ELSEVIER

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Cryptic binding sites on proteins: definition, detection, and druggability

Sandor Vajda[1,2], Dmitri Beglov[1], Amanda E Wakefield[1,2], Megan Egbert[1] and Adrian Whitty[2]

Many proteins in their unbound structures lack surface pockets appropriately sized for drug binding. Hence, a variety of experimental and computational tools have been developed for the identification of cryptic sites that are not evident in the unbound protein but form upon ligand binding, and can provide tractable drug target sites. The goal of this review is to discuss the definition, detection, and druggability of such sites, and their potential value for drug discovery. Novel methods based on molecular dynamics simulations are particularly promising and yield a large number of transient pockets, but it has been shown that only a minority of such sites are generally capable of binding ligands with substantial affinity. Based on recent studies, current methodology can be improved by combining molecular dynamics with fragment docking and machine learning approaches.

Addresses
[1] Department of Biomedical Engineering, Boston University, Boston, MA 02215, United States
[2] Department of Chemistry, Boston University, Boston, MA 02215, United States

Corresponding authors: Vajda, Sandor (vajda@bu.edu), Whitty, Adrian (awhitty@bu.edu)

Introduction

Many proteins have small-molecule binding pockets that are not easily detectable in the ligand-free structures. These cryptic sites require a conformational change to become apparent. A cryptic site can therefore be defined as a site that forms a pocket in a ligand-bound structure, but not in the unbound protein structure [1••]. It has long been well-known that proteins are dynamic objects, and that their binding sites may change conformation upon ligand binding [2]. However, finding and utilizing cryptic

or hidden binding sites has received growing attention during the last few years [1••,3•,4,5••,6,7,8,9,10], seemingly motivated by two factors. First, many biologically relevant drug targets lack appropriately sized pockets in their unbound structures to support the strong binding of drug-sized ligands [7,11]. It has been suggested that cryptic sites can provide previously undescribed pockets, potentially enabling targeting of proteins that would otherwise be considered undruggable [7], and thereby expanding the 'druggable genome' [11]. In keeping with this idea, some of the pockets that bind small molecule inhibitors of protein–proteins interactions, a class that include many such challenging targets, are cryptic [12]. Cryptic sites located away from the main functional site of a protein, but which can modulate the activity of the protein allosterically, are also potentially useful [13•], particularly if the main functional site cannot be targeted with sufficient specificity [14]. Targeting a distal site also has the potential to give a different pharmacological profile [15]. The second factor contributing to the increased interest in cryptic sites is the availability of improved methodology for identifying such sites, particularly molecular dynamics and Markov state simulation methods that are now computationally feasible [3•,8,9,16,17,18,19]. More specialized computational tools have also been developed that integrate molecular dynamics simulations [7,9] with fragment docking [5••] and machine learning approaches [1••]. Despite this high level of recent interest in cryptic sites, a review of recent publications suggests that a number of questions concerning the definition, identification, and druggability of cryptic sites, and their potential value for drug discovery, are either not fully answered, or have been answered in conflicting ways. Here we identify and discuss some of these questions, with emphasis on problem areas that need further work.

When is a binding site cryptic?

Intuitively, a binding site is cryptic if it can be identified in the ligand-bound but not in the unbound structure of a protein. This definition is far from rigorous, however, since it depends both on the method of searching for the sites, and on the particular unbound structure(s) considered. To develop a benchmark set of proteins with cryptic sites, Cimermancic *et al.* [1••] screened over 20,000 unbound–bound protein pairs from the Protein Data Bank using two pocket detection algorithms, Fpocket [20] and ConCavity [21]. Both algorithms derive

scores that reflect the putative capacity of pockets to bind small molecules. They averaged Fpocket's residue druggability scores and ConCavity's residue scores over residues within 5 Å of the ligand, to form a pocket score with values between zero (no pocket) and 1 (well-formed binding site). This composite measure primarily depends on the volume of the pocket, but also includes other factors such as residue polarity and evolutionary conservation. Cryptic sites were defined as sites with an average pocket score of less than 0.1 in the unbound form of the protein and greater than 0.4 in the bound form. Using these criteria, together with manual inspection, Cimermancic et al. [1••] selected 93 pairs in which each unbound structure had a site considered cryptic due to its low pocket score, and each bound structure had a functionally relevant ligand bound at the site. The resulting CryptoSite set is very useful for testing cryptic site prediction algorithms, and the definition of a cryptic site that these authors developed could provide the basis for a community standard as other groups test newly developed methods.

A potential issue with the approach of Cimermancic et al. [1••] is that, to determine whether a site can correctly be considered cryptic, it is important to consider the full range of conformations available to the protein in the absence of ligand. Basing the structural comparison on a single unbound structure ignores the ensemble of conformations available to the unbound protein, especially important at mobile regions such as potential cryptic sites [16]. This consideration raises the following question: Can a site be properly considered as cryptic if it is absent in just one or a very few unbound structures, even if it is fully formed in other unbound structures? Such behavior suggests that, although the protein can adopt conformations in which the pocket is absent, it also has accessible conformations in which the pocket is present. In some cases these bound-like conformations may even represent the most abundant state of the protein in the absence of ligand. An alternative, more stringent definition is for a pocket to be considered cryptic only if it is absent in all, or nearly all, unbound structures of the protein, such that it cannot be reliably identified in the absence of a bound ligand, and likely does not exist in any large fraction of the conformational states available to the unbound protein.

