Electronic Supplementary Material (ESI) for Molecular BioSystems.

Small-molecule binding sites to explore new targets in the cancer proteome

David Xu, Shadia I. Jalal, George W. Sledge Jr., and Samy O. Meroueh*

Supplementary Text

Druggable Binding Sites across all 10 Diseases. Using the previously established cutoffs, we identified genes that were overexpressed across multiple cancer types and featured druggable binding sites. We ranked these genes based on the total number of tumors that overexpressed the gene (**Fig. S1**). Using a simple PubMed query, we then counted the number of articles in which either the gene symbol or gene name was co-mentioned with the term 'cancer'. Most of the most frequently occurring differentially-expressed genes correspond to proteins of well-established cancer targets. Among them are matrix metalloproteinases (MMPs), including *MMP1*, *MMP9*, and *MMP12*, which are implicated in tumor invasion and metastasis (1). There are several protein kinases, including *TTK*, *AURKA, AURKB*, and *PLK1*, that are involved in cell signaling and well-established oncology targets (2). Some genes among this list that have not been extensively studied nor targeted in cancer. These include the serine/threonine kinase *PKMYT1* (*MYT1*) is a regulator of G2/M transition in the cell cycle, but lacks focused small molecule inhibitors that specifically target the kinase. Recent efforts in developing small molecule inhibitors involve repurposing of available kinase inhibitors to specifically target the kinase (3). A subunit of the *GINS* complex *GINS2* (*PSF2*) is involved in cell proliferation and survival in cancer cell lines (4,5). The *GINS* complex plays a role in initiating DNA replication during the cell cycle (6).

Untargeted Proteins with ENZ Binding Sites. In total, we identified 102 ENZ binding sites among the 202 proteins that were both overexpressed and correlated with patient survival. Many of these binding sites previously been targeted by small molecule inhibitors and have co-crystallized structures of the protein with their respective inhibitors. We highlight examples of proteins with ENZ binding sites that have seldom been considered in cancer and lack therapeutics (e.g. *PYCR1*, *QPRT*, *HSPA6*), or are well-studied in cancer but lack small molecule inhibitors (e.g. *PKMYT1*, *STEAP3*, *NNMT*).

The reductase *PYCR1* is involved in oxidative stress and catalyzes the final step of proline biosynthesis. Mutations in this gene have been associated with autosomal recessive cutis laxa (7). The protein forms a homodecamer structure consisting of five homodimers (**Fig. S4A**). The lone binding site is the catalytic site containing a NAD(+) molecule, which binds in the low millimolar range (8). This ENZ binding site is adjacent to the interface of the homodimer subunit. The phosphoribosyltransferase *QPRT* is involved in the catabolism of quinolinate in the de novo NAD(+) biosynthesis pathway from tryptophan (9). The protein is potential therapeutic target in malignant glioma cells (10). *QPRT* forms a hexamer structure consisting of three homodimers (**Fig. S4B**). A lone ENZ binding site was detected on the structure of the monomer, but is distant from any PPI interface on the hexamer structure. The heat shock protein *HSPA6* is part of the ubiquitous Hsp70 family involved in protein folding and protection from stress. Members of this family have been implicated in both cancer initiation and progression, specifically in the regulation of multiple signaling pathways (11). This chaperone protein is only expressed after severe stress, rather than as a 'housekeeper' (12). The ENZ binding site is on the N-terminal ATPase domain (**Fig. S4C**) at the ATP binding site.

As previously mentioned, the serine/threonine kinase *PKMYT1* (*MYT1*) is a regulator of G2/M transition in the cell cycle. Kinome-wide screenings of known inhibitors reveals therapeutics that specifically target the protein with strong binding affinities (13). Among the inhibitors (e.g. dasatinib, bosutinib, PD173955) that bind, *PKMYT1* is considered an off-target. The crystal structure of the protein kinase domain contains both the ENZ binding site shared across all members of this family, as well as an additional OTH binding site near the αC helix (**Fig. S4D**). The metalloreductase *STEAP3* is required for iron homeostasis and TLR4-mediated inflammatory response in innate immunity (14). The STEAP family share common processes in cancer growth and apoptosis (15,16). *STEAP3* specifically is able to maintain tumor growth in hypoferric conditions (17). The protein forms a homodimeric structure and features a lone ENZ binding site at the NAP(+) binding site (**Fig. S4E**). Finally, the methyltransferase *NNMT* catalyzes the methylation of nicotinamide and has been shown to promote cancer migration and survival (18,19). In renal carcinomas, *NNMT* induces invasion by activating *MMP2* (20). In cancer, *NNMT* was shown to regulate protein methylation through epigenetic remodeling (21). This mechanism is suggested to be from upregulation via *STAT3* signaling (22). There is a lone ENZ binding site on *NNMT* is at the methylation site, where a SAH molecule is bound in the crystal structure (**Fig. S4F**).

Untargeted Proteins with PPI Binding Sites. Small-molecule inhibition of protein-protein interactions has been historically challenging due to the lack of well-defined binding sites at the protein-protein interface. We explore the structure of proteins whose overexpression correlates with patient outcome to uncover potential new PPI targets that could be amenable for drug discovery. We identified 46 PPI binding sites on 40 proteins whose overexpression correlate with patient outcome. Among them, 25 binding sites occur on 24 proteins with $\log_2$FC greater than 2.0 (**Fig. S3**). The unique interfaces at these PPIs create an opportunity to develop highly specific compounds that mediate the interaction between the proteins. Indeed, a number occur on proteins that have previously been targeted in cancer therapeutics (e.g. *PLAUR-PLAU* (23) and *IL2RA-IL2* (24)). Here, we highlight examples of proteins with PPI binding sites that have not been previously targeted by small molecule inhibitors and are either seldom considered in cancer (e.g. *CASC5*, *ZBTB32*, and *CSAD*), or are well-studied in cancer but lack small molecule inhibitors (e.g. *HNF4A, MEF2B*, and *CBX2*).

The cell cycle associated protein *CASC5* (*KNL1*) is a potential target in BRCA and features both a PPI and OTH binding site on its structure (**Fig. S5A**). The protein interacts with *BUB1* and *BUB1B* to mediate microtubule attachment during mitosis (25). *CACS5* is part of the larger MIS12 protein complex in kinetochore assembly. The PPI is formed between *CASC5* and the *NSL1* subunit, which is essential as a scaffold to support additional interactions between the MIS12 complex with other protein complexes in kinetochore assembly. An additional allosteric binding site was detected that was directly adjacent to this PPI site. A transcription factor, *ZBTB32*, is overexpressed in KIRC and features one binding site on its BTB/POZ domain. This DNA-binding protein functions as an early repressor to immune processes, including the repression of MHC class II expression during B cell differentiation (26) and the proliferative burst of natural killer cells during viral infection (27). The crystal structure is of a PPI motif, with a single binding site at the known PPI interface of this protein in its dimer form (**Fig. S5B**). A ligase *CSAD* is overexpressed in KIRC and features five binding sites on its crystal structure. In mice, the protein acts as the rate limiting enzyme in taurine biosynthesis (28). Members of the group 2 decarboxylase family are involved in decarboxylation of amino acids. In rats, overexpression of CSAD from hepatocarcinogenesis resulted in production of antibodies against *CSAD* in rats (29). The homodimer structure reveals three binding sites at the PPI interface and an additional two allosteric binding sites (**Fig. S5C**). Two of the PPI binding sites are directly adjacent to one another while the third binding site is distant. In addition, the distant binding site and the PPI binding site closer to this distant binding site are occupied by alpha helices of the binding partner.

The cofactor site occupied by the bound ligand did not have a DrugScore (0.76) above the established cutoff and was not considered.

