KEYNOTE (GREEN)

Which cryptic sites are feasible drug targets?

Maria Lazou [1], **Dima Kozakov** [3,4], **Diane Joseph-McCarthy** [1,2], **Sandor Vajda** [1,2,*]

[1] Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA
[2] Department of Chemistry, Boston University, Boston, MA 02215, USA
[3] Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794, USA
[4] Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794, USA

Cryptic sites can expand the space of druggable proteins, but the potential usefulness of such sites needs to be investigated before any major effort. Given that the binding pockets are not formed, the druggability of such sites is not well understood. The analysis of proteins and their ligands shows that cryptic sites that are formed primarily by the motion of side chains moving out of the pocket to enable ligand binding generally do not bind drug-sized molecules with sufficient potency. By contrast, sites that are formed by loop or hinge motion are potentially valuable drug targets. Arguments are provided to explain the underlying causes in terms of classical enzyme inhibition theory and the kinetics of side chain motion and ligand binding.

Keywords: Ligand-binding affinity; Side chain movement; Time scale; Competitive inhibition; Induced fit; Conformational selection

Maria Lazou is a third-year PhD student in the Department of Biomedical Engineering at Boston University. Her interests are in modeling of biomolecular systems with emphasis on molecular interactions.

Dima Kozakov received an MSc in applied mathematics and physics at the Moscow Institute of Physics and Technology, and a PhD in biomedical engineering at Boston University. Currently. He is a professor in the Department of Applied Mathematics at Stony Brook University, an Affiliate Member at Laufer Center for Physical and Quantitative Biology, and an Affiliate Member at the Institute of Advance Computational Sciences. Dr Kozakov applies methods from machine learning and physics for modeling of biological macromolecules, with an emphasis on molecular interactions and drug design, and its application to understanding disease at the systems level with atomic resolution. Dr Kozakov's approaches are heavily used in the pharmaceutical industry and are consistently top-performing in blind molecular modeling completions.

Diane Joseph-McCarthy is the executive director of the Bioengineering Technology & Entrepreneurship Center and professor of biomedical engineering and chemistry at Boston University. Before that, she was a life science executive with over 20 years of drug discovery, development, and leadership experience. She was senior vice president of Discovery & Early Development at EnBiotix, a bioengineering company focused on respiratory and rare disease. She was also an associate director at AstraZeneca, where she led a global team in Predictive Science. At Wyeth, she held positions of increasing responsibility. Diane received her PhD from MIT and was a postdoctoral fellow at Harvard University/Harvard Medical School. She is a fellow of the American Institute for Medical and Biological Engineering.

Sandor Vajda is a professor of biomedical engineering and chemistry at Boston University. He received his MSc in applied mathematics and PhD in chemistry in Budapest, Hungary. He held research positions in the Department of Engineering, University of Warwick, and the Department of Chemistry, Princeton University. He has been active in method development for modeling of biological macromolecules, with a focus on the analysis of protein–protein interactions, the characterization of protein-binding sites, target identification, druggability assessment, and the development of computational methods for fragment-based drug discovery. His group developed the protein–protein docking program and server ClusPro and the FTMap family of protein-mapping servers for the identification of binding hot spots.

* *Corresponding author. Vajda, S. (vajda@bu.edu)*

https://doi.org/10.1016/j.drudis.2024.104197

Introduction

Several medically important proteins are difficult to drug or might even be undruggable because they have no functional orthosteric site that can bind drug-sized small molecules with sufficient affinity and/or specificity to compete with endogenous ligands.[p1],[p2],[p3],[p4] This outcome generally occurs if the site is too shallow and featureless, or if it is too polar or charged.[p5] For example, high-affinity active-site inhibitors of tyrosine phosphatases, such as PTP1B, would need to emulate the charged nature of the phosphorylated substrates. In addition, achieving selectivity would require very large compounds because of the similarity of residues in other phosphatases directly surrounding the site.[p6],[p7],[p8] Thus, while many high-affinity ligands are known, none have been developed into a viable drug.[p7] The difficulty of binding to functional sites highlights the desire to target other potential sites on the protein, frequently referred to as exosites.[p9] In many cases, ligands at such sites might modulate binding at the functional site by allosteric effects.[p10],[p11] Exosites even without allostery can be useful for binding proteolysis targeting chimeras (PROTACs) and molecular glues.[p12]

Given that exosites generally do not bind endogeneous ligands, they frequently do not form well-defined pockets in ligand-free protein structures and might be discovered by serendipity when some ligands, such as crystallization additives, bind to the protein.[p13] This type of exosite is referred to as a cryptic site.[p14] Thus, a binding site is generally considered cryptic if it can be identified in the ligand-bound but not in the unbound structure of a protein. However, this definition is far from rigorous or complete.[p14] In fact, proteins are dynamical objects and, hence, the volume of any ligand-binding site in the apo structure follows a distribution, and for cryptic sites this distribution is simply biased toward lower values.[p15] While cryptic sites are particularly important if the orthosteric binding site is not druggable, many cryptic pockets are also present at the surface of proteins that already exhibit known druggable orthosteric binding sites. To develop and test a method for the identification of cryptic sites, Cimermancic *et al.* created the CryptoSite set of 93 ligand-free and ligand-bound pairs of proteins from the Protein Data Bank (PDB), with closed pockets in the selected ligand-free structures.[p1] Further analysis showed that 80% of these proteins have other ligand-free structures in the PDB with at least a partially open pocket at the cryptic site.[p16] However, placing ligands from holo structures of the protein into any apo structure without substantially adjusting the conformation of such proteins would create steric clashes and, hence, drug discovery by virtual screening that relies on rigid protein structures remains difficult, especially for cryptic sites.

Finding novel cryptic sites is a major challenge in drug discovery. Experimental identification of cryptic sites generally involves screening libraries of small molecules or fragments,[p17],[p18] site-directed tethering,[p17],[p19] or the use of antibodies.[p20] All of these tools are costly and only moderately effective unless substantial preliminary information is available about the location of a potential cryptic site and for selecting an appropriate library of small molecules. A more recent technology is chemoproteomics, which addresses the limitations of conventional screening techniques to discover previously unknown pockets on the surface of proteins and identify small molecules that bind to those targets.[p12] However, chemoproteomic probes create covalent interactions and, thus, target sites that most likely bind only covalent ligands with high affinity.

Several computational methods have also been used for the identification of cryptic sites. The most obvious approaches are based on molecular dynamics (MD) simulations, and can involve Markov state models,[p21] accelerated MD,[p22] or collective variable enhanced sampling methods.[p23] Although computationally expensive, such tools can open transitional pockets.[p21],[p23],[p24],[p25] An improved version of the general approach is mixed MD, also called cosolvent MD.[p26] Mixed MD methods involve simulating proteins in a solution of water and cosolvent molecules.[p23],[p26],[p27],[p28],[p29] It has been observed that the presence of small molecules in the simulation enhances the opening of pockets.[p23],[p30] More recently, cryptic sites have been identified by protein structure prediction and other applications of artificial neural networks.[p31],[p32]

Although several methods focus on finding cryptic sites,[p33],[p34] there is more limited discussion of the potential usefulness of the sites for drug discovery. This question is important because MD and other simulation tools can open many transitional pockets for periods of time that could thus be considered potential cryptic sites.[p24],[p35],[p36] It is unlikely that all these sites can bind small molecules with high affinity and, therefore, it is reasonable to assess their potential ligandability (frequently referred to as druggability).[p37],[p38] However, most computational tools of druggability analysis are based on the analysis of pocket properties and, hence, without well-defined pockets are of limited use.[p2] Here, we discuss our protein-mapping program FTMap,[p39] which was shown to detect binding hot spots even when applied to unbound protein structures,[p16] but in case of cryptic sites frequently overestimates druggability.

