Supplementary Information Appendix

Exploring the Structural Origins of Cryptic Sites on Proteins

Dmitri Beglov[a], David R. Hall[b], Amanda Wakefield[a,c], Lingqi Luo[d], Karen N. Allen[c], Dima Kozakov[a,e,f,1], Adrian Whitty[c,1], and Sandor Vajda[a,c,]

[a]Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215;
[b]Acpharis Inc., Massachusetts 01746;
[c]Department of Chemistry, Boston University, Boston, Massachusetts 02215;
[d]Program in Bioinformatics, Boston University, Boston, Massachusetts 02215;
[e]Department of Applied Mathematics and Statistics, Stony Brook University NY, USA
[f]Laufer Center for Physical and Quantitative Biology, Stony Brook University NY, USA

Content:

References to Supplementary Information

Supplementary Figures

Figure S1. Characterization of the extended CryptoSite set.

Figure S2. Examples of hot spots near cryptic sites**.**

Figure S3. Hot spots of Cyclin Dependent Kinase 2 (CDK2).

Table S1. Mapping results for the unbound structures in the CryptoSite and extended CryptoSite sets.

No	Unbound	Bound	Lig	Hot spots (CSs)	Dist.	Comment
1 S	3CHEA 1W9PA 3CHEB 3CHFB			0(21) 0(23) 0(17) 0(22)	0.83 0.11 0.24 0.10	Chitinase B1 enzyme (1). In the unbound structure 3CHEA, W384 protrudes into the site and would clash with the ligand, superimposed into the structure, whereas W137 turns away from the ligand and makes the pocket much more open. The largest hot spot, CS0(20), is in the pocket, but shifted 1.09 Å from the ligand. The repositioning of the W52, W384, and W137 side chains can be spontaneous variation, as the conformation is seen in many unbound structures. E.g., in 3CHFB W384 is positioned well for ligand binding, but W137 is open. In 1W9PA W137 has two rotamers, one is clashing with the ligand, the other creates a good open site. *This flexibility implies that the side chains compete with ligand binding, reducing affinity.*
		2IUZB	D1H	1(16), 2(13)	0.16	In the bound structure 2IUZB the planar ligand D1H is perfectly aligned between the side chains of W52, W137, and W384.This requires a slight turn of W384, and the the closing of W317 on the ligand. Ki=2800 nM (2).
2 S	2AKAA 2XELA 2JJ9			0(17), 4(8) 0(20) 0(18)	2.11 0.69 0.81	Myosin II. In 2AKAA, the side chains of L262 and Y634 protrude into the very narrow binding site, and the largest hot spot 0(17) is at the entrance of the pocket. The side chain of L262 in the alternative unbound structures 2XELA and 2JJ9 also protrudes into the pocket. Thus, ligands have to compete with the side chains, and binding requires induced fit, which may reduce binding affinity. Changes in backbone are small.
		1YV3A	BIT	1(14)	0.15	Pocket in bound structure is very narrow, the inhibitor blebbistatin is planar (3). IC50 = 4900 nM.