Beglov et al. [22••] have investigated how broader consideration of the conformations available to the unbound protein would affect the set of cryptic sites identified by Cimermancic et al. [1••]. To each protein pair in the CryptoSite set, they added all unbound structures in the Protein Data Bank having at least 95% sequence identity. The number of such additional unbound structures varied from zero to 498 per protein, resulting in an extended CryptoSite dataset that included 4950 structures rather than the original 186. Inclusion of these additional unbound structures revealed that bound-like pockets are at least partially formed in some unbound structures for close to 50% of the 93 proteins in the CryptoSite set [22••]. For example, in the original CryptoSite set the protease beta-secretase 1 (BACE-1) is represented by unbound and bound structures 1W50 and 3IXJ. The unbound structure 1W50 has a low Fpocket druggability score because the loop comprising residues 71–74 is far from the active site, making the pocket too open to score as druggable (Figure 1a). The loop is closed down on the inhibitor in the bound structure 3IXJ, resulting in a well-formed pocket that binds the isophthalamide ligand with high affinity. The analysis of 52 structures of unbound BACE-1 in the extended CryptoSite set reveals that, in these structures, the pockets in question are almost evenly distributed between conformations resembling the unbound and bound forms, with druggability scores varying between 0.2 and 0.6 (Figure 1c). Thus, it is arguable whether this site should be considered as cryptic.

In many of the 93 proteins the analysis of the structures in the extended CryptoSite set revealed some degree of spontaneous shift toward the ligand-bound conformations at the binding site, but with the distribution of observed conformations heavily weighted toward the unbound state [22••]. For example, the original CryptoSite set includes 2GFC and 2JDS as an unbound–bound pair of structures for the cAMP-dependent protein kinase known as Protein Kinase A (PKA). In the unbound structure, 2GFC, the activation loop (which has the sequence SFG rather than the DFG segment seen in many kinases) protrudes into the active site, closing the pocket (Figure 1b). The site opens when an inhibitor binds (PDB ID 2JDS, see Figure 1b). With few exceptions, in the unbound structures the SFG loop resides in the partially hydrophobic outer region of the kinase active site (Figure 1c). Although the pocket is fully formed in a few structures, this is because these structures also contain bound allosteric modulators far from the active site, for example, myristolic acid in the structure 4DFZ. Thus, it is probably reasonable to consider this PKA pocket cryptic.

Proteins in which the cryptic site is completely missing in all unbound structures seem to be rare. A classic example of this type is TEM β-lactamase, in which an elongated cryptic site was discovered serendipitously when crystals revealed two small molecules from the crystallization buffer bound between helices 11 and 12 [23]. In the bound structure (PDB code 1PZO) the position of the shorter helix shifts, opening a substantial crevice (Figure 1d). However, no opening of the site is seen in any unbound structure. Engineered variants of the protein exist in which helix 11 is unfolded, resulting in partial pocket opening [24], but in the wild-type enzyme the site is definitely cryptic.

The conclusion as to whether a site is cryptic may also depend on the resolution of the available structures. An example is protein tyrosine phosphatase 1B (PTP1B),