The main result of this paper is the claim that the potential druggability of a cryptic site substantially depends on the mechanism of pocket opening. After analyzing numerous proteins with known cryptic sites and the ligands that bind to them, we made the surprising observation that sites formed by side chains moving out of the pocket to enable ligand binding, possibly accompanied by very small backbone conformational changes, generally do not bind small ligands with low nanomolar affinity and, hence, are not suitable for the development of noncovalent drugs. The K_d or IC_{50} values of any such molecule are at best in the low micromolar or high nanomolar range, and optimization generally has moderate impact. Although oral drugs seldom have nanomolar potency (50 nM on average),[p40] these cryptic sites are generally orders of magnitude weaker. In addition, despite the moderate potency of oral drugs, the sites need to exhibit high potential binding affinity, because the actual potency is generally reduced in the process of optimizing absorption, distribution, metabolism, and excretion–toxicity (ADMET) properties.[p40] In contrast to the sites formed by side chain motion, sites that are formed by loop or hinge motion are potentially valuable drug targets. These hypotheses were formulated by simple analysis of proteins with ligands that bind at cryptic sites. Our goal here is to explore this limitation and provide a plausible biophysical

explanation based on a straightforward argument borrowed from the classical theory of enzyme inhibition,[p41] and on an analysis of binding kinetics affected by the relative timescales of side chain and loop motion.

Druggability of cryptic sites

Early experimental approaches to determining druggability were based on the hit rates of fragment screening by nuclear magnetic resonance (NMR) or X-ray crystallography.[p42],[p43],[p44],[p45] The major advantages of experimental methods are that they can be applied to proteins without any known ligand, and can provide potential fragment hits along with the locations of binding sites. In addition, as discussed below, these are the only approaches that directly apply to cryptic sites. However, establishing druggability by such experiments is expensive and, as such, a variety of computational methods have been developed. Hajduk *et al.* used their NMR data to derive a regression-based model to predict the likely fragment hit rate for untested targets, based on structural properties of the protein-binding site, such as total and apolar surface area and pocket compactness, generating a predicted druggability score for each.[p42],[p43] More generally, statistical and machine learning-based methods correlate pocket properties, including geometry, amino acid content, and energetics, with druggability data available for traditional targets.[p46],[p47],[p48] Both approaches are trained on benchmark sets of proteins that are known to bind small molecules with high affinity and, thus, evaluate whether the novel target in question contains a binding pocket that in some sense resembles pockets on known druggable targets that bind conventional drug-like ligands.[p43],[p46],[p48] Therefore, these druggability assessment tools assume that the ligand-binding pocket is well defined and sufficiently open, which is a clear limitation for the analysis of cryptic sites if only the unbound structure of the protein is available. One possible solution is generating an ensemble of open conformations using the simulation tools described in the Introduction for the identification of cryptic sites. However, as demonstrated for the Fpocket-based druggability scores,[p49] the results heavily depend on the geometry and other properties of the resulting structures,[p15] making reliable predictions for cryptic sites difficult.

The alternative to statistical methods is virtual screening, which uses sets of probes to identify sites where such probes accumulate.[p50],[p51],[p52] Druggability analysis by such methods is a natural extension of cosolvent MD simulations.[p50],[p52] However, this approach is computationally expensive and the results still depend on the shape of the resulting pocket. Here, we discuss the use of FTMap,[p39] which was extensively utilized for the analysis of druggability[p3] and provides some binding site information without lengthy simulations, even for cryptic sites.[p16] FTMap was originally developed to identify binding hot spots, regions on the protein surface where ligand binding makes potentially large contributions to binding free energy.[p39] The program distributes small organic probe molecules of different size, shape, and polarity on the surface of the protein to be studied, finds favorable positions for each probe type, clusters the probes, and ranks the clusters on the basis of their average energy. Regions that bind several probe clusters are called consensus sites (CSs) and are the predicted binding hot spots. The hot spots are ranked based on the number of probe clusters they include, starting from the strongest hot spot, denoted as 0(*xx*), where *xx* is the number of probe clusters. FTMap maps rigid X-ray structures, which is a limitation, but it has relatively moderate sensitivity to conformational variation, and its modified version, FTFlex, can explore the impact of alternative side chain conformations in the ligand-binding site.[p53]

FTMap has been used to analyze the druggability of over 150 proteins from a variety of druggability benchmarks.[p3] It was established that a necessary condition for a site to bind any ligand is to have a strong hot spot, with 13 or more probe clusters for at least micromolar affinity (which we call 'borderline druggable'), and with at least 16 probe clusters for nanomolar affinity.[p3] We added two more conditions for druggability, which are at least one additional hot spot within 8 Å from the strongest hot spot, and at least a 10 Å diameter of the region containing the main and nearby secondary hot spots.[p3] However, these additional conditions are needed only to determine whether the site can be targeted by traditional drug-like ligands or whether there is a need for other chemotypes, such as macrocyclic compounds or peptide-derived foldamers,[p3] and, thus, will not be discussed here. The primary motivation for presenting FTMap results here is our observation that cryptic sites in ligand-free structures generally have a binding energy hot spot very close by, even in the absence of a well-formed pocket.[p16] This is because the existence of a hot spot is required for the attraction resulting in ligand binding.[p16] Given that strong hot spots predict druggability for noncryptic binding sites, we expected that the results would also provide useful information for cryptic sites. However, as we discuss below, the condition overestimates the druggability of many sites that are formed by side chain motion without substantial change in backbone conformation.

Mechanisms of cryptic site opening

We studied the types of conformational change that occur between unbound and bound structures in the 93 structures of the CryptoSite set. In 18 cases, there were no significant main-chain motions in the vicinity of the binding site, and the cause of the pocket being cryptic was that one or more side chains protruded into the site in the unbound structure. Table 1 lists the ten proteins in this group that had published ligand-binding affinity data available. PDB IDs of unbound and bound structures are shown as given in the CryptoSite set. As mentioned above, many proteins in the CryptoSite set have additional unliganded structures that are more open than those included in the original benchmark,[p16] and we added some unbound structures that had a better-formed pocket at the cryptic site.