3 LC	2GFCA 4DFZE			0(18), 1(16), 4(13) 0(26), 1(20)	0.12 0.11	cAMP-dependent protein kinase. In the unbound structure 2GFCA the DFG loop 51-56 protrudes into the cryptic site, but it is in the domain-domain interface, and mapping after domain split yields 3 strong hot spots in the DFG and ATP sites.
		2JDSA	L20	0(20), 1(16)	0.09	The DFG loop avoids the site in the bound structure 2JDSA, *and also in many unbound structures such as 4DFZE, resulting in a strong hot spot close to the ligand.* IC50=27 nM, Ki=6.3 nM (2).
4 S	1ALBA 1LIBA 5C0NB 1AB0A			0(26) 0(22) 0(25) 0(16)	0.20 0.70 1.41 0.23	Adipocyte lipid-binding protein. The side chain of F57 protrudes into the site, and compete with ligand binding. The main hot spot is not affected. F57 also protrudes into the pocket in the alternative unbound structures 1LIBA and 5C0NB. Thus, since F57 must be moved to the side, ligand binding competes with the side chain.
		1LICA	HDS	0(24), 1(20)	0.06	HDS is a lipid. Kd = 1700-–2000 nM (4).
5 LC	1NEPA 2HKAA			N/A 0(24)	 0.23	NPC2 is a lysosomal protein deficient in Niemann-Pick type C2 disease. In the unbound structure 1NEPA loop 96-103 protrudes into the ligand binding site, and there is no hot spot. Y100 and F66 deeply protrude into the site. The structure 2HKA has 3 chains in the unit cell. Chain C binds the ligand cholesterol-sulfate (C3S). Chain A of the same structure has no bound ligand, but the 96-103 loop is substantially further from the site. However, F66 still protrudes into the site and would class with the ligand, thus there is no proof the the site can become open without ligand binding.
		2HKAC	C3S	0(26)	0.23	Binds cholesterol sulfate, Kd = 30-50 nM (5).

6 LC	3MN9A 3B63B 2Q1NB			1(19), 2(12) 0(24), 1(16) 0(16)	0.07 0.16 0.14	Monomeric actin. Loop 166-171 is close to the site in the unbound structure and hence Y169 protrudes into the site. In addition, the C-end F375 of the unbound structure would clash with the ligand. However, the second strongest hot spot is not much affected. Residue 375 is missing in the alternative unbound structure 3B63B, resulting in a stronger hot spot, but otherwise the unbound structures are similar, no indication of conformational selection.
		3EKSA	CY9	1(17), 2(17)	0.31	Fungal toxin cytochalasin D.
7 S	1ALVA 1NX2A			3(15), 4(7) 0(23), 1(17), 2(16)	1.05 0.26	Calpain Domain VI (calcium-dependent protease, small subunit). The side chain of Q173 protrudes into the binding site. In addition, helix 115-127 slightly turns, which moves V125 into the site. Domain split provides 3(15). In the bound structure both Q173 and V125 move away. The alternative unbound structure 1NX2A was also mapped with domain split. It has V125 out of the site, but not Q173. Nevertheless, it has a stronger hot spot 0(23) than 1ALVA
		1NX3A	ISA	0(30)	0.20	Calpain inhibitor PD150606. Binds with high affinity (Ki=210 nM) to domains I and II (http://www.emdmillipore.com/), but no information on binding to domain VI.
8 LO	2YQCA			0(18)	0.21	Uridine-diphospho-N-acetylglucosamine pyrophosphorylase. In the unbound structure 2YQCA the pocket is more open than in bound form.
		2YQSA	UD1	0(23)	0.05	Three loops approaching the ligand molecule close on the active site when ligand is bound. *The mechanism is induced fit.* In addition, K421 acts as a co-factor. Uridine-diphospho-N-acetylglucosamine (UD1) is a reaction product (7).

