Druggability of methyl-lysine binding sites

Calvin Santiago[1], Kong T. Nguyen[1], Matthieu Schapira[1,2,*]

1 – Structural Genomics Consortium, University of Toronto, 101 College st., Toronto, ON, Canada

2 – Department of Pharmacology and Toxicology, University of Toronto, 101 College st., Toronto, ON, Canada

* Corresponding author: matthieu.schapira@utoronto.ca
tel: 1-416-978-3092
fax: 1-416-946-0880

ABSTRACT

Structural modules that specifically recognize - or read - methylated or acetylated lysine residues on histone peptides are important components of chromatin-mediated signaling and are involved in the epigenetic regulation of gene expression. Deregulation of epigenetic mechanisms is associated with disease conditions, and antagonists of bromodomains - modules that read acetyl-lysines - are efficacious in animal models of cancer and inflammation; on the other hand, little is known regarding the druggability of methyl-lysine binding modules. We conducted a systematic structural analysis of readers of methyl marks and derived a predictive druggability landscape of methyl-lysine binding modules. We show that these target classes are generally less druggable than bromodomains, but that some proteins stand as notable exceptions. We find that limited protein backbone motion or the presence of side-cavities juxtaposed to the methyl-lysine binding pocket can significantly improve druggability. We also show that druggability can be conditional on the inhibitor's ability to induce conformational rearrangements. This structure-based druggability landscape can be used to

prioritize efforts aimed at developing small molecule antagonists against readers of the histone code.

KEYWORDS

Druggability, Binding Pocket, Methyl-lysine, Inhibitors, Epigenetics

INTRODUCTION

Post-translational modifications on histone tails dictate the compaction state of chromatin, which in turn controls access of the transcription machinery to target genes [1]. The most frequent histone marks are acetylation and methylation of lysine side-chains, and arginine methylation. Distinct combinations of such marks, and their recognition by specific binding modules, signal activation or silencing of gene expression in different cell types [2; 3]. The deregulation of this refine and complex signaling system has been associated with disease states [4; 5], and has been validated as a point of pharmacological intervention: two histone deacetylase inhibitors are currently approved for the treatment of cutaneous T-cell lymphoma [6]. Other enzymes involved in writing and erasing histone marks -protein methyltransferases, lysine acetyltransferases and lysine demethylases- are expected to be valid target class for therapy. Potent and selective chemical probes have been reported in the case of protein methyltransferases, clearly demonstrating the chemical tractability of this protein family [7; 8].

Readers of histone marks also play a critical role in chromatin mediated signaling, and could be targeted by small molecule antagonists [9; 10; 11]. These binding modules consist generally of independent structural domains that are part of larger genes. Acetylated lysines are recognized by bromodomains [12; 13]. Methylated lysines and arginines are recognized by PHD zinc finger domains [14; 15], or the Royal family, composed of Tudor, MBT, PWWP, and chromodomains [16; 17; 18]. Protein interaction events are historically less chemically tractable than enzymes, but the development of low nanomolar preclinical compounds targeting bromodomains has revealed that chemically antagonizing acetyl-lysine recognition is a valid avenue to dissect chromatin biology, and probably for drug discovery [19; 20]. The recent discovery of a low micromolar and selective small molecule antagonist of the methyl-lysine reader

L3MBTL1 suggests that binding modules reading methyl marks may also be chemically tractable [21; 22].

A large body of structural data is available on these binding modules, with over 50 structures solved in complex with histone peptides [23; 24]. In this work, we analyze these structures to derive a druggability landscape across the major classes of methyl-lysine readers, and understand the structural features that drive druggability. This analysis should help prioritize medicinal chemistry efforts targeting readers of the histone code.

COMPUTATIONAL METHODS

Protein Preparation

Most binding sites analyzed in this work are represented multiple times in each PDB file, due to the redundant presence of each chain in the crystal asymmetric unit. To save computing time and to avoid the formation of unphysical binding sites between monomers, only one monomer, usually chain A and its corresponding bound peptide/small molecule ligand, was retained for further protein preparation steps, unless explicitly specified. When multiple domains with the same fold were present within a given peptide chain, only the domain in complex with a methyl-lysine was kept. For examples, in the case of L3MBT1 structures (PDB codes: 2RJF, 3P8H), two out of three MBT domains, namely the first and the third MBT domains were excluded from the protein preparation steps since it was previously shown that only the second MBT domains of L3MBT1 can bind methylated lysine [25]. Similarly, the first three MBT domains of L3MBT2 (PDB: 3F70) were removed because only the fourth MBT repeat can bind H4K20me1 [26]. All water molecules were removed, except for bromodomain structures which contain three conserved water molecules that are important for receptor-ligand recognition (molecules W3, W30 and W34 in the second bromodomain of BRD2, PDB code 3ONI).

All steps of the Maestro Protein Preparation Wizard were followed to prepare the proteins used in this study. All structures were downloaded from the PDB and loaded into Schrodinger Maestro workspace. In case of the five NMR structures included in this study (no crystal structure is available in complex with a ligand for CBX1, CBX7, AIRE, DPF3, and TAF3), the first snapshot was used. Water molecules were removed unless indicated otherwise as described above. ICM (Molsoft, San Diego) was used to add missing side-chains where necessary. Protonation states were set at pH 7.4 using Epik. H-bond assignment was optimized by protassign, including exhaustive sampling and minimizing hydrogens of altered species at neutral pH. Water orientations were also sampled for the three conserved waters molecules in bromodomains. The bound ligand-protein complexes were refined during Impref minimization with RMSD=0.35 Angstrom set for heavy atoms convergence.

We also carried out another series of protein preparation using a different procedure where the bound ligands were removed from the beginning and impref was applied to the unliganded protein. Following this procedure, the accuracy of suggested binding sites, however, were lower than that of the other procedure (data not shown). Our observation is in agreement with what was published by other authors that protein-ligand complex relaxation does improve the SiteMap results [27] and that SiteMap performs better on holo structures than apo structures [28].

Computing DScore

The bound ligand was removed from the complex after impref refinement step and the unliganded protein was used for computing a druggability score (DScore) with SiteMap (Schrodinger, New York). To find all possible binding sites in a structure, the automatic identification of potential binding sites implemented in SiteMap was selected with all default parameters (15 “site points” per reported site, report up to 5 sites, using a restrictive definition of hydrophobicity and standard grids, cropping site maps at 4 Angstroms from the nearest site point,

and using OPLS_2005 force field). For each input protein, SiteMap suggests a list of up to five potential binding sites. The site that includes the bound ligand (methylated lysine, acetylated lysine, arginine or small molecule inhibitor) was selected manually. Computed DScore values of the selected binding sites were then extracted and plotted in Figure 1. If the binding sites found by SiteMap did not include the bound ligand, a null DScore was assigned, meaning the site around the bound ligand is undruggable.