To demonstrate the important role of side chains occluding cryptic sites, we consider two slightly different examples. The first is the kinesin-like mitotic motor protein. In the CryptoSite set, this protein is represented by the unbound structure 3HQD.A and the bound structure 1Q0B.B, where the extensions of the four-letter PDB codes indicate the particular chains considered. 1Q0B.B has been co-crystallized with the small molecule monastrol, which has the PDB identifier NAT. Figure 1a shows

TABLE 1

Proteins from the CryptoSite set with cryptic sites created by side chain movement only

Name	Unbound[a]	Bound[b]	Ligand[c]	Affinity, nM[d]	FTMap[e]		Why is this site cryptic?
					Unbound	Bound	
Kinesin-like mitotic motor protein KIF11	3HQD.A 4A28.B	1Q0B.B	NAT	IC_{50} = 7000	4(9), 5(7) 0(26)	0(15), 3(14)	In 3HQD.A, E116 and W127 protrude into the cryptic site. In 4A28.B, E116 moves out of the site and W127 is missing
Myosin II	2AKA.A 2XEL.A	1YV3.A	BIT	IC_{50} = 4900	0(17), 4(8) 0(20)	1(14)	In 2AKAA, the side chains of L262 and Y634 protrude into the very narrow binding site
Chitinase B1	3CHE.A	2IUZ.B	D1H	K_i = 2800	0(20), 2 (19)	1(16), 2(13)	In 3CHEA, W384 protrudes into the site and would clash with the ligand
Adipocyte lipid-binding protein	1ALB.A	1LIC.A	HDS	K_d = 1700–2000	0(26)	0(24), 1(20)	Side chain of F57 protrudes into the site, and compete with ligand binding
Ricin	1RTC.A 1UQ4.A	1BR6.A	PT1	IC_{50} = 600 000	0(23), 1 (15) 0(33), 1 (20)	2(13), 4(11)	In 1RTCA, the side chain of Y80 protrudes into the binding site
Androgen receptor	2AX9.A	2PIQ.A	RB1	IC_{50} = 50 000	4(8)	6(8)	In 2AX9A, the side chains of M734 and K720 protrude into the site
Ribonuclease A	1RHB.A 3EUX.B	2W5K.B	NDP	K_i = 12 000	0(17), 1 (16) 0(34)	2(16)	H119 protrudes into the ligand-binding site
Pyruvate dehydrogenase kinase isoform 1	2Q8F.A	2Q8H.A	TF4	K_d = 1 000 000	N/A	2(15)	H149 protrudes into the site, but turns outward in 2Q8HA
Transthyretin	2H4E.B	3CFN.B	2AN	K_d = 4300	1(13)	0(20)	In 2H4EB, K15 and L17 protrude into the site
Exodeoxyribonuclease I	1FXX.A	3HL8.A	BBP	K_i = 32 000	N/A	5(8), 7 (5)	In 1FXXA, W245 protrudes into the surface site

[a] Protein Data Bank ID of the unbound structure in the CryptoSite database with additional unbound structures. The chain identifiers are added to the 4-letter PDB IDs.
[b] Protein Data Bank ID of the bound structure with chain identifier added to the 4-letter PDB ID.
[c] Three-letter code of the ligand in the bound structure.
[d] Affinity as reported.
[e] Number of probe clusters as determined by FTMap. For example, 0(26) means that the top hot spot includes 26 probe clusters.

1Q0B (green) and the ligand monastrol (yellow) and the unbound structure 3HQD.A (cyan), which has the side chains of E116 and W127 protruding into the ligand-binding pocket. As shown, these side chains would clash with the ligand and, thus, have to move out upon ligand binding. We mapped both the unbound and bound protein structures using FTMap. For 3HQD.A, FTMap finds two weak hot spots 4(9) and 5(7), with 9 and 7 probe clusters, respectively, at the cryptic site (Table 1). The hot spots of the bound structure (after removing the ligand) are 0(15) and (14), thus slightly stronger. As discussed in the previous section, nanomolar binding requires a hot spot with at least 16 probe clusters and, thus, FTMap correctly predicts that the cryptic site is a poor binder. However, the kinesin-like mitotic motor protein also has another unbound structure, 4A28.B, which has a more open cryptic site. In 4A28.B, the side chain of E116 (lilac) is turned out of the pocket, and the side chain of W127 is not visible in the X-ray structure (Figure 1a). The mapping of 4A28.B places the strongest hot spot 0(26) with 26 probe clusters at the cryptic site (Table 1). Nevertheless, monastrol binding at this site has very low potency (Table 1 and see below).

The second example of cryptic site formed by site chain motion 1 is myosin II (Table 1). Whereas the closing and opening of the cryptic site in myosin II are primarily caused by side chains, a small local backbone conformational change involving three residues also occurs. Figure 1b shows the structure 1YV3.A of myosin II (green) co-crystallized with a blebbistatin inhibitor (yellow), as well as the side chains of L262 and Y634 from the superimposed ligand-free structure 2AKA.A (cyan), which would clash with the ligand. Ligand binding also causes a small change in the conformation of loop 261–263, but it is important that the moving side chains have the main role in restricting the pocket rather than the backbone in the loop region. 2AKA.A has two hot spots, 0(17) and 4(8), at the cryptic site. Similarly to the kinesin-like mitotic motor protein, myosin II also has an unbound structure, 2XEL.A, which has a more open site compared with 2AKA.A. 2XEL.A has a strong hot spot 0(20) at the cryptic sites. However, as in the previous case, the site binds ligands with very low potency (Table 1). More generally, despite the low potency of the known ligands, FTMap predicts strong binding hot spots for six of the ten proteins in Table 1 and, thus, substantially overpredicts the expected druggability.

The most frequent origins of cryptic sites are not in fact side chains that protrude into the binding pocket, but loops or secondary structure elements, usually short α-helices. Table 2 lists the 27 proteins from the CryptoSite set that have binding affinity data and cryptic sites that are formed by one of the following mechanisms: (1) loops that protrude into the site, making it closed to ligand binding; (2) loops that are too open in the unbound structure, resulting in a poorly formed pocket; (3) secondary structure elements that are too close, closing the pocket; and (4) secondary structure elements that are too far, making the pocket too open. For example, in the unbound structure 2GFC.A (cyan) of the catalytic subunit of the cAMP-dependent protein kinase (PKA), loop 51–56 protrudes into the ligand-binding site