9 LC	2QFOB 1YESA 1UYLA			0(17), 2(12), 3(11) 0(28), 1(20) 0(18)	0.16 0.33 0.50	Hsp90 (heat shock protein 90) is a chaperone protein. Residues 106 to 110 form a loop in the unbound structure 2QFO that reduces the size of the pocket. The loop becomes a helix in the bound structure 2WI7A, farther away from the site. In the alternative unbound structure 1YESA the helix is partially formed, and the remaining loop region is farther from the ligand binding site, allowing for a larger pocket.
		2WI7A	KL	0(18)	0.34	Inhibitor, IC50=58 nM.
10 S	1RTCA 1UQ4A 1IFTA 1IL5B 1OBSA			0(23), 1(15), 4(9) 1(20) 0(20) 0(24) 0(22)	2.38 2.34 2.79 2.37 0.31	Ricin is a carbohydrate-binding protein. In the unbound structure 1RTCA the side chain of Y80 protrudes into site, pushing 0(23) 2.38 Å away. The same side chain enters the site in the alternative unbound structures 1UQ4A, 1IFTA, and 1IL5B. In all cases the strong hot spot is at the entrance of the pocket, overlapping with the ligand. In 1OBSA the side chain is turned, and allows for the hot spot to enter the pocket, but it still would clash with the ligand. Thus, Y80 competes with ligand binding.
		1BR6A	PT1	2(13), 4(11), 5(10)	0.34	Pteroic acid inhibitor. Hot spots are relatively weak, but cover most of the ligand. IC50 = 600,000nM (6).
11 U	1RDWX			1(18), 2(16)	3.34	Monomeric actin, alpha skeletal muscle. In the unbound structure 1RDWX the carboxyl end 373-375 is *unordered and protrudes into the site.* The main hot spot is located at 3.34 Å after domain split. No other unbound structure with hot spots nearby.
		1J6ZA	RHO	1(11)	0.20	Tetramethylrhodamine-5-maleimide, a cell-permeable thiol reactive fluorescent probe. Imaging agent, need for induced fit, poorer pocket, *weak binding.*

12 LC	1TQOA			0(27), 4(8)	0.27	Staphylococcal nuclease. Loop 114-118 protrudes into the site in the unbound structure, but would not interfere with the binding of the ligand that occupies only part of the pocket. On the other hand. Y112 of the bound structure seems to support ligand binding. The largest hot spot still overlaps with the ligand. The loop moves away in the bound structure 1TR5A, opening the pocket further
		1TR5A	THP	0(19)	0.24	Thymidine-3',5'-diphosphate inhibitor. Ki ~ 100 nM (8).
13 LO	3PUWE 5GS2A 5GS2A 4O4BA 4O4BA			0(19), 1(16) 0(26) 2(16) 0(18) 1(17)	2.11 2.82 0.40 0.39 4.03	Maltodextrin/maltose-binding protein. Several loops (11-17, 142-154) are much closer to the ligand in the bound form. In the alternative unbound structure 5GS2A the loops are very similar to those in the bound structure 1FQCA, whereas in 4O4BA those are like in 3PUWE. Thus, *conformational changes in the loop appear to be spontaneous, hot spot is moving freely between close and far.*
		1FQCA	GLO	1(19)	0.23	D-glucose in linear form. Kd = 200 nM (9).
14 LO	3L7UC 3L7UC3 L7UA			0(25) 1(19) 0(20)	2.31 0.31 0.64	Human nucleoside diphosphate kinase A. The pocket in the unbound structure 3L7UC is too open. While 1(19) overlaps with the ligand, 0(25 is at a different part of the large site. However, the pocket is even more open in 3L7UA, but the main hot spot is closer. *No unbound structure is closed, so this may be induced fit.*
		2HVDC	ADP	0(20) 1(19)	2.53 0.26	In the bound structure 2HVDC loop 52 to 60 and helix 61 to 70 move closer, creating a binding site. However, the site is an alongated crevice, and the strongest hot spot remains too far from the bound ligand. Ligand position does not overlap with strongest hot spot. ADP is reaction product, weak binding, Kd = 4000 nM.