Figure 1
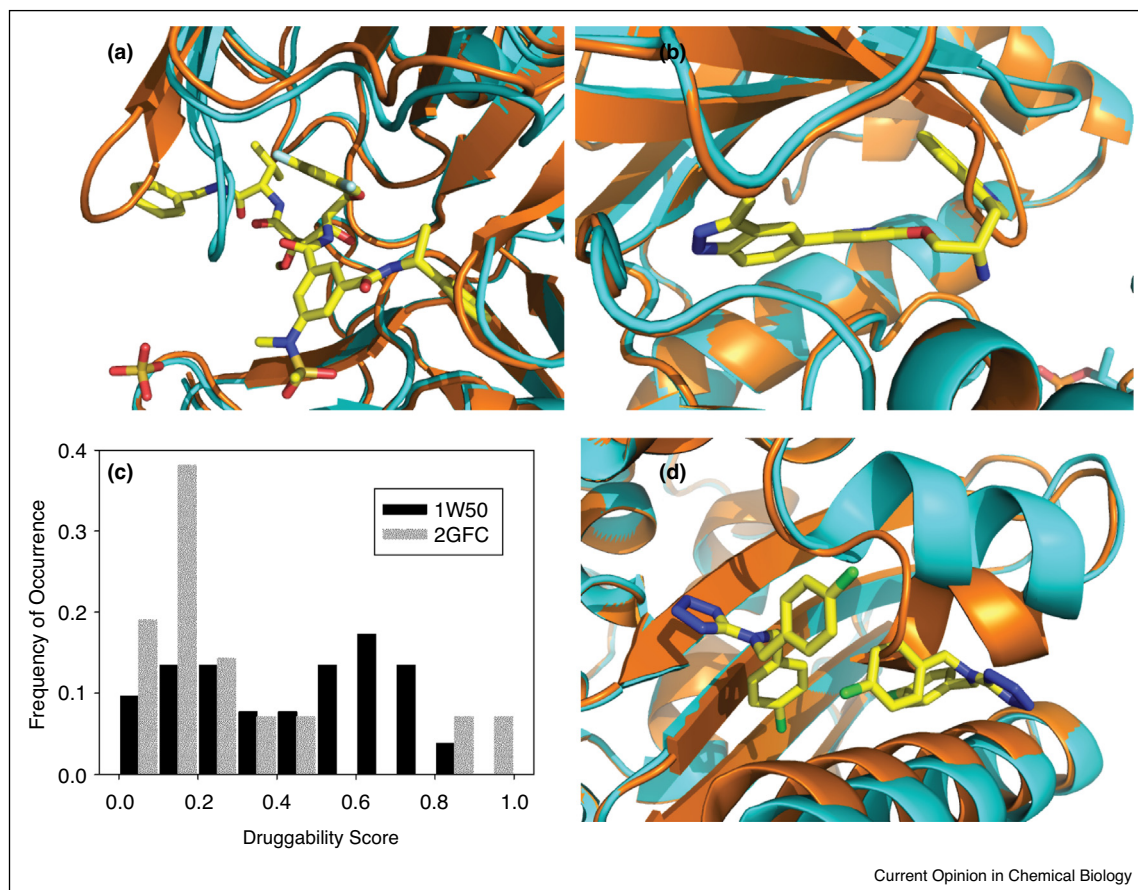
Types of cryptic sites. In all cases the bound structure is colored orange, the bound ligand is shown as yellow sticks, and the unbound structure superimposed over the bound one is colored cyan. PDB codes are shown in parenthesis. **(a)** Unbound (1W50) and ligand-bound (3IXJ) structures of beta-secretase 1 (BACE-1) protease, the latter co-crystallized with a isophthalamide inhibitor. In the unbound structure the binding pocket is too open, and the loop is closing down on the ligand upon binding. **(b)** Unbound (2GFC) and ligand-bound (2JDS) structures of cAMP-dependent protein kinase, the latter co-crystallized with the ATP-competitive inhibitor A-443654. In the unbound structure the activation loop protrudes into the binding site and would clash with the inhibitor superimposed from the bound structure. **(c)** Distributions of druggability scores in the unliganded structures of BACE-1 protease (homologs of 1W50) and cAMP-dependent protein kinase (homologs of 2GFC). Based on the druggability score, in the BACE-1 protease the pockets are almost evenly distributed between conformations resembling the unbound and bound forms. In contrast, in the cAMP-dependent protein kinase unbound-like conformations dominate. **(d)** Unbound (1JWP) and ligand-bound (1PZO) structures of TEM β-lactamase. The bound structure 1PZO includes two small inhibitor molecules bound between helices 11 and 12.

which has a cryptic allosteric site close to its C-terminus [25,26]. The unbound structures 2F6V and 1SUG of PTP1B, with resolutions of 1.7 Å and 1.95 Å, respectively, have a well-resolved C-terminal helix that covers the allosteric site, rendering it undetectable by any current pocket detection algorithm. However, the C-terminal region is very flexible; in fact, the helix is missing in several low resolution unbound structures such as 2HNP and 1T49, and is poorly resolved in 1T48 [25]. Absent the helix, the allosteric site appears accessible to ligands [27]. These problems demonstrate that the concept of cryptic sites may be difficult to rigorously define unless criteria

for selecting the ensemble of protein structures to be considered are first specified.

Cryptic sites for drug discovery

An important practical question concerning cryptic sites is their usefulness for drug discovery, frequently described in terms of druggability, a concept used with a number of different meanings [28•]. We follow the definition of Hajduk *et al.* [29], and define a protein as druggable if it is capable of binding a ligand with affinity high enough for the modulation of biological activity. In addition to conventionally drug-like molecules that comply with Lipinski's Rule-of-Five (Ro5) [30,31], the ligands considered here

include macrocycles, peptidomimetics, and other compounds with molecular weights exceeding 500 Da, but which nonetheless have some potential for good drug-like properties [32–34]. Such beyond-Ro5 compounds frequently target protein sites that are large, highly lipophilic, or highly polar, flexible, flat, or featureless [34]. While these proteins may be difficult to drug, considering them as undruggable would lead to missing potentially important targets. We note that our definition does not consider the biological significance of the target, which must be established independently of its druggability [28•].