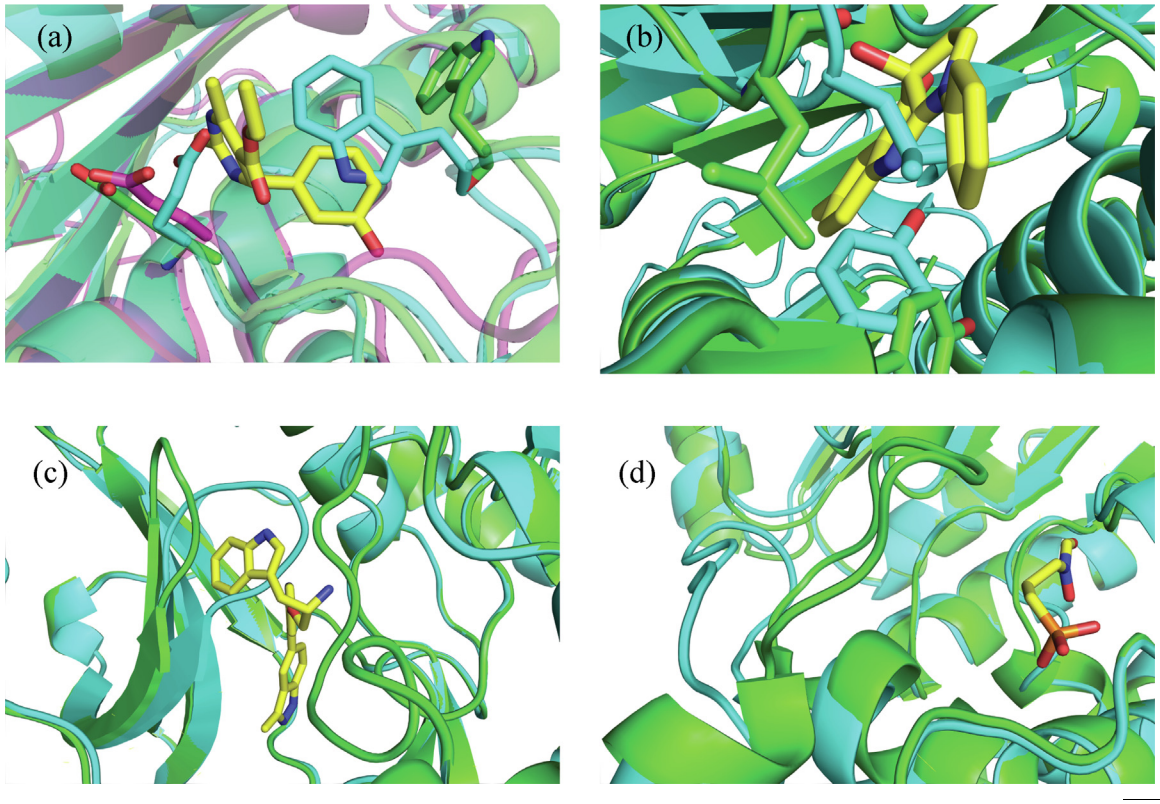

FIGURE 1

Cryptic site opening resulting from side chain and loop motion. **(a)** Structure 1Q0B.B from the Protein Data Bank (PDB) of the kinesin-like mitotic motor protein KIF11 (green) co-crystallized with monastrol (yellow), and a ligand-free structure 3HQD (cyan). In 3HQD.A, the side chains of E116 and W127 protrude into the cryptic site, whereas, in 1Q0B.B, both side chains move out of the site. However, unbound structures also exist without steric clashes with side chains. For example, in the structure 4A28.B (lilac), the side chain of E116 moves out of the site and W127 is missing. **(b)** Structure 1YV3.A of myosin II (green) co-crystallized with a blebbistatin inhibitor (yellow), and the ligand-free structure 2XEL.A (cyan). Side chains of L262 and Y634 of both proteins are shown as sticks and reveal that these side chains would clash with the ligand in the ligand-free structure, indicating that the site is cryptic. Both side chains move out of the site upon ligand binding. **(c)** Structure 2JDS.A of cAMP-dependent protein kinase (green), co-crystallized with the small ligand L20 (yellow). In the unbound structure 2GFC.A (cyan), the loop of residues 51–56 protrudes into the ligand-binding site so that even the backbone would clash with the ligand. **(d)** Structure 2EGH.B of DXP reductoisomerase (green) co-crystallized with 2-phosphoglycolic acid (yellow). In the unbound structure 1K5H.C (cyan), the loop 208–215 moves far to the left, and the site becomes too open.

such that even the backbone would directly clash with the ligand L20 (Figure 1c). 2GFC.A has two strong hot spots, 0(18) and 1 (16), at the cryptic site, indicating potential druggability. In the bound structure 2JDS.A (green), the loop turns away from the site, and the binding hot spots at the site are 0(20) and 1(16), thus even stronger than for 2GFC.A.

DXP reductoisomerase (Figure 1d) is an example of a protein in which the ligand-binding site is too open in the unbound structure, and the pocket is formed by a loop closing down on the site upon ligand binding. In the unbound structure 1K5H. C, loop 206–216 (cyan, shown on the left of Figure 1d) leaves the site too open, but the loop moves to the right and forms the pocket in the structure 2EGH.B (green), co-crystallized with 2-phosphoglycolic acid (yellow). The cryptic pocket in the unbound structure 1K5H.C has only two weak hot spots, 0(14) and 3(11), but chain B of the same protein, 1K5H.B, has a more open site and two strong hot spots, 0(19) and 1(18). FTMap predicts strong hot spots with 16 or more probe clusters for 22 of the 27 proteins in Table 2, even when applied to unbound structures. The hot spots are weak for the four proteins that also have the lowest potency compounds (beta-lactoglobulin, acid-beta-glucosidase, P38 MAP kinase, and glutamate receptor). However, most of these proteins have other unbound structures in the PDB, and the number of probe clusters vary substantially when considering different structures. Mapping the bound structure (after removing the ligands) yields more consistent results, but requires a protein structure co-crystallized with a bound ligand. Thus, while FTMap provides some information, the analysis of druggability for cryptic sites needs further development.

Impact of flexible side chains on binding affinity of cryptic sites

Comparing the data in Tables 1 and 2 shows striking differences in the affinities of available ligands. If forming the cryptic site is caused by side chain motion and possibly small movements of short loop regions with two or three residues, the binding affinity of the site is low, in the micromolar range. Although we could not find binding data for all such proteins in the CryproSite set, in no case was there evidence of strong binding

TABLE 2

Proteins from the CryptoSite set with cryptic sites formed by loop or short secondary structure motion

Name	Unbound[a]	Bound[b]	Ligand[c]	Affinity, nM[d]	FTMap[e]		Why is this site cryptic?
					Unbound	Bound	
PKA	2GFC.A 4DFZ.E	2JDS.A	L20	IC_{50} = 27, Ki = 6.3	0(18), 1 (16) 0(26), 1 (20)	0(20), 1(16)	In unbound structures, loop 51–56 protrudes into the cryptic site
DXP reductoisomerase	1K5H.C 1K5H.B	2EGH.B	FOM	K_i = 38	0(14), 3 (11) 0(19), 1 (18)	0(15), 3(11)	In unbound structures, loop 205–216 moves outward, making the pocket too open
Hepatitis C virus polymerase	3CJ0.A	2BRL.A	POO	IC_{50} = 18	1(16)	2(17)	In unbound structures, the binding site is occupied by a small α-helix, which is displaced and becomes disordered upon ligand binding
Hepatitis C virus polymerase	3CJ0.A	3FQK.B	79Z	IC_{50} = 81	0(17)	0(18), 5(8)	In unbound structures, loop 364–369 moves toward the 79Z binding site.
Tyrosine kinase domain of C-MET	1R1W.A	3F82.A	353	IC_{50} = 4.6	0(19)	0(18), 1(18)	In 1R1WA, loop 1220–1230 would clash with the ligand
Angiopoietin-1 receptor	1FVR.A	2OO8.X	RAJ	IC_{50} = 1	2(17), 3 (15)	0(16), 2(11)	In 1FVRA, loop 981–995 protrudes into the binding site
Biotin carboxylase	1BNC.B 2V5A.B 2W6P.B	2V5A.A	LZL	IC_{50} = 150	1(16), 4 (11) 0(19) 0(20)	0(23), 2(11)	In unbound structures, loop 159–169 is missing or partially unfolded, which makes the site too open
Staphylococcal nuclease	1TQO.A	1TR5.A	THP	$K_i \sim 100$	0(27), 4(8)	0(19)	In 1TQOA, loop 114–118 protrudes into the site
Glutamate racemase (GluR)	2OHG.A	2OHV.A	NHL	K_i = 16	0(22),1 (21)	0(25), 1(18)	In 2OHGA, loop 41–45 protrudes into the pocket
Serotonin *N*-acetyltransferase	1B6B.A	1KUV.A	CA5	K_i = 22	0(24)	0(19)	In 1B6BA, loop 52–63 protrudes into the binding site
Coagulation factor VII zymogen	1JBU.H 1FAK.H.	1WUN.H	P5B	IC_{50} = 62	1(16) 0(30)	0(24)	In 1JBUH, loops 184–189 and 220–222 are missing
NPC2 lysosomal protein	1NEP.A 2HKA.A	2HKA.C	C3S	K_d = 30–50	N/A 0(24)	0(26)	In the unbound structures, loop 96–103 protrudes into the ligand-binding site. 2HKA.A has no bound ligand but a strong hot spot
Hsp90	2QFO.B 1YES.A	2WI7.A	2KL	IC_{50} = 58	0(17), 2 (12), 0(28), 1 (20)	0(18)	In the unbound structures loop 106–110 reduces the pocket size
IL-2	1Z92.A	1PY2.A	FRH	IC_{50} = 60	0(23)	1(17), 2(14)	In 1Z92A, loop 30–35 protrudes into the site
Bcl-xL	3FDL.A	2YXJ.A	N3C	K_i = 0.5	0(23), 2 (15)	0(29), 1(15),	In 3FDLA, the helix 101–111 moves and opens the ligand-binding site
SARS-CoV main protease	1UK2.A 2GT7.A 3EA8.A	2GZ7.A	D3F	IC_{50} = 300	1(16), 3 (12) 0(23), 2 (16) 0(17), 1 (16)	0(21)	Loop 166–172 changes conformation. In addition, the side chains of H41, N189, and M165 protrude into the site
Maltodextrin/maltose-binding protein.	3PUW.E 5GS2.A 5GS2.A	1FQC.A	GLO	K_d = 200	0(19), 1 (16) 0(26) 2(16)	1(19)	Several loops (11–17, 142–154) are much closer to the ligand in the bound form. Conformational changes in the loop appear to be spontaneous
Thrombin	1HAG.E 1EOJ.A 1H8I.H	1GHY.H	121	K_i = 8	0(19), 2 (14) 0(27) 0(24)	0(22)	In the unbound structures, several loops are far from the pocket, making it too open
BACE-1 protease	1W50.A 1FKN.A 1XN2.C	3IXJ.C	586	IC_{50} = 0.32	1(17) 0(26) 0(16)	0(18), 1(18)	Several loops, primarily 71–74, move such that, in the unbound structure, the pocket is too open. The loops close down upon ligand binding
Uridine phosphorylase	1K3F.B 1ZL2.E 2OEC.C	1U1D.F	181	K_i = 353	0(27) 0(26) 0(17)	0(23), 1(18)	Active-site loop containing residues 225–230 acts as a lid over the pyrimidine-binding site. The loop is too open in unbound structures and moves inward in bound structures
Angiopoietin-1 receptor	1FVR.A	2OO8.X	RAJ	IC_{50} = 1	2(17), 3 (15)	0(16), 2(11),	In the unbound structure, loop 981–995 protrudes into the binding site and would clash with bound ligand