15 LC	1R1WA 1R1WA 2G15A 2G15A			0(19) 1(18) 0(21) 1(18)	1.43 0.35 1.90 0.23	Tyrosine kinase domain of the hepatocyte growth factor receptor C-MET. Loop 1220-1230 of the unbound structure 1R1WA would clash with the ligand, whereas the helix 1118-1130 is too far out. 1(18) is in the site when domains are split. Moving the side chains using FTFlex yields CS0(23), CS1(15). The alternative unbound structure 2G15A shows the extreme flexibility of the 1228-1244 loop. It was also mapped after domain split. Loops 1225-1243 and 1115-1118 are not visible in the bound structure 3F82A.
		3F82A	353	0(18), 1(18)	0.07	Inhibitor, IC50 = 4.6 nM (6).
16 LM	1G4EB 1G4EB			0(18), 1(16)	0.21 1.33	Thiamin phosphate synthase. *Loop 1156-1161 and residue 1112 are missing in unbound structure.* The loop is closing down on the ligand in the bound structure and defines the pocket.
		1G67B	POP/ TZP	2(12), 3(10)	0.23	Reaction intermediates 4-methyl-5-hydroxyethylthiazole phosphate and pyrophosphate 2. Affinity is not clear.
17 U	3F74C 1LFAA 1ZOOA 1ZOOB			3(16) 0(24), 1(16) 0(17) 0(25)	0.21 0.52 0.32 0.21	Integrin alpha-L. Disordered carboxyl end protrudes into site, which has only the 4th strongest hot spot. The bound structure 3BQMC and the unbound structure 1LFAA have very similar conformations with the carboxyl end far from the ligand binding site, which has the main hot spot for these structures.
		3BQMC	BQM	0(22), 1(19)	0.10	IC50 = 2 nM (2).
18 LC	1MY1C 4U2RC 1FTOA 4U2RA			2(13) 0(38), 1(22) 0(29), 1(16) 0(23)	0.56 0.44 0.24 1.67	Glutamate receptor subunit 2. Several loops (139-143, 68-73) protrude into the site, but move further away upon ligand binding. Alternative unbound structures 4U2RC and 1FTOA have no bound ligand, but are very close to the bound structure 1FTLA, *suggesting a conformational selection model.* The lack of strong hot spot in 1MY1C also indicates that induced fit is less likely, but also suggests weaker binding
		1FTLA	DNQ	0(27), 1(19)	0.11	IC50 = 998 nM (2).

19 I	2WGB 2WGQA 1OACA			3(8) 0(21) 0(18)	5.20 2.50 2.73	Copper amine oxidase. Interdomain cryptic site. All mappings with domain split. In addition, residue Y466 is modified in the bound structure 1D6YB to L-peptide linking. Thus, the sequences differ between unbound and bound structures.
		1D6YB	HY1	0(18)	2.70	Reaction product phenylacetaldehyde, weak binding. Mapped with domain split.
20 S	1DUBD 2DUBC			1(14), 2(12) 2(17)	0.70 0.12	Enoyl-COA hydratase. L117 of the unbound structure protrudes into the pocket. Domain split helps, but the hot spot has only 14 probe clusters. In the alternative unbound structure 2DUBC loop 114-118 is missing, emphasizing the flexibility of this region. 2DUBC has a slightly stronger hot spot, even when mapped without domain split. In the bound structure 1EY3F the loop moves further from the site.
		1EY3F	DAK	1(17)	0.19	Substrate (11)
21 LO	1PZTA			0(27), 2(13)	0.46	Beta1,4-Galactosyltransferase-I undergoes critical conformational changes upon substrate binding from an open conformation to the closed conformation. This change involves two flexible loops: the small (residues 313 – 316) and the long loop (residues 345 – 365). In the unbound structure 1PZTA loop 345-365 is far from the ligand binding site, but the strong hot spot persists. Upon substrate binding, W314 in the small flexible loop moves towards the catalytic pocket and interacts with the donor and the acceptor substrates. The middle of loop 345-365 travels over 16 Å toward the ligand. These changes contribute to strong binding and are considered due to *induced fit* (12). However, we note that 1PZT is a monomer, whereas 1PZY is a heterodimer that also includes an alpha-lactalbumin molecule. This latter supports the binding of a N-acetyl-D-glucosamine molecule, and this may affect the conformation of the long loop.
		1PZYD	UDP	0(24)	0.21	Donor substrate uridine-5'-diphosphate.