RESULTS

A druggability score (Dscore) that accounts for volume, enclosure and hydrophobicity of binding pockets was calculated for 48 structures of histone binding modules solved in complex with a substrate or a small molecule inhibitor, representing 45 distinct domains across 10 different structural folds (see Methods section for details). Bromodomains are binding acetyl-lysine, and not methyl-lysine, but the structures from the BET bromodomain subfamily were added as a druggable reference, since low nanomolar, drug-like compounds with *in vivo* efficacy were recently reported against this branch of the phylogenetic tree [19; 20]. The Dscores were calculated with SiteMap (Schrodinger, New York, NY) using the software's built-in binding site identification algorithm. This site includes and generally extends beyond the restrictive limits of the pocket defined by the bound ligand (methyl-lysine, chemical inhibitor, or arginine in the case of WDR5 and SND1). When multiple binding sites were found, only the site including the bound ligand was selected. Halgren's score cut-offs were used to evaluate druggability of a site, defined as the ability to bind with high affinity passively absorbed small molecules: binding pockets with Dscores lower than 0.83 are predicted "undruggable", pockets with Dscores above 0.98 are predicted "druggable", and Dscores between these thresholds indicate sites that are difficult to drug [27]. For simplicity, we will refer to this third, less reliable range as "druggable" as well, even though it should be kept in mind that sites falling in this

area of the druggability landscape are expected to be associated with higher failure rates in drug-discovery projects.

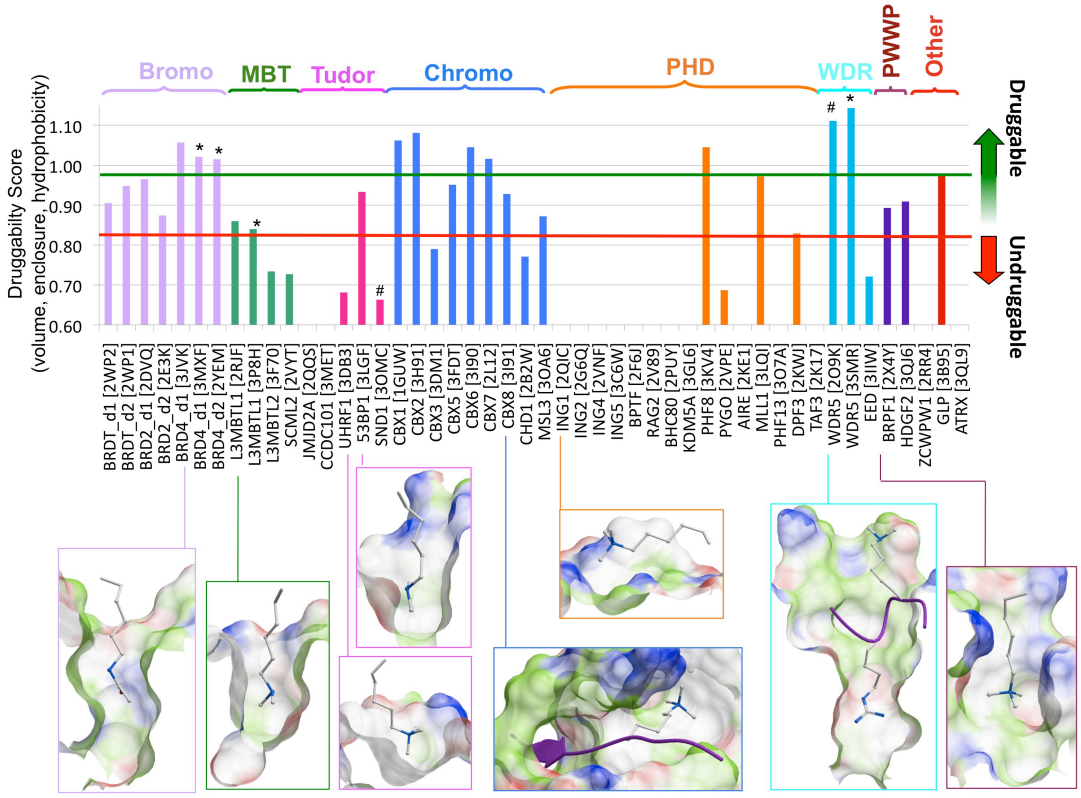


Figure 1: Structure-based druggability of histone mark binding domains. Binding pockets are occupied by a methyl-lysine (acetyl-lysine in the case of bromodomains), unless otherwise specified by a "*" (chemical inhibitor) or a "#" (arginine). Druggability scores below 0.83 indicate undruggable structures; values above 0.98 indicate very druggable structures (see text for details). Target families, name and PDB codes are indicated (in the case of proteins containing multiple acetyl-lysine binding bromodomains, the domain number is specified. Ex: BRDT1_d1 is the first bromodomain of BRDT). No bar indicates a Dscore < 0.6. Molecular surface coloring: green: hydrophobic; red: hydrogen bond acceptor; blue: hydrogen bond donor.

The druggability of methyl-lysine binding sites is highly variable

Overall, a good correlation was observed between the pocket geometry, and the Dscore assigned: deep, buried pockets had favourable Dscores, shallow pockets did not (Figure 1). A first observation is that the predicted druggability of methyl-lysine binding domains is more variable than that of the BET bromodomains. All seven structures representing 6 distinct BET bromodomains are predicted druggable, while, apart from the PWWP domain (poorly represented by only two structures), no methyl-lysine binding fold is consistently populated with druggable binding sites. Nanomolar drug-like inhibitors were developed against BET bromodomains: it comes as no surprise that the corresponding acetyl-lysine binding sites are associated with high Dscores [19; 20]. This suggests that the bromodomain fold has a topology that is favourable to drug discovery regardless of sequence (at least within the BET sub-family). The situation is not as favourable for methyl-lysine binding domains, but we note that in all cases, druggable binding sites exist, in varying proportion: most promising are PWWP domains, with the only two complex structures available being druggable. The central hole of the WD repeat canonical donut shape is also promising, as shown by the Dscores > 1.1 of WDR5 in complex with a histone peptide (PDB code 2O9K) or a small molecule ligand (PDB code 3SMR). We note here that WDR5 binds a methyl-lysine peptide, but that the methyl-lysine itself contributes little to the interaction, unlike the arginine side-chain, positioned two residues upstream, that is deeply inserted in the central cavity of the protein. In the case of EED, the pocket exploited by the methyl-lysine is not druggable (PDB code 3IIW - Dscore = 0.72), but the other face of the donut - where EZH2 is recruited - is (Supp. Figure 1, PDB code 2QXV – Dscore = 0.98). Next are chromodomains and MBT where seven out of nine and one out of three sites respectively are found druggable. In the case of the CBX proteins, the druggable site extends far beyond the methyl-lysine pocket, into the groove that is occupied by the histone tail backbone (discussed in detailed below). The ratio falls to one druggable site out of five proteins for Tudor domains, and three out of fifteen for PHD fingers. Finally, a few other folds have been solved in complex with methyl-lysines (zf-CW domain of ZCWPW1, ankyrin repeat of GLP, ADD domain of ATRX), of

which only the ankyrin repeat of GLP (PDB code 3B95) is predicted druggable. These results show that (1) BET bromodomains are the most druggable as a whole, (2) some methyl-lysine binding folds are more favourable to drug discovery programs than others, and (3) druggabilty can vary significantly from one target to another within each family.