TABLE 2 (*CONTINUED*)

Name	Unbound[a]	Bound[b]	Ligand[c]	Affinity, nM[d]	FTMap[e]		Why is this site cryptic?
					Unbound	Bound	
Bovine beta-lactoglobulin	1BSQ.A 5IO5.A 1BSY.A	1GX8.A	RTL	K_d = 430	4(9) 0(24) 0(17)	2(12), 3(9)	Loop 84–90 in the unbound structure protrudes into the RTL-binding site
Acid-beta-glucosidase	3GXD.B 2NT1.D	2WCG.A	MT5	K_i = 420	7(5) 3(16)	1(17)	In the unbound structure 3GXDB, loop 342–356 moves outward, which makes the site too open. The loop is somewhat closer in the alternative unbound structure 2NT1D
TetR-like transcriptional regulator	2WGB.A 2V57.D	2 V57.A	PRL	K_d = 79	0(25) 0(19)	0(22), 1(19)	Loop 104–115 closes the pocket in the unbound structure
Coagulation factor VII zymogen (EGF2/ Protease)	1JBU.H 1FAK.H 1KLJ.H	1WUN.H	P5B	IC_{50} = 62	1(16) 0(30) 0(16)	0(24)	Loops 184–189 and 220–222 are well formed in the bound structure, but missing in the unbound structure 1JBUH
Glutamate receptor subunit 2.	1MY1.C 4U2R.C 1FTO.A	1FTL.A	DNQ	IC_{50} = 998	2(13) 0(38), 1 (22) 0(29), 1 (16)	0(27),1 (19)	Loops 139–143 and 68–73 protrude into the site, but move further away upon ligand binding
P38 MAP kinase	2ZB1.A 1YWR.A	2NPQ.A	BOG	K_d = 3080	N/A 0(23), 2 (16)	0(21), 6(6)	Lipid-binding pocket; helix 253–261 closes down the site in the unbound structure

[a] Protein Data Bank ID of the unbound structure in the CryptoSite database with additional unbound structures. The chain identifiers are added to the 4-letter PDB IDs.
[b] Protein Data Bank ID of the bound structure with chain identifier added to the 4-letter PDB ID.
[c] Three-letter code of the ligand in the bound structure.
[d] Affinity as reported.
[e] Number of probe clusters as determined by FTMap. For example, 0(18) means that the top hot spot includes 18 probe clusters.

(Kd < 300 nM) by any ligand. Despite the very weak potency in all cases, FTMap predicts a strong hot spot for six of the ten proteins considered in Table 1; thus traditional druggability analysis has limited value. Here, we formulate a hypothesis as to why cryptic sites formed by the motion of flexible side chains are likely to have only moderate ligand-binding capability.

Our starting point is to consider that surface side chains undergo conformational changes as a result of thermal motion at room temperature on the timescale of 10^{-11}–10^{-10} s. Therefore, the cryptic site fluctuates, with high frequency, between its open state, when the moving side chains are out of the pocket, and its closed state, when the side chains move back and interact with the rest of the residues in the pocket. Given that some of the side chains spend a substantial amount of time in the latter conformations interacting with residues that also bind the ligand when the moving side chains are out of the pocket, these side chains essentially act as competitive inhibitors of ligand binding.[(p41)] Although the motion of a flexible side chain is limited to well-defined rotations, in the closed state of the site it binds to the other residues in the pocket and, thus, competes with the ligand that would bind to the same residues. This model implies that the open receptor O participates in two reversible binding reactions. The first is the association of the open receptor O with the ligand L, forming the receptor–ligand complex OL with the binding constant $K_b = [OL]/([O][L])$, where the square bracket denotes concentration. The second reaction is the association between the open receptor O with a moving side chain S that interacts with the receptor, resulting in its closed state C. The total concentration $[R_{tot}]$ of the receptor is $[R_{tot}] = [O] + [OL] + [C]$. The fraction of the receptor in the open state (i.e., available for ligand binding) is $f = [O]/[R_{tot}]$. Since $[OL] = K_b[O][L]$, this fraction f is given by Equation (1):

$$f = [O]/[R_{tot}] = 1/(1 + K_b[L] + [C]/[O]) \quad (1)$$

where K_b is the maximum binding constant when the site is fully open, that is, at $[C] = 0$, whereas $[C]/[O]$ is the concentration of the receptor in a side chain-bound (i.e., closed) state divided by the concentration of the receptor in the open state (i.e., with the moving side chain out of the pocket). Given that, in a stationary state, the concentrations are proportional to probabilities, this ratio is the probability of the receptor being in the closed state divided by the probability of the receptor being in the open state. If the pocket is open 100% of the time, then $[C]/[O] = 0$. If no ligand is present, that is, $[L] = 0$, and the site is open, say, 90% of the time, then $[C]/[O] = 0.1/0.9 = 0.11$, and if the site is open only 10% of the time, which is the more realistic case, then $[C]/[O] = 0.9/0.1 = 9.0$, which is most likely close to the largest possible value of the $[C]/[O]$ term in Equation (1).