The transcription factor *HNF4A* forms a homodimer complex to interact with DNA to control the expression of other genes. In the monomer structure, two binding sites were detected on the protein surface (**Fig. S5D**). One of these two sites is bound to a saturated fatty acid in multiple superimposed crystal structures, while the other is at the homodimer interface required for transcription factor activity. An additional PPI site at the coactivator binding site was not detected by the SiteMap program. The transcription factor has been implicated in cancer through Hippo pathway signaling (30), but lacks small molecule inhibitors that target it or other transcription factors in general. Similarly, *MEF2B* belongs to a family of transcription factors that forms a homodimeric structure that binds to DNA. Mutations in *MEF2B* in lymphomas were found to contribute to lymphomagenesis through the deregulation of *BCL6* (31). A binding site was detected at the homodimer interface of *MEF2B* directly adjacent to the DNA-binding site (**Fig. S5E**). No binding sites were detected at the PPI interface with histone deacetylase *HDAC*. As previously mentioned, *CBX2* is a component of the Polycomb protein complex and regulates gene expression during development (32) and proliferation of adult stem cells and cancer cells through chromatin modification (33). Through the expression of *CBX2*, *SMARCE1* suppresses *EGFR* transcription in lung cancer (34). Meta-analysis of *CBX2* against a variety human cancers showed significant correlations with metastatic progression and overall survival (35). The crystal structure of the chromodomain of *CBX2* reveals a conserved binding site at the PPI interface with a histone peptide (**Fig. S5F**). Members of this family share similar overall structures at this domain, but distinct residues in the peptide binding site contribute to the binding affinity to a specific histone for epigenetic modification (36).

Untargeted Proteins with OTH Binding Sites. OTH binding sites can provide an avenue to modulate either enzymatic function or protein-protein interactions of the target. Compounds that bind to OTH sites could act either in an orthosteric manner if the binding site happens to be the binding site of a substrate or protein, or allosterically if the binding site is outside an enzyme active site or protein binding site. Among the genes whose overexpression strongly correlated with patient outcome and that possessed an OTH binding site, several had never been studied in cancer before nor do they have small molecule inhibitors either in the literature or in co-crystallized complexes. We highlight four examples that span a variety of tumors: a protein of unknown

function *FAM83A*, a water channel *AQP2*, a serine protease *SERPIND1*, and a protein associated with the immune response *TNFAIP8L2*.

The protein encoded by *FAM83A* (family with sequence similarity 83, member A) is among the top-ranked candidates in both LUAD and LUSC, and features a binding site in the center of the only available crystal structure (**Fig. S6A**). The crystal structure consists of a Pfam domain with unknown function. In addition, the function and localization of this protein are still unknown, but it has been implicated in a variety of cancer-related processes. Members of the FAM83 family exhibit oncogenic properties, and *FAM83B* was shown to regulate MAPK signaling (37). *FAM83A* was also shown to confer resistance to EGFR inhibitors in breast cancer cell lines, although the mechanism is still unclear.

The membrane protein *AQP2* ranks among the most promising target in COAD and features two allosteric binding sites on its protein structure. The protein acts as a water channel in the kidney, where it traffics water between the membrane and storage vesicles (38). Mutations to this protein can result in diabetes insipidus. One of the binding sites is formed from the bundle of α-helices, and extends to an area that is directly adjacent to the PPI interface of the homo 4-mer structure (**Fig. S6B**). The other binding site is formed at the opposite end of the channel, which is closed in the crystallized structure.

Another protein harboring several OTH binding sites is the serine protease *SERPIND1.* It is overexpressed in KIRC and features five potentially allosteric sites on its serpin domain. It acts as a thrombin inhibitor in the coagulation cascade and promotes angiogenesis through activation of the AMPK signaling (39). Thrombin is well-studied in cancer and known to induce tumor invasion by acting on integrins and MMPs (40). Its three OTH binding sites are distant of the thrombin-*SERPIND1* complex, while two are adjacent to the PPI interface (**Fig. S6C**).

A tumor necrosis factor, *TNFAIP8L2* (*TIPE2*), acts as a negative regulator of both innate and adaptive immunity through its regulation of phagocytosis and oxidative burst (41). It is overexpressed in KIRC and features only one binding site. By targeting the interaction between Toll-like receptors and Rac GTPases, this protein regulates the activation of the PI3K-Rac pathway (42). In addition, it has been shown to affect Erk 1/2 signaling with respect to cell migration (43). The crystal structure of *TNFAIP8L2* consists of a helical bundle with a deep

binding site partially enclosed in the middle of the protein (**Fig. S6D**). The domain properties as well as the binding site function are relatively unknown, much like the biology of the protein itself.

Binding Sites that have been Previously Targeted with Small Molecules. We identified protein targets that have not been previously explored with small molecules or drugs by conducting a literature search for the 202 proteins whose mRNA levels were overexpressed ($\log_2$ FC ≥ 1.5) and correlated with worse patient survival (HR > 1). A protein is considered to be targeted if one of the identified binding sites contained a co-crystallized inhibitor. We identified 26 proteins with co-crystallized inhibitors bound to one of the previously identified druggable binding sites (**Table S4**). The majority of these co-crystallized binding sites were located at enzyme active sites. It is worth mentioning that other proteins among the 202 have been actively targeted, albeit at a site that did not score above the cutoffs or lacked a co-crystallized structure of the inhibitor. Interestingly, many of these co-crystallized structures occur at binding sites at or below our higher DrugScore cutoff of 1.0.

In addition, we identified several proteins that have been targeted with small molecules, which had binding sites that occurred at protein-protein interaction interfaces. Among these PPI binding sites, two have solved co-crystallized inhibitors in the binding site. The *ATAD2-HIST1H4A* interaction occurs at a bromodomain and has been subject to fragment based screening (44). Similarly, the ATP binding site is the site of the *GSG2-HIST2HA* interaction, and has been targeted by various kinase inhibitors (45). Other interactions have been previously targeted but do not have co-crystallized inhibitors, including the *PLAUR-PLAU* (46) and *PLAUR-VTN* (23) interactions, or have only been targeted by short peptides (e.g. the *EPHB2*-Ephrin ligand interaction (47)).

Analysis of the Three-Dimensional Structure of Proteins Harboring Binding Sites. In addition to exploring the biological function of individual targets, we explored whether the three-dimensional structure of proteins that contain druggable binding sites possessed a known fold (**Table S5**). Protein domains are often conserved among members of a gene family and can function independently of the rest of the protein sequence. This is most notable in protein kinases, where the ATP binding site is heavily conserved among its 518 members. Using Pfam (48), we mapped the residues surrounding each binding site to its protein domain. We find that these binding sites are mainly parts of the protein kinase, serpin, kinesin, and peptidase domains. However, a large portion of these binding sites were not mapped to a specific domain. We next

examined those proteins without co-crystallized small-molecule inhibitors and find stark contrasts to the overall distribution. In contrast to the binding site distribution on the domains of all identified proteins, many of the well-studied and often targeted protein domains are removed. Among them are the protein kinases by cancer therapeutics (2) and trypsins by both macromolecules and organic small molecules (49). Serpins, which are natural serine protease inhibitors, were among the most frequently untargeted domains. Similarly, the majority of binding sites across these targeted and untargeted proteins are classified as OTH, with only a few enzyme active sites and PPI binding sites scattered among the various protein domains. Especially in well-studied systems where the active site is known, these other binding sites provide opportunities for allosteric inhibition of the protein.

Structural Features Surrounding Binding Sites. We next looked at the secondary structure of residues that compose the individual binding sites of these proteins across their individual binding site annotations. Similar to the approach used to identify the residues around the binding site to determine the binding site's location respective to the protein's domains, we collected the secondary structure annotations of the individual residues from DSSP (50). By examining the residues around a binding site, we generalized the type of secondary structures that were used to construct the binding site itself (**Fig. S7**). The majority of binding sites identified were a mixture of secondary structures or random coils among all proteins with or without small molecule inhibitors. Combined, these two secondary structures generally making up the large majority of all binding sites in each binding site type. In addition, the distributions of binding sites across both groups of binding sites are similar. In each case, the least frequently observed secondary structure among these binding sites were the helix-like (i.e. α-helix, 3_{10} helix, or π-helix) and sheet-like structures (i.e. beta bridges and beta bulges). In both ENZ and PPI binding sites, binding sites consisting of helix-like and sheet-like structures are roughly even, while in OTH binding sites, there are many more binding sites consisting of helix-like structures than sheet-like structures. Among the binding site types, only those classified as PPI show distinctly more binding sites composed of coil-like structures than those consisting of a mixture of secondary structures. This is in contrast to the previously observed secondary structures of interaction partners in PPI binding sites, where while the majority of interacting residues of the interaction partner consisted of random coils, many more helical type residues occurred in the PPI binding sites.