Kozakov *et al.* [35] developed druggabilty conditions based on the FTMap computational method [36]. FTMap distributes small organic probe molecules of different sizes, shapes, and polarities on the surface of the protein to be studied, finds the most favorable positions for each probe type, clusters the probes, and ranks the clusters on the basis of their average energy. Regions that bind several different probe clusters are called consensus sites, and are the predicted binding energy hot spots [36]. FTMap is available through a free server [36]. It was shown by extensive benchmarking against experimental data that a binding site is potentially druggable if it harbors a binding energy hot spot that is strong enough to comprise at least 16 FTMap probe clusters [35]. To evaluate the druggability of the cryptic sites in the CryptoSite set, Beglov *et al.*, applied FTMap to all 4857 unbound and 93 bound structures in the extended set [22••]. According to the results, for 81% of the proteins at least one of the unbound structures has hot spots with 16 or more probe clusters within 5 Å of the cryptic site, and thus the sites are predicted to be druggable even when the protein is in this unbound conformation. For example, the cryptic sites in the unbound structures 1W50 of BACE-1 and 2GFC of PKA have hot spots containing 17 and 18 probe clusters, respectively. However, we note that the condition that binding 16 probe clusters implies druggability was based on the analysis of traditional drug targets [35], and may over-predict druggability when applied to cryptic sites [22••]. In particular, it was found that cryptic sites that are opened by movements of side chains only, generally have low affinities – in the micromolar range for the best reported ligands – even when FTMap shows the site to contain a hot spot comprising more than 16 probe clusters [22••].

While further work is required to establish reliable druggability criteria for cryptic sites, many such sites are definitely druggable, having been shown to bind druglike ligands with high affinity. In Table 1 we list proteins from the CryptoSite set [1••] that have known ligands that bind at the site with affinities less than 300 nM, as demonstrated by low K_D, K_i, or IC_{50} values, depending on what measure was

Table 1

Proteins in the CryptoSite set with validated high affinity cryptic sites.

Name	Unbound[a]	Bound[b]	Lig.[c]	Affinity, nM[d]	R-L distance Bound[e]	R-L distance Unb.[f]	Type[g]
mRNA-decapping enzyme DcpS	3BL9B	3BL7A	DD1	$IC_{50} = 7.6$	2.68	0.57	Allo
Hepatitis C virus polymerase	3CJ0A	2BRLA	POO	$IC_{50} = 18$	2.57	0.89	Allo
Hepatitis C virus polymerase	3CJ0A	3FQKB	79Z	$IC_{50} = 81$	2.63	1.15	Allo
Tyrosine kinase domain of C-MET	1R1WA	3F82A	353	$IC_{50} = 4.6$	2.81	0.62	Allo
TetR-like transcriptional regulator	2WGBA	2V57A	PRL	$K_D = 79$	2.84	0.48	Allo
Angiopoietin-1 receptor	1FVRA	2OO8X	RAJ	$IC_{50} = 1$	2.93	0.45	Allo
Nicotinic acetylcholine receptor	3PEOG	2BYSJ	LOB	$K_D = 0.3$	2.41	2.04	Allo
Biotin carboxylase	1BNCB	2V5AA	LZL	$IC_{50} = 150$	2.73	2.55	Allo
Staphylococcal nuclease	1TQOA	1TR5A	THP	$K_i \sim 100$	2.84	2.33	Ortho
DXP reductoisomerase	1K5HC	2EGHB	FOM	$K_i = 38$	2.49	1.69	Ortho
Glutamate Racemase (GluR)	2OHGA	2OHVA	NHL	$K_i = 16$	2.46	0.70	Ortho
SARS-CoV main protease	1UK2A	2GZ7A	D3F	$IC_{50} = 300$	2.71	1.56	Ortho
Serotonin N-acetyltransferase	1B6BA	1KUVA	CA5	$K_i = 22$	2.47	0.64	Ortho
Coagulation factor VII zymogen	1JBUH	1WUNH	5B	$IC_{50} = 62$	2.74	2.24	Ortho
NPC2 lysosomal protein	1NEPA	2HKAC	C3S	$K_D = 30–50$	3.47	0.59	Ortho
Hsp90 (heat shock protein 90)	2QFOB	2WI7A	2 K L	$IC_{50} = 58$	2.83	2.23	Ortho
Integrin alpha-L	3F74C	3BQMC	BQM	$IC_{50} = 2$	2.75	1.94	Ortho
Interleukin-2	1Z92A	1PY2A	FRH	$IC_{50} = 60$	2.81	1.39	PPI
Bcl-xL	3FDLA	2YXJA	N3C	$K_i = 0.5$	3.07	2.27	PPI

[a] Protein Data Bank ID of the unbound structure with chain identifier added to the 4-letter PDB ID.
[b] Protein Data Bank ID of the bound structure with chain identifier added to the 4-letter PDB ID.
[c] 3-letter code of the ligand in the bound structure.
[d] Affinity as reported.
[e] Shortest distance between any atom of the protein and any atom of the ligand in the bound structure.
[f] The maximum of the shortest distances between any atom of the protein and any atom of the ligand superimposed from the bound structure, maximized over all unbound structures. If this distance is less than the distance in the previous column, then the ligand would clash with the protein in all unbound structures, thus the pocket is not formed without ligand binding.
[g] Type of binding site: Allo – allosteric, ortho – orthosteric (primary binding), PPI – protein–protein interaction inhibitor.

reported. All of these proteins can therefore reasonably be considered as druggable. To assure that the analysis captured only sites that are genuinely cryptic even when all their unbound structures are considered, the proteins were subjected to a further test to ensure that all available unbound structures would clash with the ligand, and thus the pocket is fully formed only upon ligand binding (this condition eliminated both BACE-1 and PKA; indeed, for both of these proteins all unbound structures would have some clashes with the superimposed ligands).