If the site is fully open all the time, $[C]/[O] = 0$ and Equation (1) describes the typical hyperbolic saturation curve with respect to the increasing ligand concentration. In this case, Equation (1) describes the fraction of the receptor binding sites available for ligand binding.[(p41)] To assess the impact of the inhibition by the moving side chain, that is, the impact of the $[C]/[O]$ term on the saturation curve, we have to consider the range of the $K_b[L]$ term. Assuming that the ligand concentration is in the micromolar range, $[L] = 10^{-6}\ M$, where M denotes mol/l, and a micromolar binding site, $K_b = 10^6/M$, then we have $K_b[L] = 1.0$. If the binding-site affinity is in the nanomolar range, that is, $K_b = 10^9$/M, then $K_b[L] = 1000$. Therefore, the $[C]/[O]$ term in Equation (1) has a noticeable impact on f only for binding sites with moderate affinity. Given that cryptic sites created exclusively by side chain motion are close to the surface and generally have moderate affinity even when fully open, we

expect $K_b[L] < 10$; thus, it is relatively small and, hence, comparable to the potential values of the $[C]/[O]$ term. Under these conditions, a side chain acting as a competitive inhibitor can substantially reduce the fraction f of the open receptor O available for ligand binding. In Figure 2a we consider $K_b = 10^6/M$ and ligand concentrations between 10^{-7} M and 10^{-5} M, and show the value of the fraction f at different $[C]/[O]$ values, thus at different ratios of the probabilities of the pocket being in the closed and open state. The concentration $[O']$ of the free receptor, bound to neither ligand nor the moving side chain, and, hence, available for ligand binding in the presence of inhibition by moving side chain(s), is given by Equation (2):

$$[O'] = [O](1 + K_b[L])/(1 + K_b[L] + [C]/[O]) \quad (2)$$

This implies that $[O'] < [O]$ if and only if the ratio $[C]/[O]$ of pocket probabilities in closed and open conformations is not negligible relative to the $K_b[L]$ term. Figure 2b shows the $[O']/[O]$ ratio at $K_b = 10^6/$ M and four $[C]/[O]$ values, demonstrating a substantial reduction in the relative concentration of the receptor available for ligand binding in the usual ligand concentration range when considering surface sites that have moderate binding affinity, even when disregarding the impact of the fast-moving side chains.

The above discussion implies that the probability of a cryptic site being in the open conformation affects the affinity of ligand binding at the site. Although this assumption does not appear to be controversial, the kinetics and thermodynamics of ligand binding are usually studied separately in the literature. However, kinetics and thermodynamics are connected by the fundamental relationship denoted by Equation (3):

$$K_b = e^{-\frac{\Delta G}{RT}} = k_{on}/k_{off} \quad (3)$$

where ΔG is the binding free energy, R is the universal gas constant, T is the temperature, and k_{on} and k_{off} are the kinetic on and off rates of ligand binding, respectively. Our assumption is that k_{on} depends on the probability of the pocket being in the

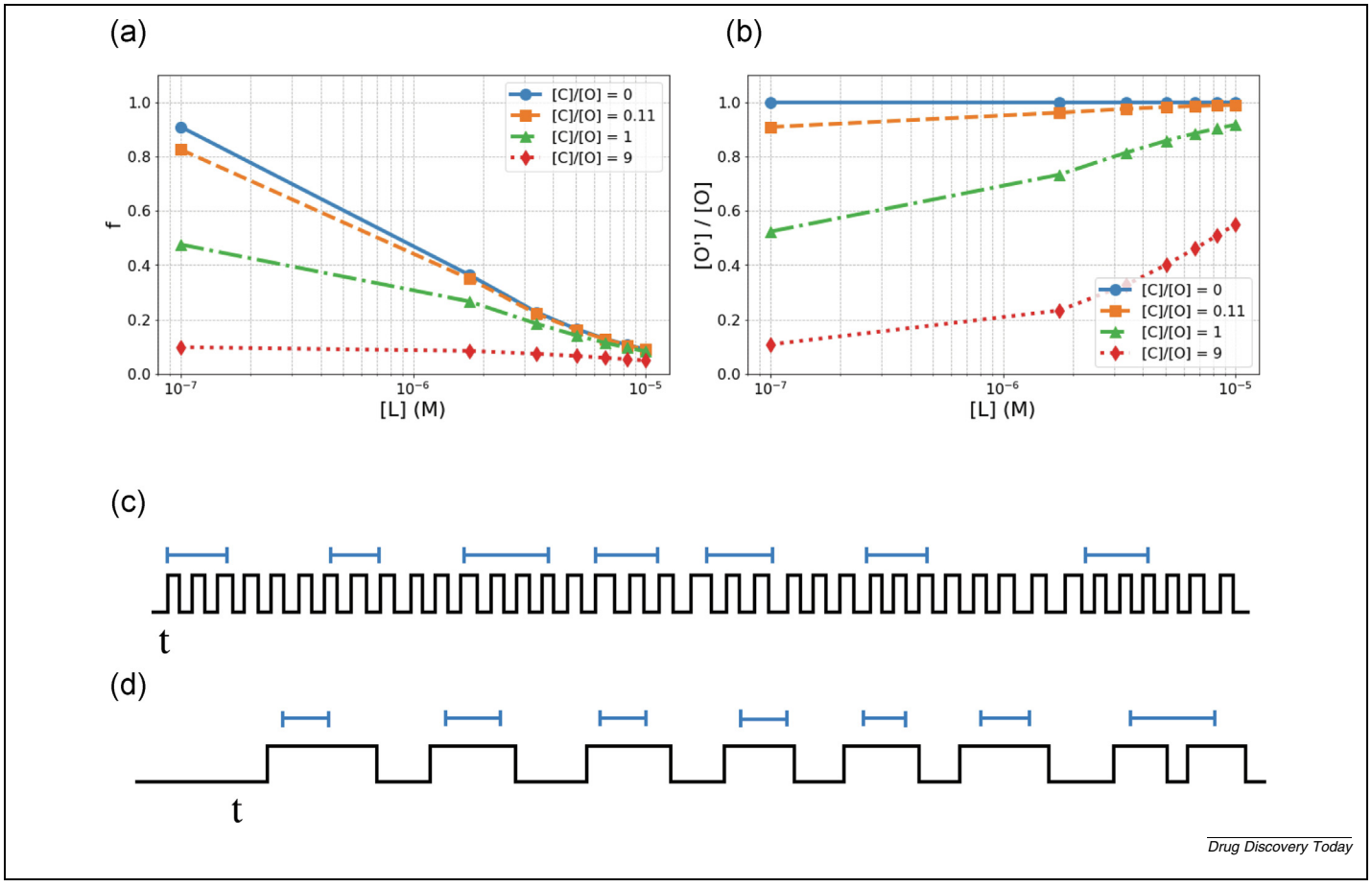

FIGURE 2

Impact of inhibition by flexible side chains and the timescales of side chain and loop motion. **(a)** Change in the $f = [O]/[R_{tot}]$ ratio at the binding constant $K_b = 10^6$/M and ligand concentrations between 10^{-7} M and 10^{-5} M for four different $[C]/[O]$ values from the fully open pocket ($[C]/[O] = 0$) to the case of a pocket open 10% of the time ($[C]/[O] = 9.0$). **(b)** Ratio $[O']/[O]$ of receptor concentrations with and without inhibition by moving side chains at different pocket-opening probabilities. The conditions are the same as in (a). **(c)** Schematic of the timescales of ligand binding (blue line) and the high frequency opening and closing of the cryptic site (i.e., upper and lower levels of the line) resulting from flexible side chains in the induced fit regime. **(d)** Schematic of the timescales of ligand binding (blue line) and the slower opening and closing of the cryptic site (i.e., upper and lower levels of the line) resulting from loop motion in the conformational selection regime.