22 SM	1XMGB			N/A		Soluble methane monooxygenase hydroxylase. The size of the deeply burried binding site is comparable to the size of the molecules used as probes by FTMap. Helix E (residues 202-211) undergoes a rearrangement in secondary structure. The absence of electron density for residues 205-215 of helix E reflects positional disorder (13).
		1XVCA	5BR	0(16)	0.40	After domain split. Very small ligand, small site (1-bromopentane).
23 LM	1IMFA			1(13), 2(12), 5(9)	0.30	Inositol monophosphatase *Loop 29 to 39 is missing in the unbound structure 1IMFA.* The loop acts as a lid and defines the pocket in the bound structure 1IMBB. Without the loop the pocket is completely open. Nevertheless, FTMap places the second and third strongest hot spots at the site. No other unbound structure with stronger hot spot has been found, so it appears that the origin of the site considered to be cryptic is the flexible loop, not visible in the X-ray structure.
		1IMBB	LIP	0(27)	0.22	Substrate l-myo-inositol-1-phosphate, strong binding.
24 S	2AX9A			4(8)	0.05	Androgen receptor. Interchange binding. With the single chainsn given, the pocket is open in both unbound and bound structures. In addition, in 2AX9A the side chains of M734 and K720 protrude into the site, there is no structure with stronger hot spot. FTFlex does not help.
		2PIQA	RB1	6(8)	0.10	Allosteric inhibitor, binds on surface, IC50 = 50,000 nM (14). The inhibitor must compete with the side chains that protrude into the site.

25 I	1EXMA 1EFTA 1TTTA			0(18), 1(17), 3(12) 1(19), 2(16) 0(20)	1.12 0.85 1.68	Elongation Factor TU. Domain-domain hinge motion. During its catalytic cycle, EF-Tu adopts two different conformations depending on the nucleotide bound: the GDP form and the GTP form. In the bound structure 1HA3B a GTP complex-like conformation of EF-Tu is observed, although GDP is bound to the nucleotide-binding site. Aurodox fixes EF-Tu on the ribosome by locking it in its GTP form. Inspection of the known structures of EF-Tu GDP and EF- Tu GppNHp shows that neither has the binding crevice for aurodox preformed. The domain 1–3 interface is very different in the GppNHp and in the GDP complexes, and in both structures, it has to be opened and rearranged to accommodate the antibiotic. Thus, aurodox binds by *ligand-induced fit* rather than by a lock-and-key mechanism (15, 16). Nevertheless, the mapping of the unbound structures 1EXMA, 1EFTA, and 1TTTA after domain split shows strong hot spots that overlap with the ligand binding site, and thus promote Aurodox binding.
		1HA3B	MAU	0(18)	0.32	Aurodox antibiotics. Hot spot is obtained after domain split.

26 LO	3KQAB 4E7DB4 E7DC			0(21), 1(17), 3(11) 0(23), 1(22) 0(24)	2.45 0.54 1.37	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (MurA). The reaction of MurA with substrate UDP-N-acetylglucosamine (UNAG) is said to proceed *through an induced-fit* mechanism with large conformational changes in the two-domain structure of the enzyme. The unliganded “open” enzyme state interacts with UNAG, forming a binary “closed” state to which the second substrate can bind (17). In the course of the open-closed transition, the loop 112-121 undergoes a drastic conformational change, closing on the UNAG-binding pocket. In the unbound structure 3KQAB the strongest hot spot is in the pocket but a bit too far. However, hot spots 3(11) and 4(11) overlap with the ligand. The main hot spot is much closer in alternative unbound structures such as 4E7DB.
		3LTHA	UD1	0(16), 1(16)	0.16	Substrate UDP-N-acetylglucosamine. Interactions with strong hot spots are responsible for the induced fit.
27 I	2BF3A 3RI7E			1(13), 4(10) 0(20)	0.40 0.22	Toluene 4-Monooxygenase Catalytic Effector Protein. The bound structure 3DHH is actually a hetero-complex of toluene 4-monoxygenase hydroxylase complexed with the effector protein.The ligand 4-bromophenol (BML) binds at the interface between the A and E chains of the complex. The alternative unbound structure 3RI7 is a similar hetero-complex, with a strong hot spot in the interface overlapping the ligand binding site. In contrast, the unbound structure 2BF3 has been crystallized as a monomer. Although it still has somewhat weaker hot spot at the cryptic site location, the unbound and bound structures have been determined under very different conditions, and the origin of conformational differences cannot be rigorously analized.
		3DHHE	BML	0(18), 1(17)	0.35	Small ligand 4-bromophenol.