Small structural variations can significantly alter druggability

To judge the sensitivity of the computed druggability index, related structures with variable Dscores were compared. Among MBT domains, the methyl-lysine binding site of L3MBTL1 was more druggable than that of L3MBTL2 or SCML2. Dscores derived from peptide-bound and inhibitor-bound L3MBTL1 structures were 0.86 and 0.84 respectively (PDB codes 2RJF and 3P8H), while peptide-bound L3MBTL2 and SCML2 produced Dscores of 0.74 and 0.73 (PDB codes 3F70 and 2VYT) (Figure 1). Superimposing the L3MBLT1 and SCML2 methyl-lysine pockets revealed that all but one residue are conserved. The only variation in sequence and structure is at T411 of L3MBTL1 (Q238 in SCML2). The pocket is open at this position in SCML2, but closed by the methyl group of the threonine's side-chain in L3MBTL1 (Figure 2). The L3MBTL1 site is consequently more enclosed and more hydrophobic than that of SCML2, which results in increased predicted druggability. These results highlight the sensitivity of the druggability function to small structural variations. Crystal structures being static snapshots of dynamic systems, this raises the possibility that different structures of the same site would be associated with contradicting druggability predictions. We note however that the first bromodomain of BRD4 was represented twice in the analysis, in complex with an acetyl-lysine [3JVK] and in complex with the inhibitor JQ1 [3MXF], with corresponding Dscores both in the most favourable region of the druggability landscape (Dscores = 1.06 and 1.02 respectively). Similarly, the second MBT domain of L3MBTL1 was crystallized in complex with a methyl-lysine [2RJF] and a chemical inhibitor [3P8H] with virtually identical Dscores of 0.86 and 0.84 respectively. Finally, the central cavity of WDR5 was solved in complex with a

methyl-lysine peptide [2O9K] and an inhibitor [3SMR], with almost identical Dscores of 1.11 and 1.14 respectively. These results suggest that, when no backbone motion is observed, as is the case in these three examples, the predicted druggability is robust.

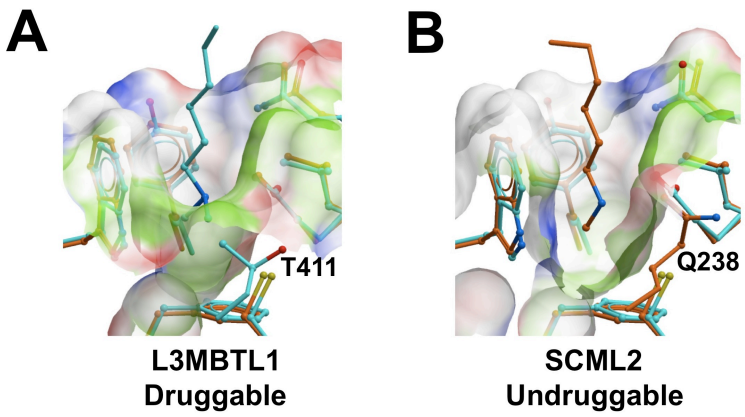


Figure 2: Small sequence variations can significantly influence Dscores. A single difference in side-chains lining the methyl-lysine site between L3MBTL1 (A, cyan – PDB code 2RJF – Dscore = 0.86) and SCML2 (B, orange – PDB code 2VYT – Dscore = 0.73) (T411 in L3MBTL1, Q238 in SCML2) results in a significant variation in predicted druggability: the methyl group of T411 produces a more enclosed and more hydrophobic pocket, two features that increase its Dscore.

Protein backbone motion has a strong impact on druggability

Binding of a substrate peptide can sometimes induce large conformational rearrangements of the methyl-lysine domain. For instance the N-terminal extremity of the CBX chromodomain is in a loose, solvated conformation in the free state, but closes in a beta-sheet arrangement onto the substrate peptide upon methyl-lysine binding [29] (Figure 3). While the apo structure is shallow and is predicted undruggable (DScore = 0.31 in the case of CBX1 - PDB code 2F2U), the backbone of the substrate peptide lays in an enclosed pocket, upstream of the methyl-lysine, that is generally predicted druggable (Figure 1). It should be noted here that, unlike other CBX proteins, CBX3 is predicted undruggable, due

to the very electronegative nature of its binding pocket. It was previously proposed that this electrostatic property is also dictating substrate selectivity [29]. In this work, we made the decision to derive druggability scores exclusively from substrate or inhibitor-bound structures, which implicitly assumes that small molecules can induce any conformational re-arrangement that may be observed in the substrate-bound form. In the case of CBX proteins, this assumption rests on the observation that proper positioning of a conserved aromatic side-chain (Phe11 in CBX6) at the methyl-lysine binding site is coupled to β-closure of the N-terminal extremity of the chromodomain backbone [29].

Conformational variability observed in the third PHD finger of MLL also illustrates how substrate binding affects the geometry and predicted druggability of a site. In the apo conformation, an insertion loop, absent from most PHD fingers, adopts an open conformation which results in a shallow, undruggable methyl-lysine binding site, undetectable by SiteMap (PDB code 3LQH). Upon substrate binding, the loop closes onto the methyl-lysine side-chain, which contributes to the formation of a druggable site (PDB code 3LQI – Dscore = 0.97) (Figure 3) [30]. A similar observation can be made upon binding of a methyl-lysine peptide to the second MBT domain of L3MBTL2: a C-terminal tail immediately following the MBT domain folds onto and extends the methyl-lysine pocket only upon substrate binding, which enhances druggabilty (apo structure [3CEY]: Dscore = 0.61 – complex structure [3F70]: Dscore=0.73) (Figure 3). We note that a similar backbone motion is prohibited in L3MBTL1 by the relative arrangement of the three MBT domains (the peptide chain immediately C-terminal to the methyl-lysine bound MBT domain is recruited by the following MBT domain and cannot fold back on the methyl-lysine pocket), and would depend in SCML2 on the unknown quaternary structure of the full-length protein. These results show that protein dynamics must be considered to assess druggability, and that compounds should be designed not only to occupy a specific site, but also to induce protein backbone motions that may be necessary to the druggability of this site.