Eight of the sites shown in Table 1 are allosteric. Since molecular mechanisms of allosteric communication are rooted in the dynamic nature of proteins, allosteric modulators frequently bind at flexible regions without pre-formed pockets [14,19]. Allosteric sites are becoming increasingly important for drug discovery [3•,9,14,19,37,38], and hence potential allostery has been a prime motivation for the analysis of cryptic sites. Table 1 also shows nine orthosteric (primary binding) sites. Five of these are enzyme active sites. In most single-chain enzymes the active site is a large cleft already present in the unbound structure [39]. However, in some cases enzyme function requires high level of flexibility, and in unbound structures the site may not be fully open. For example, staphylococcal nuclease performs relatively nonspecific cleavage of nucleic acids, and thus must be able to adapt to different substrates. Ligand binding to DXP reductoisomerase leads to domain motion and the closing of a flexible loop serving as a lid that shields the contents of active site from bulk solvent [40]. It was shown that enzymes with lid-gated active sites must operate by an induced fit mechanism [41]. Glutamate racemase is a moonlighting enzyme that serves two distinct functions [42], and the SARS-CoV main protease cleaves two different peptides. Serotonin N-acetyltransferase catalyzes acetyl transfer to a diverse array of tryptamine analogs. Thus, the ability to bind several substrates or to exclude bulk water from the active site seem to require substantial flexibility that can render the site closed in the unbound structure.

The last two items in Table 1 represent binding sites for small molecule inhibitors of protein–protein interactions, a target class of increasing importance. Such sites on the protein target frequently bind a peptide fragment of the protein partner, and hence are open to a certain degree [43]. However, enabling a small molecule to compete with the protein for binding may require expansion of the pocket to regions that can be considered cryptic. For example, interleukin-2 is well known to have a cryptic site [5••,7]. The unbound structure already has a shallow nonpolar pocket in the region that interacts with the α chain of the interleukin-2 receptor [44], and the pocket can bind small molecules with moderate affinity [45]. However, a polar extension of the pocket opens only when a longer and much higher affinity inhibitor binds [46]. A similar case is seen in KRAS, which has a small pocket in its SOS-binding interface [47], capable of

binding weak inhibitors. However, the binding of slightly higher affinity inhibitors opens a second pocket nearby [48]. The opening of additional sites by larger ligands is not limited to protein–protein interfaces. For example, in the protein methyltransferase DOT1L the expansion of inhibitor structure reveals additional binding sites that were not visible even in the SAM/SAH bound structures [49]. Thus, predicting the potential expansion of existing pockets by the binding of larger ligands is a very interesting but largely unsolved problem, and leads to a more general concept of cryptic sites.

Detection of cryptic sites

Some of the earliest reported cryptic sites were found by serendipity, through detection of distal-site binding of small molecules, many of which were components of the crystallization buffer that bound adventitiously [23]. Targeted identification of cryptic sites has traditionally involved screening libraries of small molecules or fragments [14,50], site-directed tethering [14,51], or the use of antibodies [52]. These experimental approaches require substantial effort and frequently have negative outcomes. Hence, the more recently developed computational approaches are potentially very important. Bowman and Geissler demonstrated that Markov state models built from hundreds of microseconds of molecular dynamics (MD) simulations can identify prospective cryptic sites [7]. In particular, their results showed that the known cryptic site in TEM β-lactamase is at least partially open for 53% of the simulation time, making it the most accessible of the transient pockets detected on this protein. These authors argued that the site is not seen in unbound crystal structures because the closed conformation is also a prominent state, and is likely further stabilized by crystal packing. More recently, the Bowman group designed and tested allosteric modulators binding to the allosteric site of TEM β-lactamase [3•]. Although they identified one inhibitor and two activators, the compounds had only modest effects on TEM activity.

Applying the Markov state method to TEM β-lactamase, Bowman and Geissler also reported around 50 additional pockets that were open and most likely accessible more than 10% of the time [7]. Such sites were distributed across the surface of the protein and were proposed to provide viable drug target sites [7]. This result is not without controversy. Oleinikovas *et al.* [5••] also performed MD simulations on three proteins, including TEM β-lactamase, and found that if the runs were started with the proteins in the open (i.e. bound) conformation, but without the ligands present, the pockets promptly closed. However, they observed that adding fragment sized molecules as probes occasionally caused the re-opening of cryptic sites. Mixed-solvent MD simulations that include small molecules such as benzene and isopropanol in addition to water have been also used by other groups to find cryptic sites [4,53]. Thus it appears that

even the weak binding of a small ligand can keep some of the transient surface pockets, identified by MD simulations, open for longer times. However, we strongly believe that most of these pockets are not druggable. Our argument is based on the observation that druggable cryptic sites almost invariably have a strong binding energy hot spot close by [22••]. The application of FTMap to the extended CryptoSite set showed that the number of such strong hot spots on any given protein never exceeded four [22••]. Furthermore, in addition to the cryptic site, most proteins also had known (i.e. non-cryptic) binding sites with ligands binding at one, two, or sometimes three of these strong hot spots. Thus, based on the analysis of protein structures with validated cryptic sites, Beglov *et al.* concluded that, among the pockets that are seen to open during MD simulations, only one or two may be capable of binding a ligand with substantial affinity [22••]. It is likely therefore that very few of the dozens of transient pockets seen in some Markov state simulations [7] provide viable drug targets. This may be because the opening of these transient sites is usually due to changes only in side chain conformations, and such sites typically have low ligand binding affinity [22••].