In addition, we examined the secondary structures of the residues of the binding partner inside PPI binding sites, since the physicochemical properties of the secondary structure is often used

in the design of new PPI inhibitors (51) (**Fig. S7**). Similar to the method we used to identify the secondary structure characteristics of residues on the interaction partner, we determined the secondary structure composition of the binding partner within the binding site by identifying the residues within a 5 Å radius of the SiteMap spheres and identifying the secondary structure of these residues from DSSP (50). About 27 and 46% of the residues of the binding partners in the binding site were coil-like and helical (α-helix, 3_{10} helix, or π-helix), respectively. Only 10% of the binding sites were characterized by strand-like structures (β-sheet or β-bridge). The remaining PPI binding sites were a combination of these.

Supplementary Legends

Table S1. Proteins with $\log_2$ fold change greater than 1.5 classified by disease.

Table S2. Proteins with $\log_2$ fold change greater than 1.5 and hazard ratio greater than 1 classified by disease.

Table S3. Proteins with PPI binding sites.

Table S4. Binding sites targeted by small-molecule inhibitors.

Table S5. Most frequently occurring protein domains among druggable binding sites.

Fig. S1. Proteins with druggable binding sites that are overexpressed in multiple cancer types. Proteins ($\log_2$FC ≥ 2.0) with druggable binding sites (DrugScore > 1.0) were ranked by the total number of tumors that overexpressed the specific gene. A PubMed query was used to estimate the number of times the gene was co-mentioned with 'cancer'. A heatmap shows the relative fold change for overexpressed genes across the 10 cancer types.

Fig. S2. Examples of binding site annotations. Proteins are represented in cartoon format. The monomer structure with binding sites present is in white. SiteMap sites are shown as spheres, bound ligands are shown as sticks. A, Enzyme (ENZ) site occupied by a bound inhibitor on the protein kinase domain of *AURKB* (PDB: 4af3.A). B, PPI site at the interface of *CCNE1* (PDB: 1w98.B) with *CDK2* (green). C, OTH (Non-ENZ, non-PPI) site on *ADA* (PDB: 3iar.A).

Fig. S3. Classification of enzyme types by EC codes. Binding sites that were classified as enzyme (ENZ) through manual annotation via UniProt and Catalytic Site Atlas were classified using the protein's EC codes. Binding sites were filtered using SiteScore and DrugScore greater than 0.8. Druggable binding sites feature a more stringent DrugScore cutoff of 1.0. While kinases are normally classified as part of the transferase family, here we have separated the two.

Fig. S4. Examples of untargeted proteins with ENZ binding sites. Proteins are represented in cartoon format. The monomer structure with binding sites is in white. SiteMap binding sites are shown as spheres, bound ligands are shown as ball-and-sticks. **A**, *PYCR1* (PDB: 2izz.A) as a homodimer with an ENZ binding site (peach) with bound nucleotide. **B**, *QPRT* (PDB: 3lar.E) as a hexamer with an ENZ binding site (peach). **C**, *HSPA6* (PDB: 3fe1.B) with an ENZ binding site occupied by the bound nucleotide. **D**, *PKMYT1* (PDB: 3p1a.A) with an OTH binding site (green) and an ENZ binding site occupied by a structurally aligned structure of the ATP nucleotide (PDB: 1ATP). **E**, *STEAP3* (PDB: 2vns.B) as a homodimer with an ENZ binding site occupied by the bound nucleotide. **F**, *NNMT* (PDB: 2iip.A) with an ENZ binding site occupied by the bound amino acid derivative.

Fig. S5. Examples of untargeted proteins with PPI binding sites. Proteins are represented in cartoon format. The monomer structure with binding sites is in white. SiteMap binding sites are shown as spheres, bound ligands are shown as ball-and-sticks. **A**, *CASC5* (PDB: 4nf9.B) with an OTH binding site (peach) and a PPI binding site (green) at the PPI interface with *NSL1* (PDB: 4nf9.D, yellow). **B**, *ZBTB32* (PDB: 3m5b.B) with a PPI binding site (peach) at the PrePPI predicted interface with *BCL6* (PDB: 1r29.A, pink). **C**, *CSAD* (PDB: 2jis.A) with three PPI binding sites at the homodimer interface (blue, orange, peach) and two additional allosteric OTH binding sites (yellow, green). **D**, *HNF4A* (PDB: 4iqr.F) with a PPI binding site (green) at the homodimer interface and an additional OTH binding site (peach). **E**, *MEF2B* (PDB: 1tqe.R) and its PPI binding site (peach) at the homodimer interface. **F**, *CBX2* (PDB: 3h91.A) and its PPI binding site (peach) at the interface with *HIST1H3A* (PDB: 3h91.C, magenta).

Fig. S6. Examples of untargeted proteins with OTH binding sites. Proteins are represented in cartoon format. The monomer structure with binding sites is in white. SiteMap binding sites are shown as spheres, bound ligands are shown as ball-and-sticks. **A**, *FAM83A* (PDB: 4urj.D) and its OTH binding site (blue). **B**, Homotetramer structure of *AQP2* (PDB: 1mjo.A) and its two OTH binding sites (peach, green). **C**, The protein complex between *SERPIND1* (PDB: 1jmo.A) and *F2* (1jmo.H, 1jmo.L) features five OTH binding sites (green, pink, blue, peach, yellow). **D**, TNFAIP8L2 (PDB: 2hxp.A) and its OTH binding site (peach).

Fig. S7. Secondary structure composition of residues surrounding PPI binding sites PPI binding sites. The secondary structure composition of both the binding site and the binding partner within the binding site was identified by creating a 5 Å sphere around the center of each binding site. Secondary structures were obtained from DSSP and combined based on whether the residues were primarily helix-like (i.e. α-helix, 3_{10} helix, or π-helix), sheet-like (i.e. beta bridges and beta bulges), random coils, or a mixture of these types.