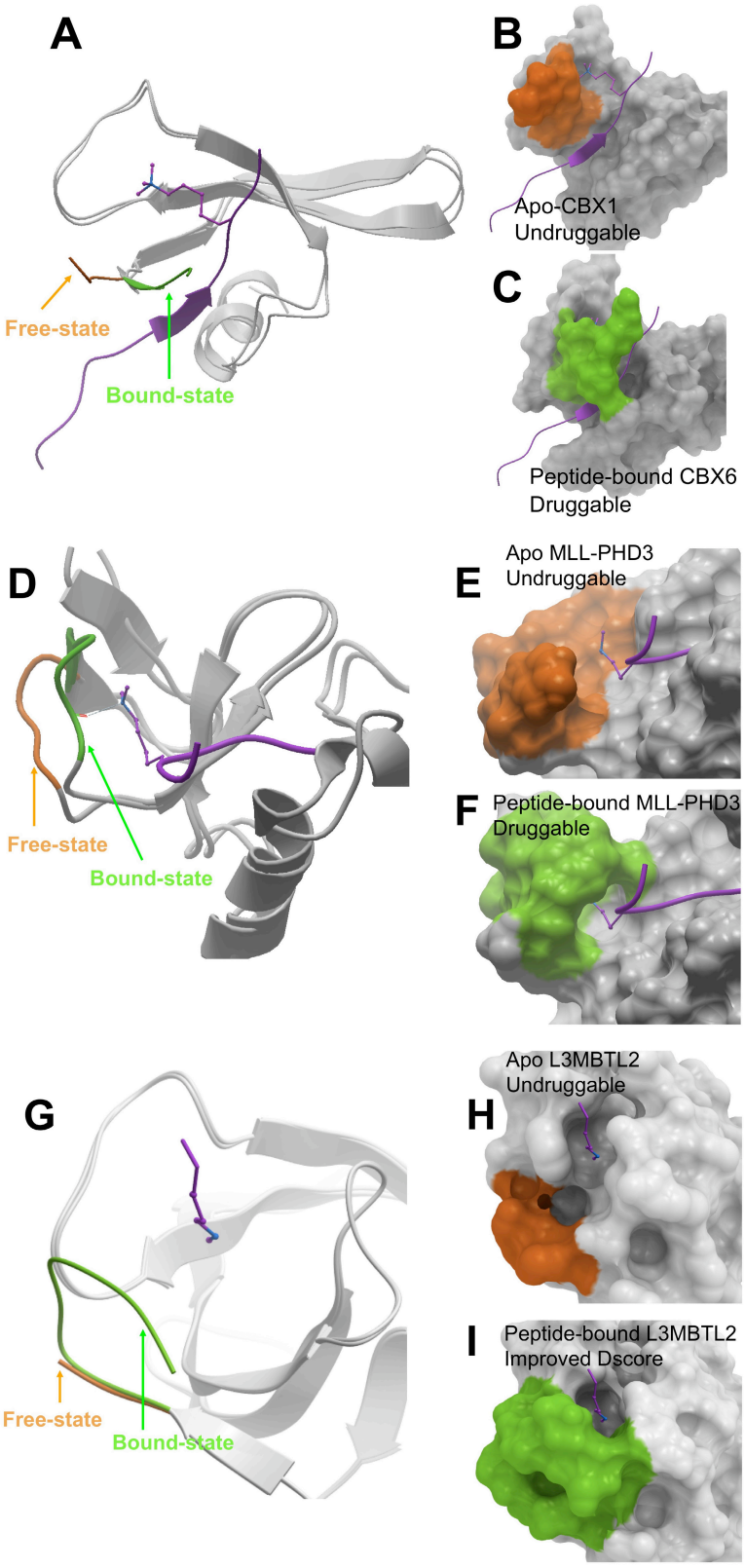

A
Free-state
Bound-state
B
Apo-CBX1
Undruggable
C
Peptide-bound CBX6
Druggable
D
Free-state
Bound-state
E
Apo MLL-PHD3
Undruggable
F
Peptide-bound MLL-PHD3
Druggable
G
Free-state
Bound-state
H
Apo L3MBTL2
Undruggable
I
Peptide-bound L3MBTL2
Improved Dscore

Figure 3: Substrate-induced pocket dynamics affects druggability. Top: the N-terminal extremity of CBX chromodomains is in an open conformation in the apo state (A,B – orange – Dscore= 0.31 - apo CBX1 [3F2U] shown) but is recruited within a beta-sheet in the substrate-bound conformation (A,C - green – Dscore=0.86 - CBX6 [3GV6] shown). The apo state is shallow and predicted undruggable, but the peptide-bound state is generally predicted druggable. **Middle**: The methyl-lysine pocket of MLL's third PHD domain features an insertion loop that adopts an open, undruggable conformation in the apo state (D,E - orange – pocket not identified by siteMap – PDB code [3LQH]). Substrate binding induces a conformational rearrangement and formation of a druggable site (D,F – green – Dscore = 0.97 – PDB code [3LQI]). **Bottom**: A C-terminal extension to the fourth MBT domain of L3MBTL2 does not contribute to the undruggable methyl-lysine pocket in the apo state (G,H - orange – Dscore = 0.61 – PDB code [3CEY]), but closes on the methyl-lysine upon substrate binding, resulting in a site with improved Dscore (G,I – green – Dscore = 0.73 – PDB code [3F70]). The substrate peptide (magenta) is shown as a reference, but is not present in the apo structures (B,E,H).

Druggability may depend on the presence of secondary cavities

The present study aims at evaluating the feasibility to develop drugs competing with methyl-lysine peptides for binding at methyl-lysine reading domains. We should note here that the sites analyzed all include, but are not restricted to the methyl-lysine binding pocket. This is illustrated for example by CBX binding sites, that are composed, in addition to the methyl-lysine binding pocket, of a channel recruiting the backbone of the substrate peptide (Figure 3C). The ankyrin repeat of GLP, an atypical methyl-lysine binding module, is another example [31] (Figure 4A,B). A mono- or di-methyl side-chain is anchored into a first cavity, composed of a canonical aromatic cage. The isolated pocket is predicted undruggable (Dscore = 0.67). Immediately located on opposite sides of the methyl-lysine pocket are two distinct cavities, one occupied by the side-chain of Thr11 in the structure in complex with a dimethylated H3K9 peptide, and the other empty (PDB code 3B95). The binding site detection algorithm implemented in SiteMap identifies a druggable site composed of the three pockets (Dscore =

0.98, Figure 4B). Another example is provided by the third PHD finger of MLL (PDB code 3LQI) [30]. As we have seen above, peptide binding induces a large conformational rearrangement that closes an insertion loop onto the methyl-lysine side-chain (Figure 3E,F). The resulting pocket is more enclosed than in the apo conformation, but yet, not predicted druggable (Figure 4C, Dscore = 0.71). A second conformational rearrangement observed upon histone peptide binding occurs in the vicinity of Ala1 of histone 3, the side-chain of which becomes buried in a side cavity (Figure 4D). The binding site identified by SiteMap spans both H3K4 and H3A1 pockets, and is predicted druggable (Dscore = 0.97) while the isolated methyl-lysine pocket is not (Dscore = 0.71). These results show that exploiting exclusively the docking site of the methyl-lysine side-chain may often not be sufficient to develop small molecule antagonists: potent compounds should extend towards side cavities.