Conclusions

A binding site is cryptic if it is formed only in ligand-bound structures, but it is difficult to define the concept more rigorously unless the ensemble of unbound structures to be considered is first specified. Cimermancic *et al.* [1••] developed a benchmark set that is useful for testing detection methods, but includes only pairs of specific unbound–bound structures. Nevertheless, the set includes a number of proteins with genuine cryptic sites that are meaningful drug targets, demonstrating the value of searching for such sites. Methods for identifying cryptic sites that are based on MD simulations need further validation [7], since they identify many transient pockets that are most likely not druggable [22••]. Cimermancic *et al.* [1••] demonstrated that, in addition to molecular dynamics, reliable cryptic site prediction requires the combined use of residue physico-chemical properties, pocket shape, sequence conservation, and fragment docking as features in a machine learning algorithm.

Conflict of interest statement

Nothing declared.

Acknowledgements

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as

- • of special interest
- •• of outstanding interest

1. Cimermancic P, Weinkam P, Rettenmaier TJ, Bichmann L,
•• Keedy DA, Woldeyes RA, Schneidman-Duhovny D,
Demerdash ON, Mitchell JC, Wells JA *et al.*: **CryptoSite: expanding the druggable proteome by characterization and prediction of cryptic binding sites**. *J Mol Biol* 2016, **428**:709-719.
A benchmark set of 93 unbound–bound protein pairs is created in which each unbound structure has a site considered cryptic due to its low pocket score, and each bound structure has a functionally relevant ligand bound at the site. The cryptic sites are comprehensively characterized in terms of their sequence, structure, and dynamics attributes. Relying on this characterization, machine learning was used to predict cryptic sites with relatively high accuracy. The resulting CryptoSite method was employed to predict cryptic sites in the entire structurally characterized human proteome, increasing the size of the potentially 'druggable' human proteome from ~40% to ~78% of disease-associated proteins. The CryptoSite Web server is available at http://salilab.org/cryptosite.

2. Laskowski RA, Gerick F, Thornton JM: **The structural basis of allosteric regulation in proteins**. *FEBS Lett* 2009, **583**:1692-1698.

3. Hart KM, Moeder KE, Ho CMW, Zimmerman MI, Frederick TE,
• Bowman GR: **Designing small molecules to target cryptic pockets yields both positive and negative allosteric modulators**. *PLOS ONE* 2017, **12** e0178678.
The geometry of a cryptic pocket was used to identify an inhibitor and two activators of TEM β-lactamase. To identify hits, a library of compounds was first virtually screened against either the crystal structure of the known cryptic pocket or an ensemble of structures obtained by Markov State Models based on molecular dynamics simulations. The hit compounds were screened experimentally and were shown to have modest effects on TEM activity. The approach is proposed for targeting proteins whose crystal structures lack obvious druggable pockets, and for identifying both inhibitory and activating small-molecule modulators.

4. Kimura SR, Hu HP, Ruvinsky AM, Sherman W, Favia AD: **Deciphering cryptic binding sites on proteins by mixed-solvent molecular dynamics**. *J Chem Inf Model* 2017, **57**:1388-1401.

5. 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.
The nature and dynamic properties of cryptic sites were investigated by performing molecular dynamics simulations on three pharmacologically relevant targets, including TEM β-lactamase. It was found that if the computations were started with the proteins in the open (i.e. bound) conformation, but without the ligands present, the pockets promptly closed. However, adding fragment sized molecules as probes occasionally caused the re-opening of cryptic sites. The observed mechanism of cryptic site formation is suggestive of an interplay between induced-fit and conformational selection. Employing this insight, a novel Hamiltonian Replica Exchange-based method Sampling Water Interfaces through Scaled Hamiltonians (SWISH) was developed for the identification of cryptic sites.

6. Chio CM, Lim CS, Bishop AC: **Targeting a cryptic allosteric site for selective inhibition of the oncogenic protein tyrosine phosphatase Shp2**. *Biochemistry* 2015, **54**:497-504.

7. 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.

8. Durrant JD, McCammon JA: **Molecular dynamics simulations and drug discovery**. *BMC Biol* 2011, **9**:71.

9. 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.

10. Udi Y, Fragai M, Grossman M, Mitternacht S, Arad-Yellin R, Calderone V, Melikian M, Toccafondi M, Berezovsky IN, Luchinat C *et al.*: **Unraveling hidden regulatory sites in structurally homologous metalloproteases**. *J Mol Biol* 2013, **425**:2330-2346.