Supplementary References

1. Deryugina, E. I., and Quigley, J. P. (2006) Matrix metalloproteinases and tumor metastasis. *Cancer Metastasis Rev.* **25**, 9-34
2. Zhang, J., Yang, P. L., and Gray, N. S. (2009) Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9**, 28-39
3. Wichapong, K., Rohe, A., Platzer, C., Slynko, I., Erdmann, F., Schmidt, M., and Sippl, W. (2014) Application of docking and QM/MM-GBSA rescoring to screen for novel Myt1 kinase inhibitors. *J. Chem. Inf. Model.* **54**, 881-893
4. Zhang, X., Zhong, L., Liu, B. Z., Gao, Y. J., Gao, Y. M., and Hu, X. X. (2013) Effect of GINS2 on proliferation and apoptosis in leukemic cell line. *Int. J. Med. Sci.* **10**, 1795-1804
5. Gao, Y., Wang, S., Liu, B., and Zhong, L. (2013) Roles of GINS2 in K562 human chronic myelogenous leukemia and NB4 acute promyelocytic leukemia cells. *Int. J. Mol. Med.* **31**, 1402-1410
6. Kang, Y. H., Galal, W. C., Farina, A., Tappin, I., and Hurwitz, J. (2012) Properties of the human Cdc45/Mcm2-7/GINS helicase complex and its action with DNA polymerase epsilon in rolling circle DNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 6042-6047
7. Dimopoulou, A., Fischer, B., Gardeitchik, T., Schroter, P., Kayserili, H., Schlack, C., Li, Y., Brum, J. M., Barisic, I., Castori, M., Spaich, C., Fletcher, E., Mahayri, Z., Bhat, M., Girisha, K. M., Lachlan, K., Johnson, D., Phadke, S., Gupta, N., Simandlova, M., Kabra, M., David, A., Nijtmans, L., Chitayat, D., Tuysuz, B., Brancati, F., Mundlos, S., Van Maldergem, L., Morava, E., Wollnik, B., and Kornak, U. (2013) Genotype-phenotype spectrum of PYCR1-related autosomal recessive cutis laxa. *Mol. Genet. Metab.* **110**, 352-361
8. Meng, Z., Lou, Z., Liu, Z., Li, M., Zhao, X., Bartlam, M., and Rao, Z. (2006) Crystal structure of human pyrroline-5-carboxylate reductase. *J. Mol. Biol.* **359**, 1364-1377
9. Malik, S. S., Patterson, D. N., Ncube, Z., and Toth, E. A. (2014) The crystal structure of human quinolinic acid phosphoribosyltransferase in complex with its inhibitor phthalic acid. *Proteins* **82**, 405-414
10. Sahm, F., Oezen, I., Opitz, C. A., Radlwimmer, B., von Deimling, A., Ahrendt, T., Adams, S., Bode, H. B., Guillemin, G. J., Wick, W., and Platten, M. (2013) The endogenous tryptophan metabolite and NAD+ precursor quinolinic acid confers resistance of gliomas to oxidative stress. *Cancer Res.* **73**, 3225-3234
11. Sherman, M. Y., and Gabai, V. L. (2014) Hsp70 in cancer: back to the future. *Oncogene*
12. Hageman, J., van Waarde, M. A., Zylicz, A., Walerych, D., and Kampinga, H. H. (2011) The diverse members of the mammalian HSP70 machine show distinct chaperone-like activities. *Biochem. J* **435**, 127-142
13. Davis, M. I., Hunt, J. P., Herrgard, S., Ciceri, P., Wodicka, L. M., Pallares, G., Hocker, M., Treiber, D. K., and Zarrinkar, P. P. (2011) Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **29**, 1046-1051
14. Zhang, F., Tao, Y., Zhang, Z., Guo, X., An, P., Shen, Y., Wu, Q., Yu, Y., and Wang, F. (2012) Metalloreductase Steap3 coordinates the regulation of iron homeostasis and inflammatory responses. *Haematologica* **97**, 1826-1835
15. Grunewald, T. G., Bach, H., Cossarizza, A., and Matsumoto, I. (2012) The STEAP protein family: versatile oxidoreductases and targets for cancer immunotherapy with overlapping and distinct cellular functions. *Biol. Cell* **104**, 641-657
16. Gomes, I. M., Maia, C. J., and Santos, C. R. (2012) STEAP proteins: from structure to applications in cancer therapy. *Mol. Cancer Res.* **10**, 573-587

17. Isobe, T., Baba, E., Arita, S., Komoda, M., Tamura, S., Shirakawa, T., Ariyama, H., Takaishi, S., Kusaba, H., Ueki, T., and Akashi, K. (2011) Human STEAP3 maintains tumor growth under hypoferric condition. *Exp. Cell Res.* **317**, 2582-2591
18. Kim, J., Hong, S. J., Lim, E. K., Yu, Y. S., Kim, S. W., Roh, J. H., Do, I. G., Joh, J. W., and Kim, D. S. (2009) Expression of nicotinamide N-methyltransferase in hepatocellular carcinoma is associated with poor prognosis. *J. Exp. Clin. Cancer Res.* **28**, 20
19. Wu, Y., Siadaty, M. S., Berens, M. E., Hampton, G. M., and Theodorescu, D. (2008) Overlapping gene expression profiles of cell migration and tumor invasion in human bladder cancer identify metallothionein 1E and nicotinamide N-methyltransferase as novel regulators of cell migration. *Oncogene* **27**, 6679-6689
20. Tang, S. W., Yang, T. C., Lin, W. C., Chang, W. H., Wang, C. C., Lai, M. K., and Lin, J. Y. (2011) Nicotinamide N-methyltransferase induces cellular invasion through activating matrix metalloproteinase-2 expression in clear cell renal cell carcinoma cells. *Carcinogenesis* **32**, 138-145
21. Ulanovskaya, O. A., Zuhl, A. M., and Cravatt, B. F. (2013) NNMT promotes epigenetic remodeling in cancer by creating a metabolic methylation sink. *Nat. Chem. Biol.* **9**, 300-306
22. Tomida, M., Ohtake, H., Yokota, T., Kobayashi, Y., and Kurosumi, M. (2008) Stat3 up-regulates expression of nicotinamide N-methyltransferase in human cancer cells. *J. Cancer Res. Clin. Oncol.* **134**, 551-559
23. Liu, D., Zhou, D., Wang, B., Knabe, W. E., and Meroueh, S. O. (2015) A New Class of Orthosteric uPAR.uPA Small-Molecule Antagonists Are Allosteric Inhibitors of the uPAR.Vitronectin Interaction. *ACS Chem. Biol.* **10**, 1521-1534
24. Thanos, C. D., Randal, M., and Wells, J. A. (2003) Potent small-molecule binding to a dynamic hot spot on IL-2. *J Am Chem Soc* **125**, 15280-15281
25. Krenn, V., Wehenkel, A., Li, X., Santaguida, S., and Musacchio, A. (2012) Structural analysis reveals features of the spindle checkpoint kinase Bub1-kinetochore subunit Knl1 interaction. *J. Cell Biol.* **196**, 451-467
26. Yoon, H. S., Scharer, C. D., Majumder, P., Davis, C. W., Butler, R., Zinzow-Kramer, W., Skountzou, I., Koutsonanos, D. G., Ahmed, R., and Boss, J. M. (2012) ZBTB32 is an early repressor of the CIITA and MHC class II gene expression during B cell differentiation to plasma cells. *J. Immunol.* **189**, 2393-2403
27. Beaulieu, A. M., Zawislak, C. L., Nakayama, T., and Sun, J. C. (2014) The transcription factor Zbtb32 controls the proliferative burst of virus-specific natural killer cells responding to infection. *Nat. Immunol.* **15**, 546-553
28. Ueki, I., and Stipanuk, M. H. (2007) Enzymes of the taurine biosynthetic pathway are expressed in rat mammary gland. *J. Nutr.* **137**, 1887-1894
29. Kishimoto, T., Kokura, K., Nakadai, T., Miyazawa, Y., Wakamatsu, T., Makino, Y., Nakamura, T., Hara, E., Oda, K., Muramatsu, M., and Tamura, T. (1996) Overexpression of cysteine sulfinic acid decarboxylase stimulated by hepatocarcinogenesis results in autoantibody production in rats. *Cancer Res.* **56**, 5230-5237
30. Fitamant, J., Kottakis, F., Benhamouche, S., Tian, H. S., Chuvin, N., Parachoniak, C. A., Nagle, J. M., Perera, R. M., Lapouge, M., Deshpande, V., Zhu, A. X., Lai, A., Min, B., Hoshida, Y., Avruch, J., Sia, D., Camprecios, G., McClatchey, A. I., Llovet, J. M., Morrissey, D., Raj, L., and Bardeesy, N. (2015) YAP Inhibition Restores Hepatocyte Differentiation in Advanced HCC, Leading to Tumor Regression. *Cell reports*
31. Ying, C. Y., Dominguez-Sola, D., Fabi, M., Lorenz, I. C., Hussein, S., Bansal, M., Califano, A., Pasqualucci, L., Basso, K., and Dalla-Favera, R. (2013) MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. *Nat. Immunol.* **14**, 1084-1092