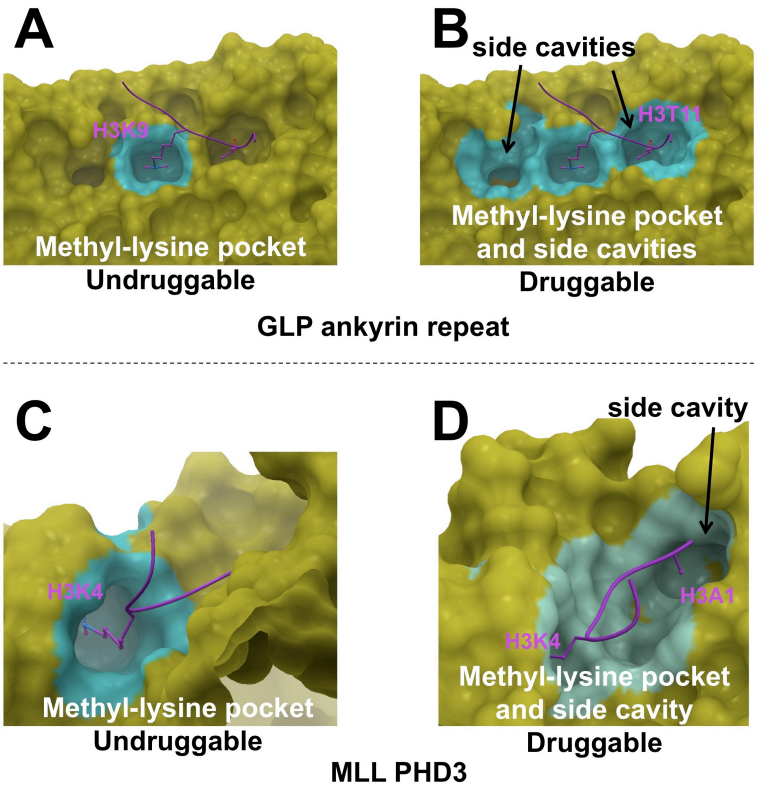


Figure 4: Side cavities can contribute significantly to druggabiliy. Top: the methyl-lysine binding pocket of the GLP ankyrin repeat (A, cyan) is undruggable (Dscore = 0.67). Including two side cavities results in a druggable site (B, Dscore = 0.98). Bottom:

the methyl-lysine pocket of MLL's third PHD finger (C, cyan) in isolation is undruggable (Dscore = 0.71). Accounting for a neighbouring side cavity results in a druggable site (D, Dscore = 0.97).

Unknown inter-domain arrangements may improve druggability

Apart from MLL, the PHD finger of the demethylase PHF8 stands out as remarkably druggable relative to other members of the PHD family (Figure 1). This domain adopts a canonical PHD fold organized around 2 zinc ions. When artificially isolated from a larger structure that includes both PHF8's PHD and JMJ domains in complex with a thrimethylated H3K4 peptide [32] (PDB code 3KV4), the PHD finger is characterized by a shallow, undruggable binding site (Dscore = 0.45) (Figure 5A). A similar surface is observed in most available PHD finger structures (Supp. Figure 2). However, the complete structure reveals that the JMJ domain of PHF8 folds onto the methyl-lysine binding site of the PHD finger, thereby generating a very enclosed and druggable cleft (Dscore = 1.05) (Figure 5B). The only other demethylase for which a structure includes both PHD and JMJ domains is JHDM1D (PDB code 3KV5), where a different quaternary arrangement is observed, that leaves the methyl-lysine site of the PHD finger open to solvent, and undruggable. Other PHD fingers have been solved fused to bromodomains (e.g. BPTF, PDB code 2F6J), but there again, the two domains are in a relative arrangement that leaves the methyl-lysine binding site undruggable. This example clearly illustrates the risks of deriving a putative druggability index from a structural domain isolated from the full-length protein. This is particularly true for readers of histone marks, that are often organized as structural modules within larger multi-modular proteins, and where intra-molecular interactions between domains are likely.

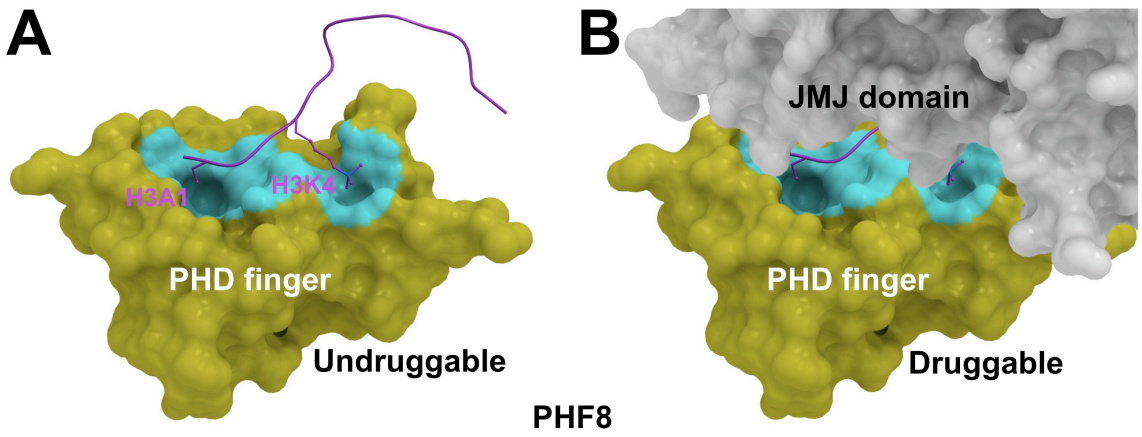


Figure 5: Effect of intramolecular interactions between domains. When artificially isolated from the neighboring JMJ domain, the PHD finger of PHF8 features an undruggable binding site organized around shallow cavities occupied by methyl-lysine and alanine side-chains (A – Dscore = 0.45), but quaternary structural arrangements bring the JMJ domain of PHF8 on top of the methyl-lysine binding site, which becomes buried, and druggable (B – Dscore = 1.05) (PDB code 3KV4).

DISCUSSION

A number of lessons can be drawn from the druggability landscape outlined in this work. First the Royal family of methyl-lysine binding domains as a whole is based on a topology that is not as favourable to the development of potent inhibitors as the BET family of bromodomains. This general observation should not be taken as an indication that methyl-lysine binding sites are not druggable. On the contrary, our data shows that each protein family is populated, to varying extents, with druggable targets (Figure 1). In some cases, such as the ankyrin repeat of GLP, we predict that compounds will bind potently only when exploiting side cavities in addition to the methyl-lysine binding cleft (Figure 4). In other cases, such as CBX proteins or the third PHD domain of MLL, potency will be conditional on the ability of the compound to induce significant backbone conformational shifts, mimicking the effect of histone peptides (Figure 3). The fact that the observed druggability of BET bromodomains, WD repeats and

PWWP domains depends neither on side cavity, nor on ligand-induced conformational rearrangements make them all the more promising target classes.

