubMed: 22145575]
41. Murray CW, et al. Experiences in fragment-based drug discovery. *Trends Pharmacol Sci*. 2012; 33:224–232. [PubMed: 22459076]
42. Hubbard RE, Murray JB. Experiences in fragment-based lead discovery. *Methods Enzymol*. 2011; 493:509–531. [PubMed: 21371604]
43. Murray CW, et al. Fragment-based drug discovery applied to Hsp90. Discovery of two lead series with high ligand efficiency. *J Med Chem*. 2010; 53:5942–5955. [PubMed: 20718493]
44. Andersen OA, et al. Structure-based dissection of the natural product cyclopentapeptide chitinase inhibitor argifin. *Chem Biol*. 2008; 15:295–301. [PubMed: 18355729]
45. Lange G, et al. Requirements for specific binding of low affinity inhibitor fragments to the SH2 domain of pp60Src are identical to those for high affinity binding of full length inhibitors. *J Med Chem*. 2003; 46:5184–5195. [PubMed: 14613321]
46. Van Molle I, et al. Dissecting fragment-based lead discovery at the von Hippel-Lindau protein:Hypoxia inducible factor 1 α protein-protein interface. *Chem Biol*. 2012; 19:1300–1312. [PubMed: 23102223]
47. Barelier S, et al. Fragment-based deconstruction of Bcl-x(L) inhibitors. *J Med Chem*. 2010; 53:2577–2588. [PubMed: 20192224]

48. Babaoglu K, Shoichet BK. Deconstructing fragment-based inhibitor discovery. *Nat Chem Biol.* 2006; 2:720–723. [PubMed: 17072304]
49. Barelier S, et al. Substrate deconstruction and the nonadditivity of enzyme recognition. *J Am Chem Soc.* 2014; 136:7374–7382. [PubMed: 24791931]
50. Fry DC, et al. Deconstruction of a Nutlin: Dissecting the binding determinants of a potent protein-protein interaction inhibitor. *ACS Med Chem Lett.* 2013; 4:95–100.
51. Kozakov D, et al. Ligand deconstruction: Why some fragment binding positions are conserved and others are not. *Proc Natl Acad Sci U S A.* 2015; 112:E2585–2594. [PubMed: 25918377]
52. Teotico DG, et al. Docking for fragment inhibitors of AmpC β -lactamase. *Proc Natl Acad Sci U S A.* 2009; 106:7455–7460. [PubMed: 19416920]
53. Hall RJ, et al. Efficient exploration of chemical space by fragment-based screening. *Prog Biophys Mol Biol.* 2014; 116:82–91. [PubMed: 25268064]
54. Bohnuud T, et al. Evidence of conformational selection driving the formation of ligand binding sites in protein-protein interfaces. *PLoS Comput Biol.* 2014; 10:e1003872. [PubMed: 25275445]
55. Oltersdorf T, et al. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature.* 2005; 435:677–681. [PubMed: 15902208]
56. Tse C, et al. ABT-263: A potent and orally bioavailable Bcl-2 family inhibitor. *Cancer Res.* 2008; 68:3421–3428. [PubMed: 18451170]
57. Tao ZF, et al. Discovery of a potent and selective BCL-XL inhibitor with in vivo activity. *ACS Med Chem Lett.* 2014; 5:1088–1093. [PubMed: 25313317]
58. Berman HM, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28:235–242. [PubMed: 10592235]
59. Huth JR, et al. Discovery and design of novel Hsp90 inhibitors using multiple fragment-based design strategies. *Chem Biol Drug Des.* 2007; 70:1–12. [PubMed: 17630989]
60. Haider MK, et al. Predicting fragment binding poses using a combined MCSS MM-GBSA approach. *J Chem Inf Model.* 2011; 51:1092–1105. [PubMed: 21528911]
61. Woodhead AJ, et al. Discovery of (2,4-dihydroxy-5-isopropylphenyl)-[5-(4-methylpiperazin-1-ylmethyl)-1,3-dihydrois oindol-2-yl]methanone (AT13387), a novel inhibitor of the molecular chaperone Hsp90 by fragment based drug design. *J Med Chem.* 2010; 53:5956–5969. [PubMed: 20662534]
62. Ernst JT, et al. Identification of novel Hsp90 α/β isoform selective inhibitors using structure-based drug design. Demonstration of potential utility in treating CNS disorders such as Huntington's disease. *J Med Chem.* 2014; 57:3382–3400. [PubMed: 24673104]
63. Bollag G, et al. Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature.* 2010; 467:596–599. [PubMed: 20823850]
64. Tsai J, et al. Discovery of a selective inhibitor of oncogenic B-RAF kinase with potent antimelanoma activity. *Proc Natl Acad Sci U S A.* 2008; 105:3041–3046. [PubMed: 18287029]
65. Bollag G, et al. Vemurafenib: The first drug approved for BRAF-mutant cancer. *Nat Rev Drug Discov.* 2012; 11:873–886. [PubMed: 23060265]
66. Saxty G, et al. Identification of inhibitors of protein kinase B using fragment-based lead discovery. *J Med Chem.* 2007; 50:2293–2296. [PubMed: 17451234]
67. Breitenlechner CB, et al. Structure-based optimization of novel azepane derivatives as PKB inhibitors. *J Med Chem.* 2004; 47:1375–1390. [PubMed: 14998327]
68. Li Q, et al. Synthesis and structure-activity relationship of 3,4'-bispyridinylethylenes: Discovery of a potent 3-isoquinoliny pyridine inhibitor of protein kinase B (PKB/AKT) for the treatment of cancer. *Bioorg Med Chem Lett.* 2006; 16:2000–2007. [PubMed: 16413780]
69. Frederickson M, et al. Fragment-based discovery of mexiletine derivatives as orally bioavailable inhibitors of urokinase-type plasminogen activator. *J Med Chem.* 2008; 51:183–186. [PubMed: 18163548]
70. Katz BA, et al. Engineering inhibitors highly selective for the S1 sites of Ser190 trypsin-like serine protease drug targets. *Chem Biol.* 2001; 8:1107–1121. [PubMed: 11731301]
71. West CW, et al. Identification of orally bioavailable, non-amidine inhibitors of Urokinase Plasminogen Activator (uPA). *Bioorg Med Chem Lett.* 2009; 19:5712–5715. [PubMed: 19703768]