11. Hopkins AL, Groom CR: **The druggable genome**. *Nat Rev Drug Disc* 2002, **1**:727-730.

12. Whitty A, Kumaravel G: **Between a rock and a hard place?** *Nat Chem Biol* 2006, **2**:112-118.

13. 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.
A recent review of computational and experimental approaches to iden-
tify hidden allosteric sites in proteins with a focus on examples of the
successful use of such techniques. It is suggested that the discovery of
hidden allosteric sites offers a new avenue for facilitating drug design by
greatly expanding the repertoire of available drug targets, contributing to
the search for allosteric drugs for the treatment of human diseases. The
targets discussed are GPCRs, K-Ras4B, interleukin-1 receptor 1, and
TEM β-lactamase.

14. Hardy JA, Wells JA: **Searching for new allosteric sites in
    enzymes**. *Curr Opin Struct Biol* 2004, **14**:706-715.

15. Arkin MR, Whitty A: **The road less traveled: modulating signal
    transduction enzymes by inhibiting their protein–protein
    interactions**. *Curr Opin Chem Biol* 2009, **13**:284-290.

16. Fenwick RB, Esteban-Martin S, Salvatella X: **Understanding
    biomolecular motion, recognition, and allostery by use of
    conformational ensembles**. *Eur Biophys J* 2011, **40**:1339-1355.

17. DuBay KH, Bowman GR, Geissler PL: **Fluctuations within folded
    proteins: implications for thermodynamic and allosteric
    regulation**. *Acc Chem Res* 2015, **48**:1098-1105.

18. Bowman GR, Geissler PL: **Extensive conformational
    heterogeneity within protein cores**. *J Phys Chem B* 2014,
    118:6417-6423.

19. Schueler-Furman O, Wodak SJ: **Computational approaches to
    investigating allostery**. *Curr Opin Struct Biol* 2016, **41**:159-171.

20. 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.

21. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA:
    **Predicting protein ligand binding sites by combining
    evolutionary sequence conservation and 3D structure**. *PLoS
    Comput Biol* 2009, **5** e1000585.

22. Beglov D, Hall DR, Wakefield AE, Luo L, Allen KN, Kozakov D,
 •• Whitty A, Vajda S: **Exploring the structural origins of cryptic
    sites on proteins**. *Proc Natl Acad Sci U S A* 2018.
The proteins in the benchmark set collected in the CryptoSite paper
were studied using the computational hot spot identification tool
FTMap. The results show that, in ligand-free structures, the cryptic
sites that are potential drug targets generally have a strong binding
energy hot spot very close by. Since each protein has only a few such
hot spots, only a limited number of transitional pockets opened by
Markov state methods would be able to bind small molecules with
appropriate affinity. This limitation contradicts some more optimistic
predictions concerning the importance of such pockets for drug
discovery.

23. Horn JR, Shoichet BK: **Allosteric inhibition through core
    disruption**. *J Mol Biol* 2004, **336**:1283-1291.

24. Dellus-Gur E, Toth-Petroczy A, Elias M, Tawfik DS: **What makes a
    protein fold amenable to functional innovation? Fold polarity
    and stability trade-offs**. *J Mol Biol* 2013, **425**:2609-2621.

25. Wiesmann C, Barr KJ, Kung J, Zhu J, Erlanson DA, Shen W,
    Fahr BJ, Zhong M, Taylor L, Randal M *et al.*: **Allosteric inhibition
    of protein tyrosine phosphatase 1B**. *Nat Struct Mol Biol* 2004,
    11:730-737.

26. Li S, Zhang J, Lu S, Huang W, Geng L, Shen Q, Zhang J: **The
    mechanism of allosteric inhibition of protein tyrosine
    phosphatase 1B**. *PLOS ONE* 2014, **9** e97668.

27. Krishnan N, Koveal D, Miller DH, Xue B, Akshinthala SD, Kragelj J,
    Jensen MR, Gauss CM, Page R, Blackledge M *et al.*: **Targeting
    the disordered C terminus of PTP1B with an allosteric
    inhibitor**. *Nat Chem Biol* 2014, **10**:558-566.

28. 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.
A comprehensive review of the development of druggability concepts and
prediction methods. An application for a frequent mutation of p53 tumor
suppressor is presented, illustrating the complementarity of druggability

prediction approaches, the remaining challenges, and potential new drug
development perspectives.

29. Hajduk PJ, Huth JR, Fesik SW: **Druggability indices for protein
    targets derived from NMR-based screening data**. *J Med Chem*
    2005, **48**:2518-2525.

30. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ: **Experimental
    and computational approaches to estimate solubility and
    permeability in drug discovery and development settings**. *Adv
    Drug Del Rev* 2001, **46**:3-26.

31. Lipinski CA: **Rule of five in 2015 and beyond: target and ligand
    structural limitations, ligand chemistry structure and drug
    discovery project decisions**. *Adv Drug Deliv Rev* **(101)**:2016:
    34-41.