32. Zhen, C. Y., Duc, H. N., Kokotovic, M., Phiel, C. J., and Ren, X. (2014) Cbx2 stably associates with mitotic chromosomes via a PRC2- or PRC1-independent mechanism and is needed for recruiting PRC1 complex to mitotic chromosomes. *Mol. Biol. Cell* **25**, 3726-3739
33. Pethe, P., Nagvenkar, P., and Bhartiya, D. (2014) Polycomb group protein expression during differentiation of human embryonic stem cells into pancreatic lineage in vitro. *BMC Cell Biol.* **15**, 18
34. Papadakis, A. I., Sun, C., Knijnenburg, T. A., Xue, Y., Grernrum, W., Holzel, M., Nijkamp, W., Wessels, L. F., Beijersbergen, R. L., Bernards, R., and Huang, S. (2015) SMARCE1 suppresses EGFR expression and controls responses to MET and ALK inhibitors in lung cancer. *Cell Res.* **25**, 445-458
35. Clermont, P. L., Sun, L., Crea, F., Thu, K. L., Zhang, A., Parolia, A., Lam, W. L., and Helgason, C. D. (2014) Genotranscriptomic meta-analysis of the Polycomb gene CBX2 in human cancers: initial evidence of an oncogenic role. *Br. J. Cancer* **111**, 1663-1672
36. Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G. A., Li, Z., Vedadi, M., Schapira, M., Min, J., and Arrowsmith, C. H. (2011) Recognition and specificity determinants of the human cbx chromodomains. *J. Biol. Chem.* **286**, 521-529
37. Cipriano, R., Miskimen, K. L., Bryson, B. L., Foy, C. R., Bartel, C. A., and Jackson, M. W. (2014) Conserved oncogenic behavior of the FAM83 family regulates MAPK signaling in human cancer. *Mol. Cancer Res.* **12**, 1156-1165
38. Frick, A., Eriksson, U. K., de Mattia, F., Oberg, F., Hedfalk, K., Neutze, R., de Grip, W. J., Deen, P. M., and Tornroth-Horsefield, S. (2014) X-ray structure of human aquaporin 2 and its implications for nephrogenic diabetes insipidus and trafficking. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 6305-6310
39. Ikeda, Y., Aihara, K., Yoshida, S., Iwase, T., Tajima, S., Izawa-Ishizawa, Y., Kihira, Y., Ishizawa, K., Tomita, S., Tsuchiya, K., Sata, M., Akaike, M., Kato, S., Matsumoto, T., and Tamaki, T. (2012) Heparin cofactor II, a serine protease inhibitor, promotes angiogenesis via activation of the AMP-activated protein kinase-endothelial nitric-oxide synthase signaling pathway. *J. Biol. Chem.* **287**, 34256-34263
40. Radjabi, A. R., Sawada, K., Jagadeeswaran, S., Eichbichler, A., Kenny, H. A., Montag, A., Bruno, K., and Lengyel, E. (2008) Thrombin induces tumor invasion through the induction and association of matrix metalloproteinase-9 and beta1-integrin on the cell surface. *J. Biol. Chem.* **283**, 2822-2834
41. Wang, Z., Fayngerts, S., Wang, P., Sun, H., Johnson, D. S., Ruan, Q., Guo, W., and Chen, Y. H. (2012) TIPE2 protein serves as a negative regulator of phagocytosis and oxidative burst during infection. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 15413-15418
42. Sun, H., Zhuang, G., Chai, L., Wang, Z., Johnson, D., Ma, Y., and Chen, Y. H. (2012) TIPE2 controls innate immunity to RNA by targeting the phosphatidylinositol 3-kinase-Rac pathway. *J. Immunol.* **189**, 2768-2773
43. Zhang, Y. H., Yan, H. Q., Wang, F., Wang, Y. Y., Jiang, Y. N., Wang, Y. N., and Gao, F. G. (2014) TIPE2 inhibits TNF-alpha-induced hepatocellular carcinoma cell metastasis via Erk1/2 downregulation and NF-kappaB activation. *Int. J. Oncol.*
44. Harner, M. J., Chauder, B. A., Phan, J., and Fesik, S. W. (2014) Fragment-based screening of the bromodomain of ATAD2. *J. Med. Chem.* **57**, 9687-9692
45. Chaikuad, A., Tacconi, E. M., Zimmer, J., Liang, Y., Gray, N. S., Tarsounas, M., and Knapp, S. (2014) A unique inhibitor binding site in ERK1/2 is associated with slow binding kinetics. *Nat. Chem. Biol.* **10**, 853-860
46. Mani, T., Wang, F., Knabe, W. E., Sinn, A. L., Khanna, M., Jo, I., Sandusky, G. E., Sledge, G. W., Jr., Jones, D. R., Khanna, R., Pollok, K. E., and Meroueh, S. O. (2013) Small-molecule inhibition of the uPAR.uPA interaction: synthesis, biochemical, cellular, in vivo pharmacokinetics and efficacy studies in breast cancer metastasis. *Biorg. Med. Chem.* **21**, 2145-2155

47. Chrencik, J. E., Brooun, A., Recht, M. I., Nicola, G., Davis, L. K., Abagyan, R., Widmer, H., Pasquale, E. B., and Kuhn, P. (2007) Three-dimensional structure of the EphB2 receptor in complex with an antagonistic peptide reveals a novel mode of inhibition. *J. Biol. Chem.* **282**, 36505-36513
48. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J., and Punta, M. (2014) Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222-230
49. Kobayashi, H., Suzuki, M., Kanayama, N., and Terao, T. (2004) A soybean Kunitz trypsin inhibitor suppresses ovarian cancer cell invasion by blocking urokinase upregulation. *Clin. Exp. Metastasis* **21**, 159-166
50. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637
51. Azzarito, V., Long, K., Murphy, N. S., and Wilson, A. J. (2013) Inhibition of [alpha]-helix-mediated protein-protein interactions using designed molecules. *Nat. Chem.* **5**, 161-173

Symbol	Name	Count	Cancer Publications	BRCA	COAD	GBM	HNSC	KIRC	LUAD	LUSC	THCA	TNBC	UCEC
MMP9	Matrix metalloproteinase-9	8	6,227										
PKMYT1	Membrane-associated tyrosine- and threonine-specific cdc2-inhibitory kinase	8	14										
AURKB	Aurora kinase B	7	500										
BUB1	Mitotic checkpoint serine/threonine-protein kinase BUB1	7	397										
CCNA2	Cyclin-A2	7	312										
MELK	Maternal embryonic leucine zipper kinase	7	63										
MMP1	Interstitial collagenase	7	546										
MMP12	Macrophage metalloelastase	7	113										
PLK1	Serine/threonine-protein kinase PLK1	7	775										
TOP2A	DNA topoisomerase 2-alpha	7	367										
AURKA	Aurora kinase A	6	686										
BLM	Bloom syndrome protein	6	1,123										
CCNB1	G2/mitotic-specific cyclin-B1	6	2,213										
CDC20	Cell division cycle protein 20 homolog	6	296										
CDC25A	M-phase inducer phosphatase 1	6	545										
CHAT	Choline O-acetyltransferase	6	355										
EPHA8	Ephrin type-A receptor 8	6	14										
F12	Coagulation factor XII	6	212										
F2	Prothrombin	6	3,337										
GINS2	DNA replication complex GINS protein PSF2	6	11										
IDO1	Indoleamine 2,3-dioxygenase 1	6	111										
KIF11	Kinesin-like protein KIF11	6	144										
KIF15	Kinesin-like protein KIF15	6	9										
KIF23	Kinesin-like protein KIF23	6	25										
MAD2L1	Mitotic spindle assembly checkpoint protein MAD2A	6	221										
MMP3	Stromelysin-1	6	379										
NEK2	Serine/threonine-protein kinase Nek2	6	115										
SERPINB4	Serpin B4	6	7										
TTK	Dual specificity protein kinase TTK	6	155										

Fig. S1. Proteins with druggable binding sites that are overexpressed in multiple cancer types.