We have seen with the PHD finger of PHF8 that deriving druggabilty values from structural domains extracted from the context of the full-length protein comes with the risk of overlooking important intramolecular interactions between distinct domains can be (Figure 5). In addition, genes involved in epigenetic signaling are often acting within multi-protein complexes, resulting in protein-protein interactions that may affect the accessibility and topology of binding sites, a reality that can only be ignored in the current analysis, considering the lack of structures of protein complexes. An example - that may only be a crystallographic artefact - illustrates how protein-protein interactions can dramatically impact druggability of methyl-lysine binding sites. The PHD finger of ING4, as most family members, is shallow and undruggable, as can be seen from a structure in complex with an H3K4me3 peptide (PDB code 2VNF - Dscore = 0.53) (Supplementary Figure 3A) [33]. Intriguingly, the arrangement of the protein in the crystal lattice places a second ING4-H3K4me3 complex on top of the methyl-lysine binding site of the first monomer. The corresponding bound methyl-lysine is sandwiched between the two domains, in a very enclosed and druggable cavity (Supplementary Figure 3B). The high local concentration of histone tails in chromatin is likely to bring diverse reading modules in proximity, and may induce unforeseen protein interaction events, such as the one observed here. Another example, again with unknown biological relevance, is provided by the structure of L3MBTL1 in complex with the inhibitor UNC669 (PDB code 3P8H). The structure reveals that the antagonist occupies an aromatic cage that is otherwise responsible for the recruitment of the methyl-lysine side-chain (Supplementary Figure 3C) [22]. The corresponding binding site is in the difficult-to-drug range of the druggability index (Dscore = 0.84). The crystal lattice of the complex structure places another L3MBTL1 monomer in the immediate vicinity of the bound ligand. This dimeric arrangement results in a more enclosed and druggable binding site (Supplementary Figure 3D - Dscore = 1.06). We note that

L3MBTL1 dimerization was already observed in a histone peptide-bound state, and histone binding was proposed to mediate dimerization events [25]. These examples clearly show how interactions between structural modules can significantly impact the druggability of a site. If these modules belong to the same protein, as in PHF8, screening against a full-length construct is probably a necessity. If the domains come from different proteins, biochemical reconstitution of the complex, or cell based assays should come to mind, but these considerations are beyond the scope of the present work.

CONCLUSIONS

Methyl-lysine binding domains are important actors of chromatin-mediated signaling, are also involved in non-epigenetic mechanisms, and represent potential targets for drug design. Efforts to synthesize tool compounds to probe the biology of these proteins, or develop clinical candidates are only worthwhile if the structural topology of the methyl-lysine binding site allows potent binding of drug-like molecules. We have shown here that most families of methyl-lysine binding domains are populated with both druggable and undruggable members. Based on structural data available so far, the druggability of CBX proteins depends on backbone motion, while WD repeats and PWWP domains seem to be structurally more favourable than PHD fingers or Tudor domains. This landscape will gain in resolution as more complex structures are solved, and can be used to prioritize screening efforts.

ACKNOWLEDGEMENTS

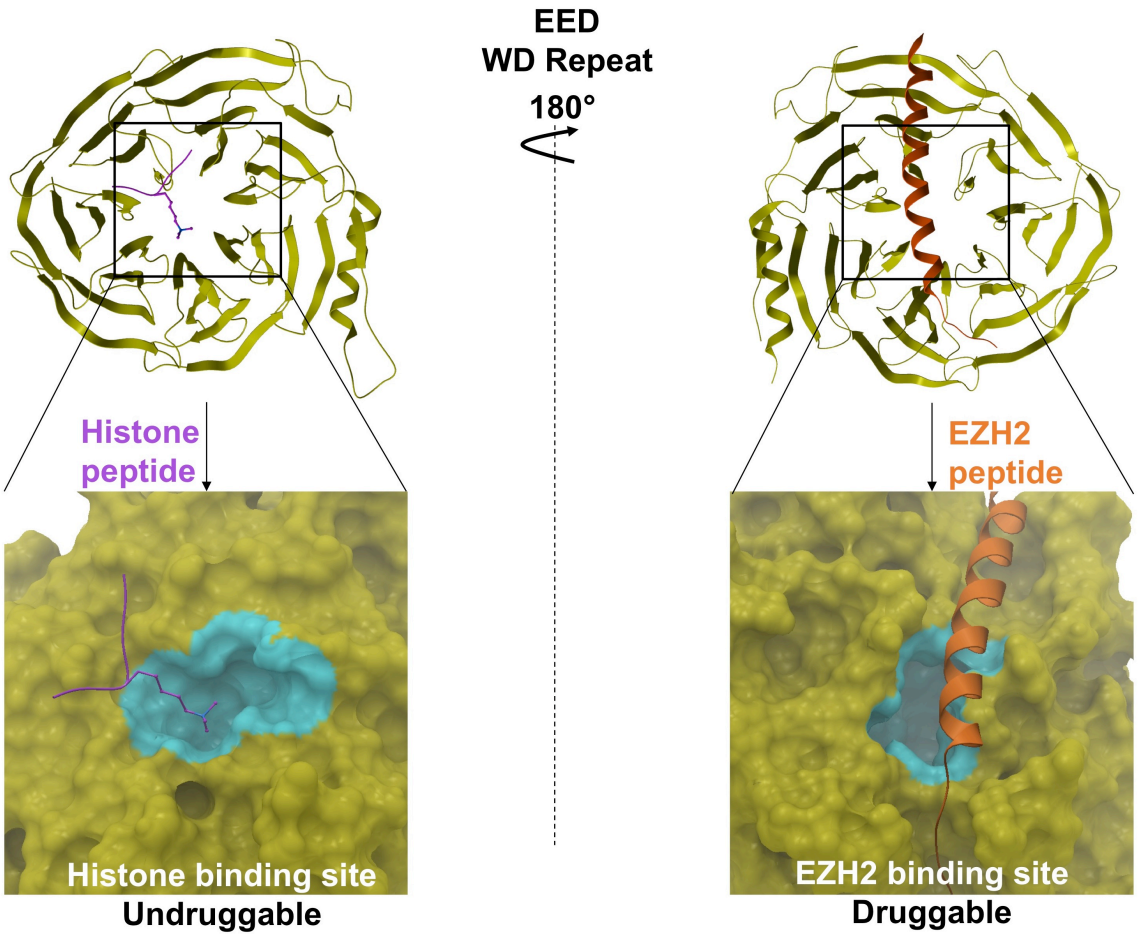
The Structural Genomics Consortium is a registered charity (number 1097737) that receives funds from Canadian Institutes for Health Research, Canadian Foundation for Innovation, Genome Canada through the Ontario Genomics Institute, GlaxoSmithKline, Eli Lilly, Pfizer, Novartis Research Foundation, Life

Technologies, Ontario Innovation Trust, Ontario Ministry for Research and Innovation, and Wellcome Trust.