72. Coyne AG, et al. Drugging challenging targets using fragment-based approaches. *Curr Opin Chem Biol.* 2010; 14:299–307. [PubMed: 20223699]
73. Geschwindner S, et al. Discovery of a novel warhead against β -secretase through fragment-based lead generation. *J Med Chem.* 2007; 50:5903–5911. [PubMed: 17985861]
74. Congreve M, et al. Application of fragment screening by X-ray crystallography to the discovery of aminopyridines as inhibitors of β -secretase. *J Med Chem.* 2007; 50:1124–1132. [PubMed: 17315857]
75. Murray CW, et al. Application of fragment screening by X-ray crystallography to β -secretase. *J Med Chem.* 2007; 50:1116–1123. [PubMed: 17315856]
76. Wang YS, et al. Application of fragment-based NMR screening, X-ray crystallography, structure-based design, and focused chemical library design to identify novel μ M leads for the development of nM BACE-1 (β -site APP cleaving enzyme 1) inhibitors. *J Med Chem.* 2010; 53:942–950. [PubMed: 20043700]
77. Zhu ZN, et al. Discovery of cyclic acylguanidines as highly potent and selective β -site amyloid cleaving enzyme (BACE) inhibitors: Part I-Inhibitor design and validation. *J Med Chem.* 2010; 53:951–965. [PubMed: 20043696]
78. May PC, et al. Robust central reduction of amyloid- β in humans with an orally available, non-peptidic β -secretase inhibitor. *J Neurosci.* 2011; 31:16507–16516. [PubMed: 22090477]
79. Chen JJ, et al. Development of 2-aminooxazoline 3-azaxanthenes as orally efficacious β -secretase inhibitors for the potential treatment of Alzheimer's disease. *Bioorg Med Chem Lett.* 2015; 25:767–774. [PubMed: 25613679]
80. Wyatt PG, et al. Identification of n-(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1h-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. *J Med Chem.* 2008; 51:4986–4999. [PubMed: 18656911]
81. Finlay MR, et al. Imidazole piperazines: SAR and development of a potent class of cyclin-dependent kinase inhibitors with a novel binding mode. *Bioorg Med Chem Lett.* 2008; 18:4442–4446. [PubMed: 18617397]
82. Cheng AC, et al. Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol.* 2007; 25:71–75. [PubMed: 17211405]
83. Fauman EB, et al. Structure-based druggability assessment--Identifying suitable targets for small molecule therapeutics. *Curr Opin Chem Biol.* 2011; 15:463–468. [PubMed: 21704549]
84. Istvan ES, Deisenhofer J. Structural mechanism for statin inhibition of HMG-CoA reductase. *Science.* 2001; 292:1160–1164. [PubMed: 11349148]
85. Harel M, et al. Crystal structure of thioflavin T bound to the peripheral site of Torpedo californica acetylcholinesterase reveals how thioflavin T acts as a sensitive fluorescent reporter of ligand binding to the acylation site. *J Am Chem Soc.* 2008; 130:7856–7861. [PubMed: 18512913]
86. Lexa KW, Carlson HA. Improving protocols for protein mapping through proper comparison to crystallography data. *J Chem Inf Model.* 2013; 53:391–402. [PubMed: 23327200]
87. Bakan A, et al. Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules. *J Chem Theory Comput.* 2012; 8:2435–2447. [PubMed: 22798729]
88. Raman EP, et al. Inclusion of multiple fragment types in the site identification by ligand competitive saturation (SILCS) approach. *J Chem Inf Model.* 2013; 53:3384–3398. [PubMed: 24245913]
89. Yu W, et al. Site-identification by ligand competitive saturation (SILCS) assisted pharmacophore modeling. *J Comput Aided Mol Des.* 2014; 28:491–507. [PubMed: 24610239]
90. Hall DH, et al. Site-identification by ligand competitive saturation (SILCS) assisted pharmacophore modeling. *J Comput Aided Mol Des.* 2011; 28:491–507.