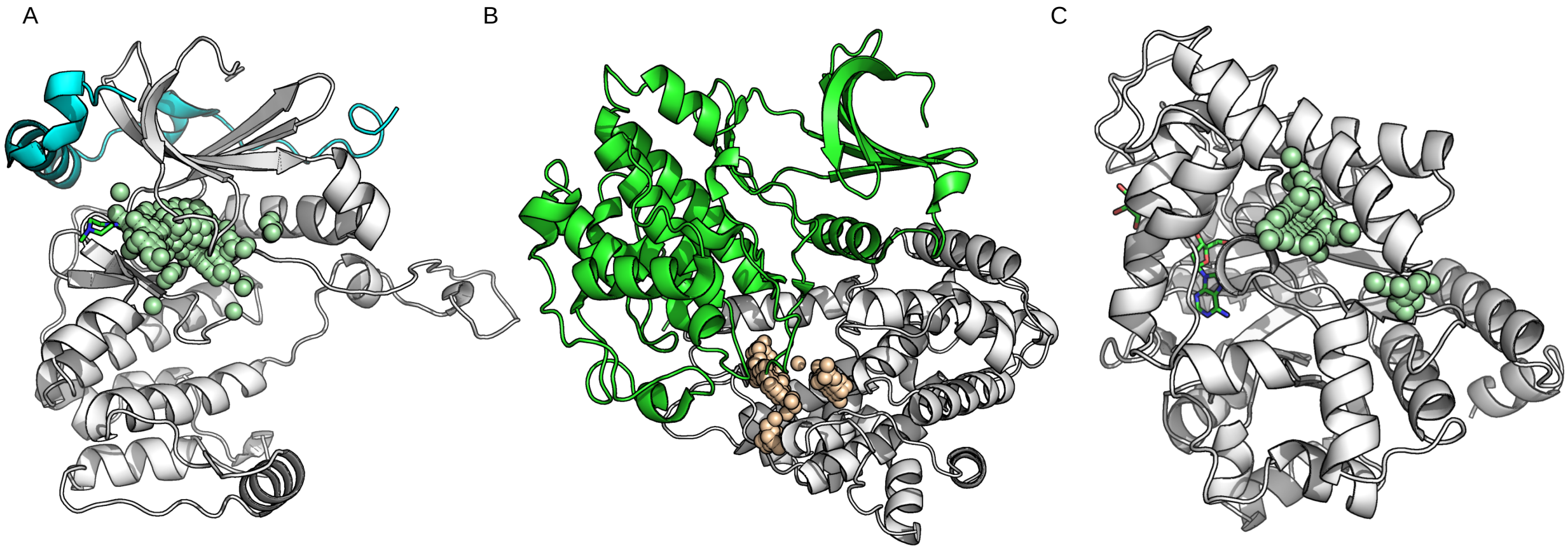

Fig. S2. Examples of binding site annotations.

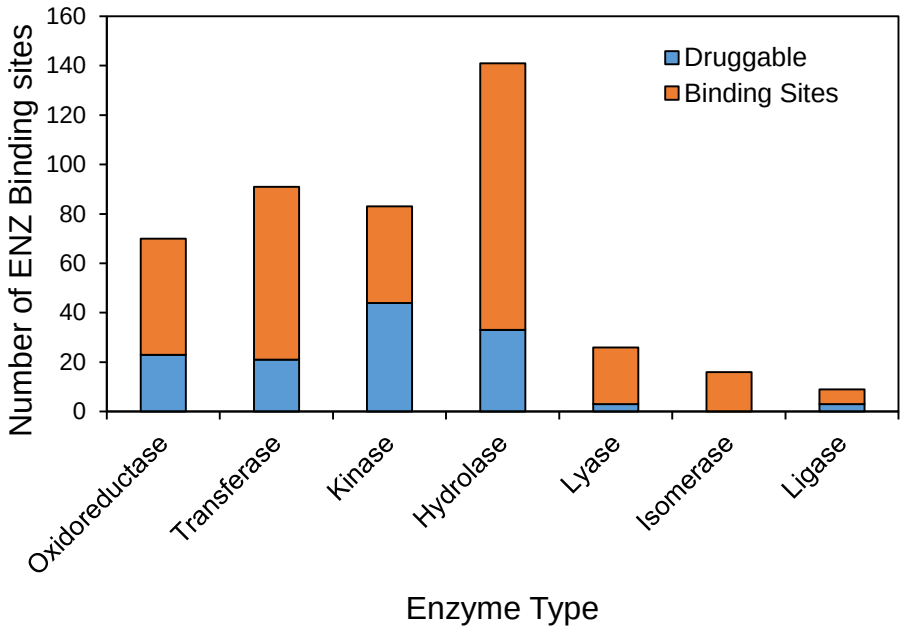

Fig. S3. Classification of enzyme types by EC codes.

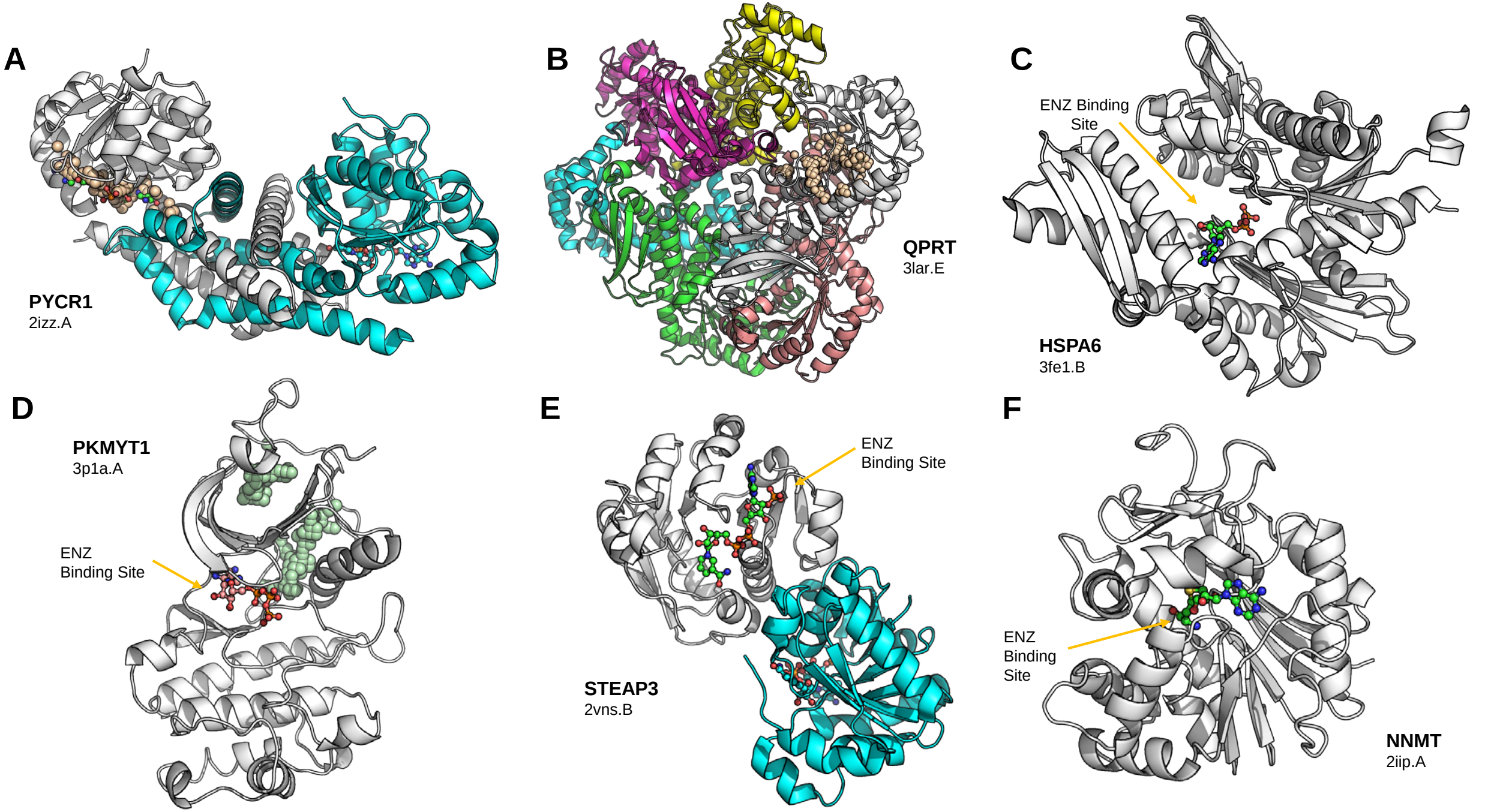

Fig. S4. Examples of untargeted proteins with ENZ binding sites.