open state available for binding, and, hence, without a change in k_{off}, the reduction of the probability reduces the binding constant K_b. The same conclusion was reported in a recent study, which observed that the potency of blebbistatin binding to the motor domain of different myosin isoforms is correlated with the probability of cryptic site opening, determined by MD simulations and a pocket volume detection algorithm.[p54] In another relevant study, the F143W mutation was introduced into BcL-xL. Although the mutated residue did not interact with any ligand, it allosterically stabilized the cryptic peptide-binding site in the open state, and improved its affinity for the known ligands.[p55]

Why cryptic sites formed by loop motion can have higher affinity

In the previous section, we showed that the low probability of pocket opening can reduce the fraction of receptor sites available for ligand binding and, hence, the binding affinity of cryptic sites formed by the high frequency motion of surface side chains. The question is why similar arguments do not apply to the proteins shown in Table 2 that have cryptic sites formed by loop or secondary structure motion. To begin with, such binding pockets can be deeper and generally have higher affinity, and, according to our analysis in the previous section, because of the much larger $K_b[L]$ term in Equation (2), the impact of the $[C]/[O]$ term becomes less significant. However, the main difference results from the different kinetic mechanism of ligand binding. The high-frequency side chain motion, on the timescale of 10^{-11}–10^{-10} s, implies that the side chains move in and out of the binding site; thus, they collide and compete with the ligand, indicating that the induced fit model of molecular recognition applies.[p56] In fact, ligand binding, from forming an encounter complex to the final bound conformation, is relatively slow, on the timescale of 10^{-8}–10^{-1} s (Figure 2C). During this period, the ligand collides and competes with the moving side chains many times. By contrast, the timescale of loop motion, 10^{-9}–10^{-6} s, partially overlaps with the scale of ligand binding, and the motion of secondary structure elements is even slower. This implies that the pocket is open for relatively long periods of times that are sufficient for binding, involving neither collisions nor competition with the residues in the binding sites (Figure 2d). Thus, the binding follows the conformational selection rather than the induced fit model. This change in the timescale implies that the competitive inhibition by the moving side chains becomes less significant.

The impact of the timescales of conformational variations on binding kinetics has been studied by Zhou by modeling the biophysical mechanism of ligand recognition.[p57] We think that the results of his paper are very relevant to our discussion and, hence, directly quote from its abstract:

'In receptor-ligand binding, a question that generated considerable interest is whether the mechanism is induced fit or conformational selection. This question is addressed here by a solvable model, in which a receptor undergoes transitions between active and inactive forms. The inactive form is favored while unbound but the active form is favored while a ligand is loosely bound. As the active-inactive transition rates increase, the binding mechanism gradually shifts from conformational selection to induced fit. The timescale of conformational transitions thus plays a crucial role in controlling binding mechanisms.'[p57]

With application to cryptic sites, the active and inactive forms are the open and closed conformations of the pocket, respectively and, with the site being cryptic, the closed state is favored while unbound. Thus, the conclusion reached by Zhou clearly applies to the problem considered here, and explains why the competitive inhibition model becomes less relevant at the slower rate of conformational transitions. In fact, the slower motion of the loops implies a shift toward the conformational selection regime, in which the cryptic pocket is open for longer time intervals, enabling ligand binding without collisions and, thus, without competition by the side chains (Figure 2d).

Concluding remarks

Our goal here is to explain the observation that cryptic sites formed exclusively by side chain opening and possibly small movements of short backbone segments provide limited binding affinity. The observation is important, because MD, the main computational tool to find cryptic sites, tends to open such pockets first. This is definitely the case in relatively short simulations, because the movement of longer loops and secondary structure elements occur on much longer timescales and require substantially longer simulations. It is intuitively clear that the probability of pocket opening impacts the on-rate of ligand binding and, because of the fundamental relationship connecting kinetics and thermodynamics (Equation (3), affects the ligand-binding affinity. Such impact has been observed in the literature.[p54],[p55] The main argument we use here is that a fast-moving side chain that spends a substantial fraction of time in the pocket interacting with the other residues competes with ligands for binding and, hence, acts as a competitive inhibitor. This inhibition occurs only in cryptic sites with moderate binding affinity for the ligand, because strong ligand binding makes the impact of the competition negligible. However, the examples presented here from the CryptoSite set show that such inhibition is real and significant, even for sites that are predicted to have a strong binding hot spot. Previously published kinetics arguments imply that such inhibition occurs in the induced fit regime of ligand binding because of high-frequency side chain motion, and it is less relevant in the conformational selection regime of the much slower loop motion.[p57]

There are two caveats that limit the applicability of our results. First, ligand binding by covalent interactions is irreversible, and is beyond the scope of this paper. Thus, our results might not apply to the sites identified by chemoproteomics. The second exception is the class of cryptic sites that directly extend a well-defined binding pocket that is already open in ligand-free structures. In such cases, ligands that bind to the 'permanent' pocket extend toward the cryptic region and force it open.[p58] Many small molecule-binding sites in protein–protein interfaces represent this type cryptic sites. For example, the inhibitor binding site of interleukin 2 (IL-2) in the interface with IL-2Rα includes a largely polar and rigid pocket, and a highly adaptive hydrophobic region that forms a pocket only upon inhibitor binding.[p59] A similar binding mechanism describes imatinib binding to the Abl tyrosine kinase. The inhibitor binds both to

the preformed ATP pocket and an adjacent allosteric pocket, which is induced by the ligand. In such cases, it might not matter whether the cryptic pocket opening involves side chain or loop motion, because the permanent part of the pocket can already provide sufficient binding free energy.

CRediT authorship contribution statement

M.L.: visualization, investigation. **D.K.**: review and editing. **D.J-M.**: review and editing. **S.V.**: writing – original draft, visualization, conceptualization, validation, resources.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by grants R35GM118078 and RM1135136 from the National Institute of General Medical Sciences.

Declaration of Interests

The authors declare no competing interest.