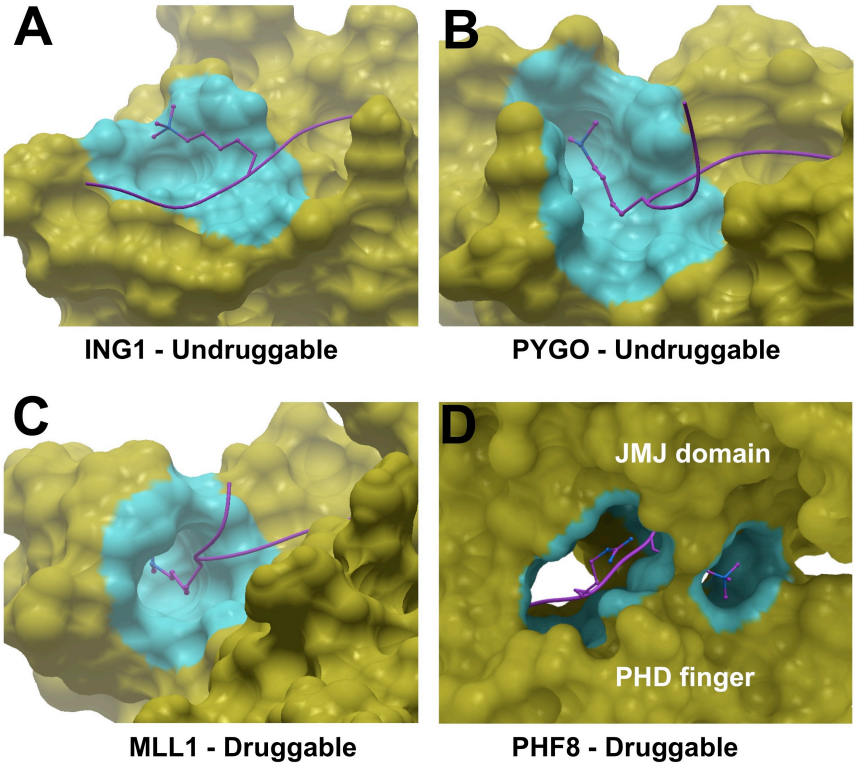
REFERENCES

[1] S.I. Grewal and D. Moazed, Science, 301 (2003) 798.
[2] T. Jenuwein and C.D. Allis, Science, 293 (2001) 1074.
[3] B.D. Strahl and C.D. Allis, Nature, 403 (2000) 41.
[4] T.K. Kelly, D.D. De Carvalho and P.A. Jones, Nat Biotechnol, 28 (2010) 1069.
[5] A. Portela and M. Esteller, Nat Biotechnol, 28 (2010) 1057.
[6] H.M. Prince, M.J. Bishton and S.J. Harrison, Clin Cancer Res, 15 (2009) 3958.
[7] S.R. Daigle, E.J. Olhava, C.A. Therkelsen, C.R. Majer, C.J. Sneeringer, J. Song, L.D. Johnston, M.P. Scott, J.J. Smith, Y. Xiao, L. Jin, K.W. Kuntz, R. Chesworth, M.P. Moyer, K.M. Bernt, J.C. Tseng, A.L. Kung, S.A. Armstrong, R.A. Copeland, V.M. Richon and R.M. Pollock, Cancer Cell, 20 (2011) 53.
[8] M. Vedadi, D. Barsyte-Lovejoy, F. Liu, S. Rival-Gervier, A. Allali-Hassani, V. Labrie, T.J. Wigle, P.A. Dimaggio, G.A. Wasney, A. Siarheyeva, A. Dong, W. Tempel, S.C. Wang, X. Chen, I. Chau, T.J. Mangano, X.P. Huang, C.D. Simpson, S.G. Pattenden, J.L. Norris, D.B. Kireev, A. Tripathy, A. Edwards, B.L. Roth, W.P. Janzen, B.A. Garcia, A. Petronis, J. Ellis, P.J. Brown, S.V. Frye, C.H. Arrowsmith and J. Jin, Nat Chem Biol, 7 (2011) 566.
[9] A.J. Ruthenburg, H. Li, D.J. Patel and C.D. Allis, Nat Rev Mol Cell Biol, 8 (2007) 983.
[10] S.D. Taverna, H. Li, A.J. Ruthenburg, C.D. Allis and D.J. Patel, Nat Struct Mol Biol, 14 (2007) 1025.
[11] M. Vermeulen, H.C. Eberl, F. Matarese, H. Marks, S. Denissov, F. Butter, K.K. Lee, J.V. Olsen, A.A. Hyman, H.G. Stunnenberg and M. Mann, Cell, 142 (2010) 967.
[12] R. Sanchez and M.M. Zhou, Curr Opin Drug Discov Devel, 12 (2009) 659.
[13] S.R. Haynes, C. Dollard, F. Winston, S. Beck, J. Trowsdale and I.B. Dawid, Nucleic Acids Res, 20 (1992) 2603.
[14] C.A. Musselman and T.G. Kutateladze, Mol Interv, 9 (2009) 314.
[15] L.A. Baker, C.D. Allis and G.G. Wang, Mutat Res, 647 (2008) 3.
[16] M.A. Adams-Cioaba and J. Min, Biochem Cell Biol, 87 (2009) 93.
[17] J. Kim, J. Daniel, A. Espejo, A. Lake, M. Krishna, L. Xia, Y. Zhang and M.T. Bedford, EMBO Rep, 7 (2006) 397.
[18] S. Maurer-Stroh, N.J. Dickens, L. Hughes-Davies, T. Kouzarides, F. Eisenhaber and C.P. Ponting, Trends Biochem Sci, 28 (2003) 69.
[19] P. Filippakopoulos, J. Qi, S. Picaud, Y. Shen, W.B. Smith, O. Fedorov, E.M. Morse, T. Keates, T.T. Hickman, I. Felletar, M. Philpott, S. Munro, M.R. McKeown, Y. Wang, A.L. Christie, N. West, M.J. Cameron, B. Schwartz, T.D. Heightman, N. La Thangue, C.A. French, O. Wiest, A.L. Kung, S. Knapp and J.E. Bradner, Nature, 468 (2010) 1067.