Trends Box

- Binding energy hot spots, smaller regions of binding sites that contribute a disproportionate amount to the free energy of binding any ligand, can be determined computationally from ligand-free structures of protein targets.
- Analysis of binding hot spots provides insights into the likelihood of success in identifying fragment hits, and the prospects that a particular fragment hit can be advanced into a high affinity, druglike ligand.
- The strength of the top-ranked hot spot determines what fragment hit rate can be expected.
- The degree of overlap with the top hot spot determines whether a fragment hit has a binding mode that is robust enough to facilitate elaboration into a higher affinity ligand.
- The presence of secondary hot spots close to CC1 is additionally required to support binding of a full-sized ligand with high affinity for the target, and the strength, number and spatial arrangement of hot spots determine the potential binding modes.

Outstanding Questions

- While the binding hot spots are less sensitive to conformational changes than binding sites are, the shape and importance of hot spots may be affected by ligand binding. Which structure is the best to map if several structures are available?
- How the mapping results change if a different set of small molecules is used as probes? Does the currently used set provide sufficient information?
- Some of the hot spots partially overlap. Should these be considered separate or part of the same hot spot when analyzing druggability and fragment conservation?
- The strength of hot spots (i.e., the number of probe clusters) determines the maximum potential contribution to the free energy when a ligand binds at the hot spot. Can this relationship be better quantified?
- Hot spots provide information on binding properties of the protein surface. Can this information be converted into structure-based pharmacophores and used for drug discovery more directly?

Figure 1.

Hot spots and selected ligands of HSP-90 and B-Raf. Top row: HSP-90 ligands. The HSP-90 hot spots shown, based on the mapping on an HSP-90 apo structure (PDB ID 1yes) are CC1 (cyan, 22 probe clusters), CC2 (magenta, 13 clusters), CC3 (salmon, 12 clusters), and CC5 (green, 9 clusters). Hot spots CC1, CC2, and CC5 overlap. **(a)** Fragment 2-methyl-4-diethylamide-phenol from PDB ID 2xdl ($K_d = 790,000$ nM) overlaps with hot spots CC1, CC2, and CC5. **(b)** Inhibitor from PDB ID 2xab ($K_d = 0.54$ nM) retains the binding mode of the fragment, but overlaps better with CC1, resulting in much improved affinity. **(c)** Inhibitor from PDB ID 2xjx ($K_d = 0.71$ nM) has been developed to have better pharmaceutical properties, but the added moiety did not interact with any hot spot, and the affinity did not improve. **(d)** Benzolactam inhibitor from PDB ID 4o0b ($K_d = 5$ nM). The inhibitor only partially overlaps with CC5, but protrudes into CC3. The reduced K_d value demonstrates that interactions with regions of the protein surface between the hot spots do not contribute much to the binding free energy. Bottom row: BRAF ligands. The BRAF hot spots shown, based on the mapping of a BRAF(V600E) structure co-crystallized with the ligand PLX4720 (PDB ID 3c4c) are CC1 (cyan, 16 probe clusters), CC2 (magenta, 15 clusters), CC3 (yellow, 15 clusters), and CC4 (salmon, 9 clusters). **(e)** Fragment N-phenyl-1H-pyrrolo[2,3-b]pyridin-3-amine, co-crystallized with Pim-1 kinase domain rather than BRAF, from PDB ID 3c4e ($IC_{50} = 100,000$ nM) overlaps with CC1 and CC3. **(f)** Inhibitor PLX4720, co-crystallized with BRAF(V600E), from PDB ID 3c4c. PLX4720 retains the core 1H-Pyrrolo[2,3-b]pyridine fragment, but the fragment position is slightly shifted relative to its position in Pim-1. The overlap with CC3 is substantially improved, and PLX4720 reaches into both CC2 and CC4, resulting in much higher affinity ($IC_{50} = 13$ – 160 nM). **(g)** PLX3203, co-crystallized with BRAF(V600E), from PDB ID 4fk3, retains the position of the core fragment in CC1 and CC3, and improves overlap with CC2 and CC4. **(h)** Vemurafenib (PLX4032) from PDB ID 3og7 also retains the position of the core, and further optimizes overlap with CC4 ($IC_{50} = 21$ – 100 nM).