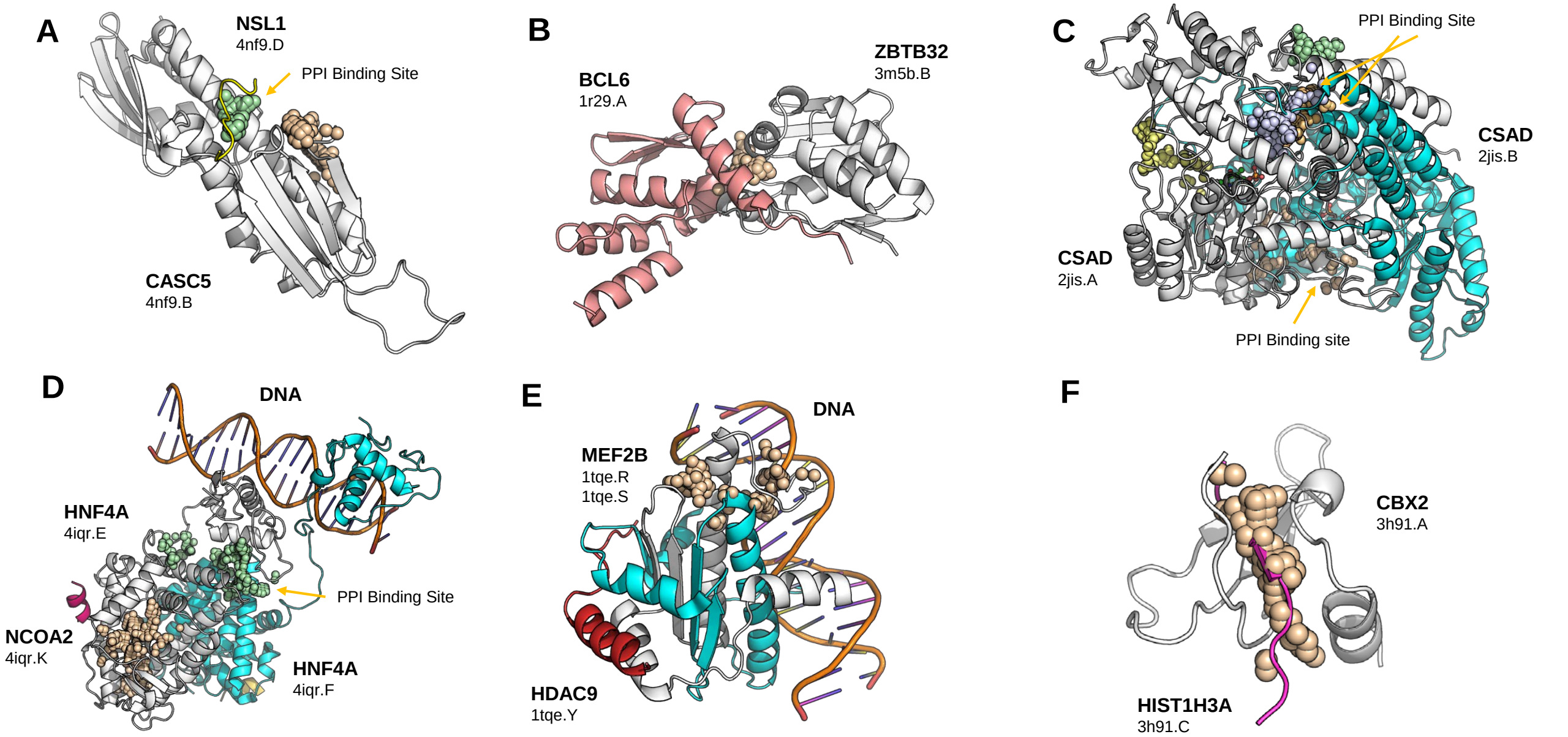

Fig. S5. Examples of untargeted proteins with PPI binding sites.

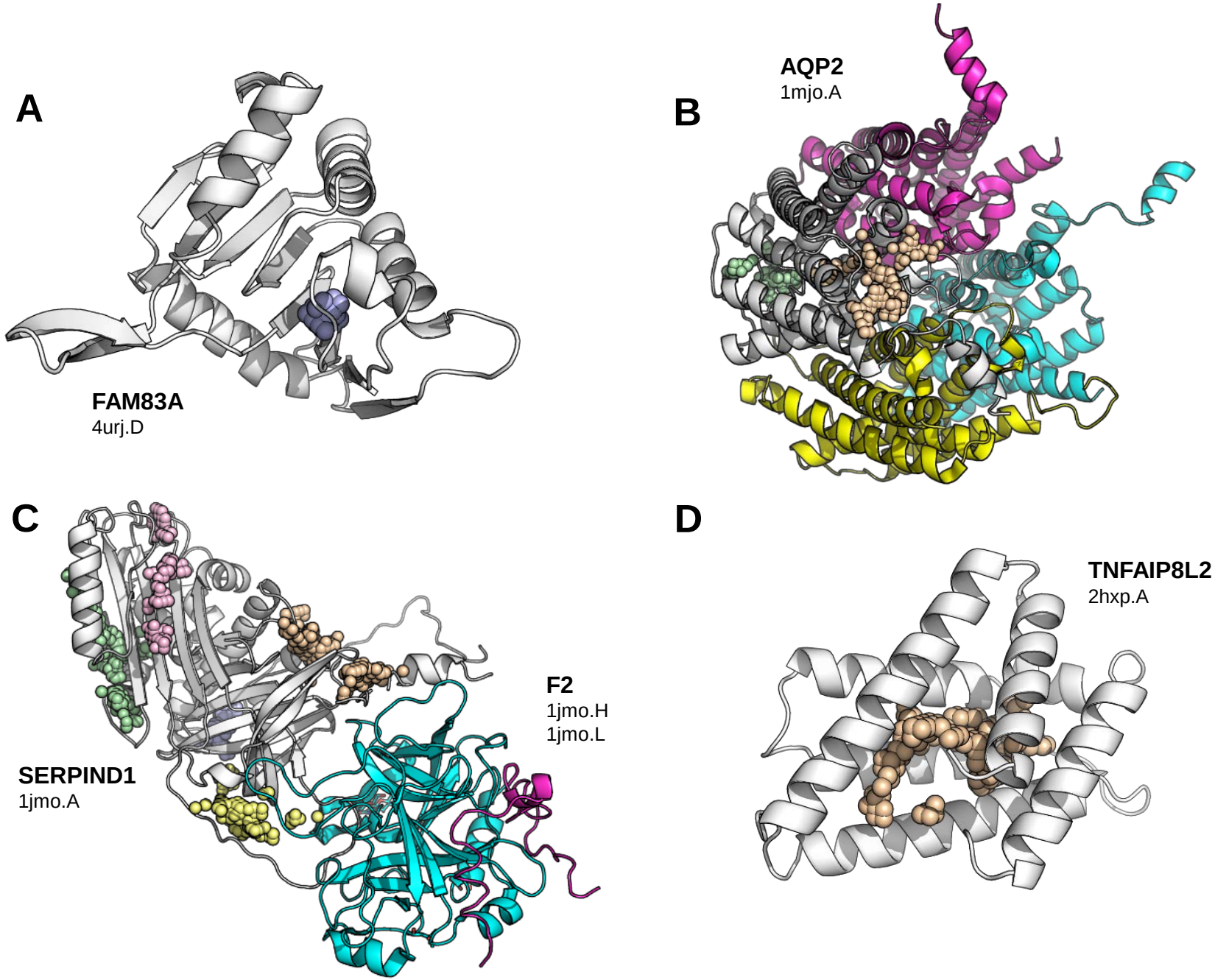

Fig. S6. Examples of untargeted proteins with OTH binding sites.