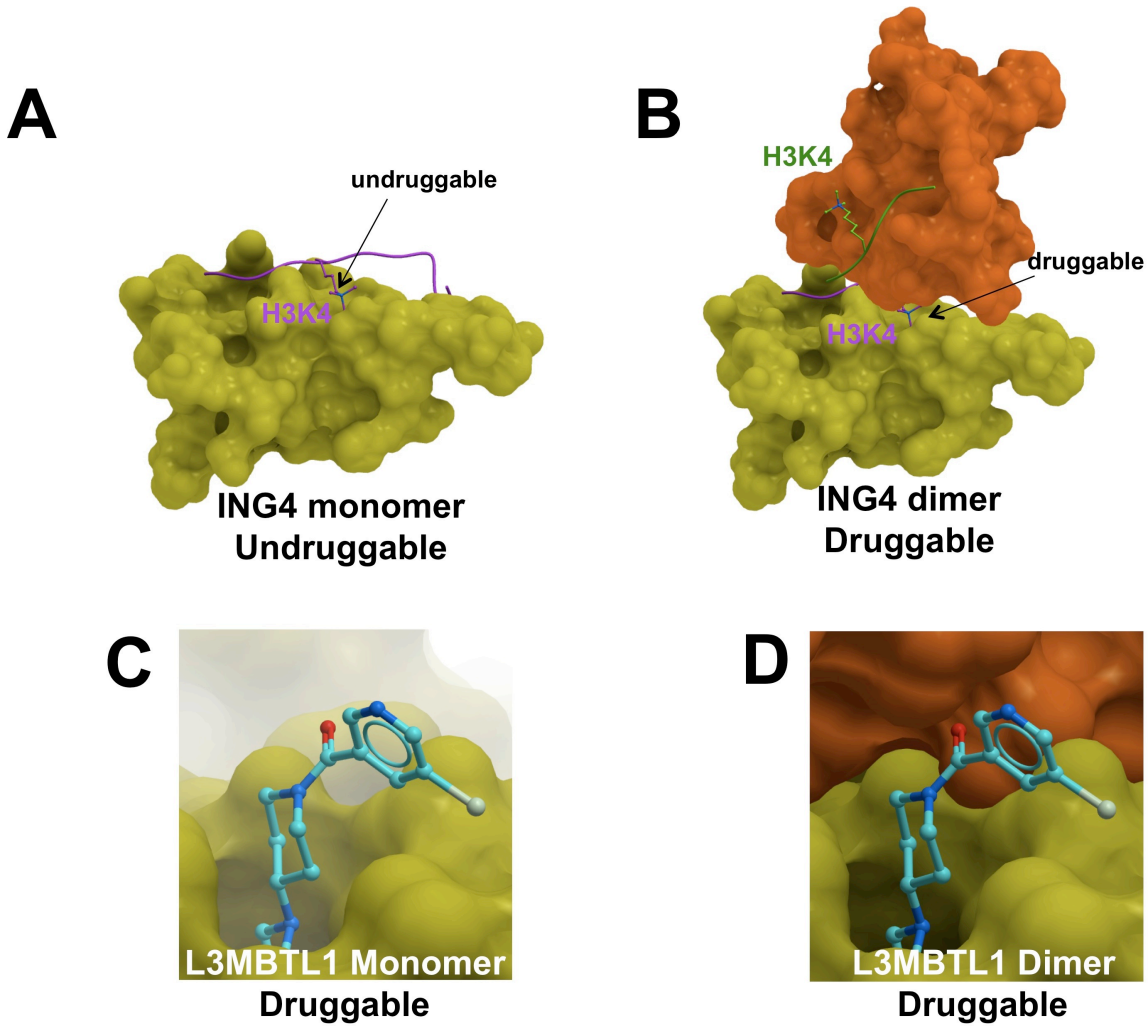
[20] E. Nicodeme, K.L. Jeffrey, U. Schaefer, S. Beinke, S. Dewell, C.W. Chung, R. Chandwani, I. Marazzi, P. Wilson, H. Coste, J. White, J. Kirilovsky, C.M. Rice, J.M. Lora, R.K. Prinjha, K. Lee and A. Tarakhovsky, Nature, 468 (2010) 1119.
[21] J.M. Herold, L.A. Ingerman, C. Gao and S.V. Frye, Curr Chem. Genomics, 4 (2011) 51.
[22] J.M. Herold, T.J. Wigle, J.L. Norris, R. Lam, V.K. Korboukh, C. Gao, L.A. Ingerman, D.B. Kireev, G. Senisterra, M. Vedadi, A. Tripathy, P.J. Brown, C.H. Arrowsmith, J. Jin, W.P. Janzen and S.V. Frye, J Med Chem, 54 (2011) 2504.
[23] M. Wang, M.W. Mok, H. Harper, W.H. Lee, J. Min, S. Knapp, U. Oppermann, B. Marsden and M. Schapira, Bioinformatics, 26 (2010) 2629.
[24] K.L. Yap and M.M. Zhou, Crit Rev Biochem Mol Biol, 45 (2010) 488.
[25] J. Min, A. Allali-Hassani, N. Nady, C. Qi, H. Ouyang, Y. Liu, F. MacKenzie, M. Vedadi and C.H. Arrowsmith, Nat Struct Mol Biol, 14 (2007) 1229.
[26] Y. Guo, N. Nady, C. Qi, A. Allali-Hassani, H. Zhu, P. Pan, M.A. Adams-Cioaba, M.F. Amaya, A. Dong, M. Vedadi, M. Schapira, R.J. Read, C.H. Arrowsmith and J. Min, Nucleic Acids Res, 37 (2009) 2204.
[27] T.A. Halgren, J Chem Inf Model, 49 (2009) 377.
[28] P. Schmidtke, C. Souaille, F. Estienne, N. Baurin and R.T. Kroemer, J Chem Inf Model, 50 (2010) 2191.
[29] L. Kaustov, H. Ouyang, M. Amaya, A. Lemak, N. Nady, S. Duan, G.A. Wasney, Z. Li, M. Vedadi, M. Schapira, J. Min and C.H. Arrowsmith, J Biol Chem, 286 (2010) 521.
[30] Z. Wang, J. Song, T.A. Milne, G.G. Wang, H. Li, C.D. Allis and D.J. Patel, Cell, 141 (2010) 1183.
[31] R.E. Collins, J.P. Northrop, J.R. Horton, D.Y. Lee, X. Zhang, M.R. Stallcup and X. Cheng, Nat Struct Mol Biol, 15 (2008) 245.
[32] J.R. Horton, A.K. Upadhyay, H.H. Qi, X. Zhang, Y. Shi and X. Cheng, Nat Struct Mol Biol, 17 (2010) 38.
[33] A. Palacios, I.G. Munoz, D. Pantoja-Uceda, M.J. Marcaida, D. Torres, J.M. Martin-Garcia, I. Luque, G. Montoya and F.J. Blanco, J Biol Chem, 283 (2008) 15956.



Supplementary Figure 1. The methyl-lysine binding site of EED is at an undruggable face of the canonical donut shape (left – Dscore =0.72 – PDB code 3IIW). On the opposite side, the EZH2 binding site is druggable (right – Dscore=0.98 – PDB code 2QXV).



Supplementary Figure 2. PHD binding sites are generally undruggable, as illustrated by ING1 (A – PDB code 2QIC – no pocket detected) or PYGO (B – PDB code 2VPE – Dscore = 0.69). Two exceptions are MLL1, where ligand-induced backbone motion generates a druggable site (C – PDB code 3LQI – Dscore = 0.97), and PHF8, where the JMJ domain closes onto the methyl-lysine binding site to form a very druggable pocket (D – PDB code 3KV4 – Dscore = 1.05).



Supplementary Figure 3: Homodimerization can alter druggability. Top: The PHD domain of ING4 (yellow) has a shallow and undruggable methyl-lysine binding site (A – Dscore=0.53). A second monomer (orange) bound to a distinct H3K4me3 peptide (green) is observed in the crystal structure, lying on top of the first methyl-lysine binding site, which becomes buried and druggable (B - Dscore=1.01) (PDB code 2VNF). **Bottom: L3MBTL1 dimerization observed in the inhibitor-bound structure significantly increases druggability.** The inhibitor occupies the methyl-lysine binding pocket of one monomer (C – Dscore = 0.86). The presence of a second L3MBTL1 subunit (orange) in the crystal structure results in a more enclosed site with higher Dscore (D – Dscore = 1.06) (PDB code 3P8H).