32. Doak BC, Zheng J, Dobritzsch D, Kihlberg J: **How beyond rule of
    5 drugs and clinical candidates bind to their targets**. *J Med
    Chem* 2016, **59**:2312-2327.

33. DeGoey DA, Chen HJ, Cox PB, Wendt MD: **Beyond the rule of 5:
    lessons learned from AbbVie's drugs and compound
    collection**. *J Med Chem* 2017. Epub ahead of print.

34. Doak BC, Kihlberg J: **Drug discovery beyond the rule of 5 –
    opportunities and challenges**. *Expert Opin Drug Discov* 2017,
    12:115-119.

35. Kozakov D, Hall DR, Napoleon RL, Yueh C, Whitty A, Vajda S: **New
    frontiers in druggability**. *J Med Chem* 2015, **58**:9063-9088.

36. Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L,
    Xia B, Beglov D, Vajda S: **The FTMap family of web servers for
    determining and characterizing ligand-binding hot spots of
    proteins**. *Nat Protoc* 2015, **10**:733-755.

37. Wagner JR, Lee CT, Durrant JD, Malmstrom RD, Feher VA,
    Amaro RE: **Emerging computational methods for the rational
    discovery of allosteric drugs**. *Chem Rev* 2016, **116**:6370-6390.

38. Acker TM, Gable JE, Bohn MF, Jaishankar P, Thompson MC,
    Fraser JS, Renslo AR, Craik CS: **Allosteric inhibitors,
    crystallography, and comparative analysis reveal network of
    coordinated movement across human herpesvirus proteases**.
    J Am Chem Soc 2017, **139**:11650-11653.

39. Laskowski RA, Luscombe NM, Swindells MB, Thornton JM:
    Protein clefts in molecular recognition and function. *Protein
    Sci* 1996, **5**:2438-2452.

40. Murkin AS, Manning KA, Kholodar SA: **Mechanism and inhibition
    of 1-deoxy-D-xylulose-5-phosphate reductoisomerase**.
    Bioorg Chem 2014, **57**:171-185.

41. Sullivan SM, Holyoak T: **Enzymes with lid-gated active sites
    must operate by an induced fit mechanism instead of
    conformational selection**. *Proc Natl Acad Sci U S A* 2008,
    105:13829-13834.

42. Sengupta S, Ghosh S, Nagaraja V: **Moonlighting function of
    glutamate racemase from *Mycobacterium tuberculosis*:
    racemization and DNA gyrase inhibition are two independent
    activities of the enzyme**. *Microbiology* 2008, **154**:2796-2803.

43. Sawyer N, Watkins AM, Arora PS: **Protein domain mimics as
    modulators of protein–protein interactions**. *Acc Chem Res*
    2017, **50**:1313-1322.

44. Kozakov D, Hall DR, Chuang GY, Cencic R, Brenke R, Grove LE,
    Beglov D, Pelletier J, Whitty A, Vajda S: **Structural conservation
    of druggable hot spots in protein–protein interfaces**. *Proc Natl
    Acad Sci U S A* 2011, **108**:13528-13533.

45. Arkin MR, Randal M, DeLano WL, Hyde J, Luong TN, Oslob JD,
    Raphael DR, Taylor L, Wang J, McDowell RS *et al.*: **Binding of
    small molecules to an adaptive protein–protein interface**. *Proc
    Natl Acad Sci U S A* 2003, **100**:1603-1608.

46. Thanos CD, Randal M, Wells JA: **Potent small-molecule binding
    to a dynamic hot spot on IL-2**. *J Am Chem Soc* 2003, **125**:
    15280-15281.

47. Maurer T, Garrenton LS, Oh A, Pitts K, Anderson DJ, Skelton NJ,
    Fauber BP, Pan B, Malek S, Stokoe D *et al.*: **Small-molecule
    ligands bind to a distinct pocket in Ras and inhibit SOS-**

mediated nucleotide exchange activity. *Proc Natl Acad Sci U S A* 2012, **109**:5299-5304.

48. Sun Q, Burke JP, Phan J, Burns MC, Olejniczak ET, Waterson AG, Lee T, Rossanese OW, Fesik SW: **Discovery of small molecules that bind to k-ras and inhibit sos-mediated activation**. *Ang Chem Int Ed* 2012, **51**:6140-6143.

49. Basavapathruni A, Jin L, Daigle SR, Majer CR, Therkelsen CA, Wigle TJ, Kuntz KW, Chesworth R, Pollock RM, Scott MP *et al.*: **Conformational adaptation drives potent, selective and durable inhibition of the human protein methyltransferase DOT1L**. *Chem Biol Drug Des* 2012, **80**:971-980.

50. 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.

51. Erlanson DA, Wells JA, Braisted AC: **Tethering: fragment-based drug discovery**. *Annu Rev Biophys Biomol Struct* 2004, **33**: 199-223.

52. Lawson AD: **Antibody-enabled small-molecule drug discovery**. *Nat Rev Drug Disc* 2012, **11**:519-525.

53. Ghanakota P, Carlson HA: **Moving beyond active-site detection: MixMD applied to allosteric systems**. *J Phys Chem B* 2016, **120**:8685-8695.