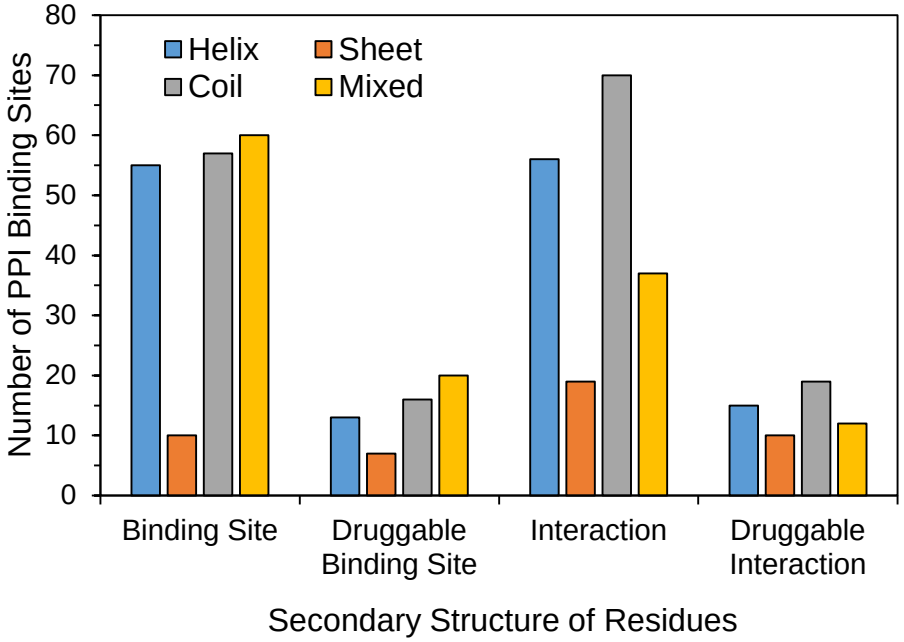

Fig. S7. Secondary structure composition of residues surrounding PPI binding sites PPI binding sites.

Supplementary Table S3. Proteins with PPI Binding Sites

Symbol	Name	Cancer Publications	Binding site	Interaction Partner		
				PDB	Symbol	Name
C3	Complement C3	1,437	2WIIA2‡	2WIIB	C3	Complement C3 Alpha Chain
C3	Complement C3	1,437	2WIIB5‡	2WIIA	C3	Complement C3 Beta Chain
CASC5	Protein CASC5	14	4NF9B2	4NF9D	NSL1	Kinetochore-associated protein NSL1 homolog
CBX2	Chromobox protein homolog 2	15	3H91A1	3H91C	HIST1H3A	Histone H3.1
CCNE1	G1/S-specific cyclin-E1	259	1W98B1‡	1W98A	CDK2	Cyclin-dependent kinase 2
CDA	Cytidine deaminase	951	1MQ0B1	1MQ0A	CDA	Cytidine deaminase
GINS4	DNA replication complex GINS protein SLD5	3	2E9XH2	2E9XE	GINS1	DNA replication complex GINS protein PSF1
GSG2	Serine/threonine-protein kinase haspin	7	3DLZA1†‡	4OUCB	HIST2H3A	Histone H3.2
HIST1H2BO	Histone H2B type 1-O	1	3AV1D1	3AV1C	HIST1H2AB	Histone H2A type 1-B/E
HNF4A	Hepatocyte nuclear factor 4-alpha	162	4IQRE2	4IQRF	HNF4A	Hepatocyte nuclear factor 4-alpha
IGFBP4	Insulin-like growth factor-binding protein 4	173	2DSRG1	2DSRI	IGF1	Insulin-like growth factor I
IL2RA	Interleukin-2 receptor subunit alpha	980	2ERJE1	2ERJH	IL2	Interleukin-2
ITGA5	Integrin alpha-5	111	3VI4A1	3VI4B	ITGB1	Integrin beta-1
MAD2L1	Mitotic spindle assembly checkpoint protein MAD2A	214	2V64F1‡	2V64G	MBP1	
MEF2B	Myocyte-specific enhancer factor 2B	18	1TQER1	1TQES	MEF2B	Myocyte-specific enhancer factor 2B
NCF1C	Putative neutrophil cytosol factor 1C	0	1KQ6A1‡	1O7KB	NCF1C	Putative neutrophil cytosol factor 1C
NCF1C	Putative neutrophil cytosol factor 1C	0	1NG2A1, 2‡	1OV3C		Peptide
PLK1	Serine/threonine-protein kinase PLK1	630	1Q4KB1	1Q4KE		Peptide
RHCG	Ammonium transporter Rh type C	10	3HD6A1, 4, 5	PrePPI	RHAG	Ammonium transporter Rh type A
RNASE2	Non-secretory ribonuclease	6	1K2AA1	2BEXB	RNH1	Ribonuclease inhibitor
TDO2	Tryptophan 2,3-dioxygenase	45	4PW8F1	4PW8E	TDO2	Tryptophan 2,3-dioxygenase
TF	Serotransferrin	2,479	3V8XB4	3V8XA	EIFE4	Eukaryotic translation initiation factor 4E
ZBTB32	Zinc finger and BTB domain-containing protein 32	3	3M5BB1	PrePPI	BCL6	B-cell lymphoma 6 protein

† Has a binding site that has small molecule inhibitors
‡ Binding site DS >= 1.0

Supplementary Table S4. Binding Sites Targeted by Small-Molecule Inhibitors

Symbol	Name	Binding site	SS	DS	Type	Reference
ADAM8	Disintegrin and metalloproteinase domain-containing protein 8	4DD8A1	1.02	0.89	ENZ	4DD8
ADAMTS4	A disintegrin and metalloproteinase with thrombospondin motifs 4	2RJPC1	1.05	1.07	ENZ	2RJP
AKR1B10	Aldo-keto reductase family 1 member B10	4JIIX1	1.10	1.09	ENZ	1ZUA
AURKA	Aurora kinase A	2J4ZB1	1.03	1.04	ENZ	2J4Z
AURKB	Aurora kinase B	4AF3A2	1.05	1.07	ENZ	4AF3
CDC20	Cell division cycle protein 20 homolog	4GGDB2	0.87	0.93	OTH	4N14
CHEK1	Serine/threonine-protein kinase Chk1	2R0UA1	1.02	1.05	ENZ	2R0U
CYP2D6	Cytochrome P450 2D6	3QM4A2	1.12	1.09	OTH	3TBG
GSG2	Serine/threonine-protein kinase haspin	3DLZA1	1.03	1.04	ENZ/PPI	4QTC
JAK3	Tyrosine-protein kinase JAK3	3LXLA1	1.04	0.99	ENZ	3LXL
KLK4	Kallikrein-4	4K8YA1	0.99	0.95	ENZ	4K8Y
MELK	Maternal embryonic leucine zipper kinase	4UMUA1	1.03	1.02	ENZ	4UMU
NEK2	Serine/threonine-protein kinase Nek2	2XK4A1	0.98	0.99	ENZ	2XK4
PARP15	Poly [ADP-ribose] polymerase 15	3GEYA1	1.05	1.08	ENZ	3GEY
PLAU	Urokinase-type plasminogen activator	2VNTD1	0.98	0.98	ENZ	2VNT
PLK1	Serine/threonine-protein kinase PLK1	2OWBA1	1.04	1.06	ENZ	2OWB
PLK4	Serine/threonine-protein kinase PLK4	3COKA1	1.05	0.98	ENZ	4JXF
TK1	Thymidine kinase, cytosolic	2ORVB1	1.06	0.88	ENZ	2ORV
TTK	Dual specificity protein kinase TTK	2ZMDA1	1.03	1.03	ENZ	3HMO
TYMS	Thymidylate synthase	1HW4A1	0.89	0.88	ENZ	1HVY

Supplementary Table S5. Most Frequently Occurring Protein Domains Among Druggable Binding Sites

Name	Proteins	Binding sites	ENZ	PPI	OTH
Binding sites on proteins with no small molecule inhibitor					
None	9	21	0	3	18
Serpin	6	20	1	0	19
Kinesin	6	15	5	0	10
Immunoglobulin	5	7	0	2	5
Protein kinase	5	18	5	0	13
Peptidase	5	11	4	1	7
Protein-tyrosine phosphatase	4	9	1	0	8
Von Willebrand factor type A	4	11	0	2	9
Interleukin	3	4	0	0	4
Cyclin	3	4	0	1	3
Trypsin	3	7	2	0	5
Binding sites on all proteins					
Protein kinase	15	38	15	1	23
None	9	21	0	3	18
Serpin	6	20	1	0	19
Kinesin	6	15	5	0	10
Peptidase	6	13	5	1	8
Trypsin	5	9	4	0	5
Immunoglobulin	5	7	0	2	5
Von Willebrand factor type A	4	11	0	2	9
Protein-tyrosine phosphatase	4	9	1	0	8
Cytochrome P450	3	13	0	0	13
Interleukin	3	4	0	0	4
Cyclin	3	4	0	1	3