Figure 2.

Hot spots and selected ligands of protein kinase B/A (PKB/PKA) and urokinase-type plasminogen activator (uPA). Top row: PKB/PKA ligands. The PKB/PKA hot spots, based on the mapping a PKA-PKB chimera co-crystallized with (2R)-2-(4-chlorophenyl)-2-phenyl-ethanamine (PDB ID 2uw8) are CC1 (cyan, 21 probe clusters), CC3 (yellow, 14 clusters), CC7 (orange, 5 clusters), and CC8 (green, 3 clusters). **(a)** 3-methyl-4-phenyl-1h-pyrazole fragment (binding to the PKA-PKB chimera) from PDB ID 2uw3 (IC_{50} = 80,000 nM) overlaps very well with CC1, but does not reach any other hot spot. The high level of overlaps predicts the conservation of fragment binding mode. **(b)** Extension of the fragment into CC3 yields the potent inhibitor from PDB ID 2uw5 (IC_{50} = 34 nM). Again, the X-ray structure was determined for a PKA-PKB chimera. **(c)** Different chemotype but fairly similar binding mode in the potent inhibitor binding to PKA from PDB ID 1sve (IC_{50} = 5 nM). The inhibitor interacts well with CC1, and reaches all other hot spots. **(d)** Potent inhibitor of PKA from PDB ID 2f7e (IC_{50} = 2.1 nM) retains the binding mode in CC1 and has very good overlap with CC3. It appears that this is sufficient for strong binding, in spite of no interactions with CC7 and CC8. Bottom row: uPA ligands. The uPA hot spots, based on the mapping of a uPA apo structure (PDB ID 2o8t) are CC1 (cyan, 29 probe clusters), CC2 (magenta, 26 clusters), CC3 (yellow, 13 clusters), and CC6 (blue, 3 clusters). **(e)** Fragment from PDB ID 2vin (IC_{50} = 1,000,000 nM) overlaps only with CC1, but the high level of overlap predicts conservation of the binding mode. **(f)** Fragment based inhibitor from PDB ID 2viw (IC_{50} = 72 nM) extends into CC2 and somewhat into CC3. **(g)** Inhibitor with different chemotype from PDB ID 1gj9 (K_i = 11 – 140 nM), resulting in better overlaps with CC1 and CC2. **(h)** An orally bioavailable, non-amidine Inhibitor with a third chemotype and binding mode from PDB ID 3ig6 (IC_{50} = 25 nM) has good overlaps with CC1 and CC2, and reaches into CC3.

Figure 3.