References

1. Cimermancic P et al. CryptoSite: expanding the druggable proteome by characterization and prediction of cryptic binding sites. *J Mol Biol*. 2016;428:709–719.
2. Abi Hussein H, Geneix C, Petitjean M, Borrel A, Flatters D, Camproux AC. Global vision of druggability issues: applications and perspectives. *Drug Discov Today*. 2017;22:404–415.
3. Kozakov D, Hall DR, Napoleon RL, Yueh C, Whitty A, Vajda S. New frontiers in druggability. *J Med Chem*. 2015;58:9063–9088.
4. Henley MJ, Koehler AN. Advances in targeting 'undruggable' transcription factors with small molecules. *Nat Rev Drug Discov*. 2021;20:669–688.
5. Wakefield AE, Kozakov D, Vajda S. Mapping the binding sites of challenging drug targets. *Curr Opin Struct Biol*. 2022;75, 102396.
6. Wiesmann C et al. Allosteric inhibition of protein tyrosine phosphatase 1B. *Nat Struct Mol Biol*. 2004;11:730–737.
7. Barr AJ. Protein tyrosine phosphatases as drug targets: strategies and challenges of inhibitor development. *Future Med Chem*. 2010;2:1563–1576.
8. Keedy DA et al. An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. *eLife*. 2018;7, e36307.
9. Perot S, Sperandio O, Miteva MA, Camproux AC, Villoutreix BO. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov Today*. 2010;15:656–667.
10. del Sol A, Tsai CJ, Ma B, Nussinov R. The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure*. 2009;17:1042–1050.
11. Nussinov R, Ma BY, Tsai CJ. Multiple conformational selection and induced fit events take place in allosteric propagation. *Biophys Chem*. 2014;186:22–30.
12. Spradlin JN, Zhang E, Nomura DK. Reimagining druggability using chemoproteomic platforms. *Acc Chem Res*. 2021;54:1801–1813.
13. Horn JR, Shoichet BK. Allosteric inhibition through core disruption. *J Mol Biol*. 2004;336:1283–1291.
14. Vajda S, Beglov D, Wakefield AE, Egbert M, Whitty A. Cryptic binding sites on proteins: definition, detection, and druggability. *Curr Opin Chem Biol*. 2018;44:1–8.
15. Sun Z, Wakefield AE, Kolossvary I, Beglov D, Vajda S. Structure-based analysis of cryptic-site opening. *Structure*. 2020;28:223–235.
16. Beglov D et al. Exploring the structural origins of cryptic sites on proteins. *Proc Natl Acad Sci U S A*. 2018;115:E3416–E3425.
17. Hardy JA, Wells JA. Searching for new allosteric sites in enzymes. *Curr Opin Struct Biol*. 2004;14:706–715.
18. Ludlow RF, Verdonk ML, Saini HK, Tickle IJ, Jhoti H. Detection of secondary binding sites in proteins using fragment screening. *Proc Natl Acad Sci U S A*. 2015;112:15910–15915.
19. Erlanson DA, Wells JA, Braisted AC. Tethering: fragment-based drug discovery. *Annu Rev Biophys Biomol Struct*. 2004;33:199–223.
20. Lawson AD. Antibody-enabled small-molecule drug discovery. *Nat Rev Drug Disc*. 2012;11:519–525.
21. Bowman GR, Bolin ER, Hart KM, Maguire BC, Marqusee S. Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proc Natl Acad Sci U S A*. 2015;112:2734–2739.
22. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys*. 2004;120:11919–11929.
23. Oleinikovas V, Saladino G, Cossins BP, Gervasio FL. Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *J Am Chem Soc*. 2016;138:14257–14263.
24. Bowman GR, Geissler PL. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc Natl Acad Sci U S A*. 2012;109:11681–11686.
25. Sztain T, Amaro R, McCammon JA. Elucidation of cryptic and allosteric pockets within the SARS-CoV-2 main protease. *J Chem Inf Model*. 2021;61:3495–3501.
26. Ghanakota P, Carlson HA. Driving structure-based drug discovery through cosolvent molecular dynamics. *J Med Chem*. 2016;59:10383–10399.
27. Smith RD, Carlson HA. Identification of cryptic binding sites using MixMD with standard and accelerated molecular dynamics. *J Chem Inf Model*. 2021;61:1287–1299.
28. Martinez-Rosell G, Lovera S, Sands ZA, De Fabritiis G. PlayMolecule CrypticScout: predicting protein cryptic sites using mixed-solvent molecular simulations. *J Chem Inf Model*. 2020;60:2314–2324.
29. Lal Gupta P, Carlson HA. Cosolvent simulations with fragment-bound proteins identify hot spots to direct lead growth. *J Chem Theory Comput*. 2022;18:3829–3844.
30. Kalenkiewicz A, Grant BJ, Yang CY. Enrichment of druggable conformations from apo protein structures using cosolvent-accelerated molecular dynamics. *Biology*. 2015;4:344–366.
31. Meller A, Bhakat S, Solieva S, Bowman GR. Accelerating cryptic pocket discovery using AlphaFold. *J Chem Theory Comput*. 2023;19:4355–4363.
32. Meller A et al. Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network. *Nat Commun*. 2023;14:1177.
33. Rehman AU et al. Hidden allosteric sites and de-novo drug design. *Expert Opin Drug Discov*. 2022;17:283–295.
34. Lu S, Ji M, Ni D, Zhang J. Discovery of hidden allosteric sites as novel targets for allosteric drug design. *Drug Discov Today*. 2018;23:359–365.
35. Eyrisch S, Helms V. Transient pockets on protein surfaces involved in protein-protein interaction. *J Med Chem*. 2007;50:3457–3464.
36. 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.
37. Surade S, Blundell TL. Structural biology and drug discovery of difficult targets: the limits of ligandability. *Chem Biol*. 2012;19:42–50.
38. Edfeldt FNB, Folmer RHA, Breeze AL. Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discov Today*. 2011;16:284–287.
39. 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.
40. Gleeson MP, Hersey A, Montanari D, Overington J. Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat Rev Drug Discov*. 2011;10:197–208.
41. Palmer T, Bonner PL. *Enzymes: biochemistry, biotechnology, clinical chemistry*. Amsterdam: Elsevier; 2007.
42. Hajduk PJ, Huth JR, Tse C. Predicting protein druggability. *Drug Discov Today*. 2005;10:1675–1682.

43. Hajduk PJ, Huth JR, Fesik SW. Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem*. 2005;48:2518–2525.
44. Mattos C, Ringe D. Locating and characterizing binding sites on proteins. *Nat Biotechnol*. 1996;14:595–599.
45. Allen KN et al. An experimental approach to mapping the binding surfaces of crystalline proteins. *J Phys Chem*. 1996;100:2605–2611.
46. Cheng AC et al. Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol*. 2007;25:71–75.
47. Halgren TA. Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model*. 2009;49:377–389.
48. Krasowski A, Muthas D, Sarkar A, Schmitt S, Brenk R. DrugPred: a structure-based approach to predict protein druggability developed using an extensive nonredundant data set. *J Chem Inf Model*. 2011;51:2829–2842.
49. 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.
50. Seco J, Luque FJ, Barril X. Binding site detection and druggability index from first principles. *J Med Chem*. 2009;52:2363–2371.
51. Bekar-Cesaretli AA, Khan O, Nguyen T, Kozakov D, Joseph-Mccarthy D, Vajda S. Conservation of hot spots and ligand binding sites in protein models by AlphaFold2. *J Chem Inf Model*. 2024;64:960–973.
52. Bakan A, Nevins N, Lakdawala AS, Bahar I. Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules. *J Chem Theory Comput*. 2012;8:2435–2447.
53. Grove LE, Hall DR, Beglov D, Vajda S, Kozakov D. FTFlex: accounting for binding site flexibility to improve fragment-based identification of druggable hot spots. *Bioinformatics*. 2013;29:1218–1219.
54. Meller A et al. Drug specificity and affinity are encoded in the probability of cryptic pocket opening in myosin motor domains. *eLife*. 2023;12, e83602.
55. Mizukoshi Y et al. Targeting the cryptic sites: NMR-based strategy to improve protein druggability by controlling the conformational equilibrium. *Sci Adv*. 2020;6, eabd0480.
56. Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol*. 2009;5:789–796.
57. Zhou HX. From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys J*. 2010;98:L15–L17.
58. Shan Y, Mysore VP, Leffler AE, Kim ET, Sagawa S, Shaw DE. How does a small molecule bind at a cryptic binding site? *PLoS Comput Biol*. 2022;18, e1009817.
59. Braisted AC et al. Discovery of a potent small molecule IL-2 inhibitor through fragment assembly. *J Am Chem Soc*. 2003;125:3714–3715.