Hot spots and selected ligands of β -secretase (BACE-1) and cyclin dependent kinase 2 (CDK2). Top row: BACE-1 ligands. The BACE-1 hot spots, based on the mapping of a BACE-1 apo structure (PDB ID 1w50) are CC1 (cyan, 19 probe clusters), CC3 (yellow, 16 clusters), CC5 (white, 7 clusters), CC7 (orange, 5 clusters) and CC8 (green, 5 clusters). The hot spot structure is very diffuse, and some hot spots partially overlap. **(a)** 4-butoxy-3-chlorobenzyl imidothiocarbamate fragment from PDB ID 3kmx ($K_d = 15,000$ nM) only partially overlaps with any hot spot. As discussed in the paper, the overlap with CC1 is not sufficient for the conservation of the binding mode, and the fragment was used only to capture some of the important interactions. **(b)** Inhibitor from PDB ID 3l5f ($IC_{50} = 605$ nM) overlaps better with CC1, CC3, and CC7, and reaches CC8, resulting in higher affinity. **(c)** The potent inhibitor LY2811376 from PDB ID 4ybi retains the binding mode of the inhibitor in **3b**, and also reaches into CC8. **(d)** The inhibitor by Amgen from PDB ID 4xkx also retains the binding mode. Bottom row: CDK2 ligands. The CDK2 hot spots, from the mapping of a CDK2 apo structure (PDB ID 4ek3) are CC1 (cyan, 24 probe clusters), CC3 (yellow, 8 clusters), and CC5 (white, 6 clusters). **(e)** Fragment from PDB ID 2vtm ($IC_{50} = 1,000,000$ nM) overlaps only with CC1, and the level of overlap seems to be too low to yield conservation of the binding mode. **(f)** Inhibitor containing parts of the fragment from PDB ID 2vtn ($IC_{50} = 850$ nM) extends into CC3. **(g)** Inhibitor from PDB ID 2vtp ($IC_{50} = 3$ nM) was developed from the previous compound, but it also extends into CC5, resulting in much improved binding affinity. **(h)** An imidazole piperazine inhibitor with another binding mode from PDB ID 2vv9 ($IC_{50} = 17$ nM). The inhibitor overlaps well with CC1 and CC3, and slightly extends toward CC5.

Table 1

Probe clusters in hot spots and the number of ligand-hot spot interactions in drug targets

Target (PDB ID mapped)	Ligands ^a	Hot spots ^b	Hot spot ^c (probe clusters ^d , ligands ^e)
HSP-90 (1yes)	167	4	1(22,160), 2(13,165) ^f , 3(12,76), 5(9,159) ^f
B-RAF (3c4c)	35	5	1(16,35), 2(15,32), 3(15,34), 4(9,24), 5(8,17)
PKB (2uw8)	85	4	1(21,84), 3(14, 46), 7(5,12), 8(3,68) ^g
uPA (2o8t)	74	4	1(29,67), 2(26,22), 3(13,9), 6(3,11)
BACE (1w50)	288	5	1(19,279), 3(16,288), 5(7,196), 7(5,273), 8(5,207)
CDK2 (4ek3)	280	3	1(24,280), 3(8,270) ^f , 5(6,186)
ACE-1 (2iul)	12	2	1(23,12), 2(13,11),
Acetylcholine esterase (1ea5)	45	4	1(21,40), 2(15,25), 3(11,23), 8(7,5)
Aldose reductase (1ads)	55	4	1(30,55), 3(13,12), 4(10,54) ^f , 7(3,41)
cAbl kinase (2g2f)	23	3	1(23,23), 3(14,20) ^f , 4(12,22) ^f
COX-2 (5cox)	28	2	2(17,28) ^h , 6(7,27)
EGFR kinase (1m14)	31	4	1(16,31), 2(13,30), 3(13,8), 4(10,26) ⁱ
Factor Xa (1lpz)	119	2	1(28,119), 2(17,118)
Fungal Cyp51 (1h5z)	9	3	1(17,9), 2(16,8) ^f , 4(12,8)
HIV-1 protease (1odw)	257	6	1(21,244), 2(20,254), 3(16,238), 4(10, 251), 5(9, 241), 6(8,254)
HMG CoA reductase (1hw8) ^j	18	2	1(14,18), 3(11,18)
MDM2 (1z1m) ^j	31	3	1(21,31), 3(16,31), 6(5,29) ^k
Neuraminidase (1ivg)	13	2	1(27,13), 3(12,13)
P38 kinase (2zb0)	170	3	1(16,168), 4(9,84), 7(6,150)
Phosphodiesterase 4D (3sl3)	28	3	1(27,28), 3(14,25), 5(9,28)
Phosphodiesterase 5A (1t9r)	15	2	1(24,15), 3(16,15)
Thrombin (2uuf)	214	4	1(31,210), 2(16,178), 5(5,208), 7(4,13)
Cathepsin K (1mem)	37	3	1(24,37), 2(19,28), 5(11,13)

^aNumber of ligands in the PDB;^bNumber of hot spots in the ligand binding site;^cRank of consensus cluster, where CCN is simply given as N;^dNumber of probe clusters in the consensus cluster;^eNumber of ligands that are within 1 Å of any probe in the particular hot spot;^fHot spot partially overlaps with CC1;^gLocated between CC1 and CC3;^hCC1 is located at the heme site;ⁱHomotetramer, no ligand-free structure;^jModel 4 of NMR structure;^kPartially overlaps with CC3.