28 CT	1CLLA 3EWTA 1IWQA 2HQWA 4M1LA			1(16) 0(31) 0(22) 0(19) 0(21)	0.95 0.13 0.16 0.19 0.22	Calmodulin (CaM), totally different unbound and bound structures. Residues 75 to 83 are missing in unbound, but the site still has the second strongest hot spot 1(16) close to the substrate binding site. The bound structure 1CTRA has a bound Ca^{2+} ion. Binding of Ca^{2+} induces marked conformational changes in CaM, including the exposure of hydrophobic clefts, which are important in the binding of CaM inhibitors and target enzymes (18). The alternative unbound structure 3EWTA has been cocrystallized with a peptide, which binds far from the cryptic site. Nevertheless, 3EWTA has a structure similar to that of the bound structure 1CTRA, and mapped with domain split yields a strong hot spot at the cryptic site.
		1CTRA	TFP	0(19)	0.25	Substrate (trifluoperazine), Kd = 1000 nM (2).
29 S	3H5RA 3H5RC 3H9QC			2(16) 1(16) 0(18)	1.25 0.59 2.02	The enzyme MccB catalyses a posttranslational modification of the MccA heptapeptide during the biosynthesis of microcin C7 (MccC7), a 'Trojan horse' antibiotic (19). In the unbound structure 3H5RA, D214 protrudes into the site. The pocket still has the hot spot 2(16), but it is not the strongest one. Chain C of the same unbound structure has a hot spot closer to the ligand binding site.
		3H9JD	APC	1(20)	0.35	Alpha,beta-methyleneadenosine-5'-triphosphate. Mapped after domain split. Only two phosphate groups of the alpha,beta-methyleneadenosine-5'-triphosphate ligand is visible in the bound X-ray structure. Appears to be an artifact, since most of the ligand is missing in chain D.

30 LC	1NI6D 1NI6A 1NI6B 1S8CB			1(17), 2(14) 0(25), 1(20) 0(26) 0(24)	1.07 0.87 4.84 4.62	Heme Oxygenase. Accommodation of the 5- trifluoromethylpyridin-2-yl group of the ligand requires a significant shift in a proximal helix, inducing the formation of a hydrophobic pocket (20). In the unbound structure 1NI6D residues 31-40 form a loop and a short helix that protrudes into the ligand binding pocket. The strongest hot spots, 0(27) and 1(17), overlap with the heme group, but 1(17) is also close to the inhibitor Q80. In contrast, in the bound structure 3HOKB residues 31-40 form part of a single large helix. There are four chains in 1NI6. Chain A has neither bound Q80 nor bound heme and the helix 31-40 also turns into the pocket, but its mapping places the strongest hot spot closer to the site occupied by the ligand. According to the authors of the X-ray structure 3HOK, the protein is an example of binding by "*induced fit*" in which the subpocket is formed by interaction with (in order to accommodate) the 5-trifluoromethylpyridin-2-yl substituent of the inhibitor (20). Induced fit certainly plays a role in changing the conformation of the 31-40 segment. While the structure of 1NI6A shows that strong hot spots can form spontaneously, the loop and helix of residues 31-40 would still clash with the ligand. *Thus, strong hot spots attract the ligand, but binding still requires induced fit, and the latter is responsible for the moderate binding affinity.*
		3HOKB 3HOKB	Q80	0(20) 2(14), 3(12), 5(8)	1.50 0.11	Inhibitor, IC50 = 2100 nM. Although the strongest hot spot 0(20) is only 1.5 Å from the ligand, it overlaps with the HEM group which binds in the same pocket. The strongest hot spots that actually overlap with the large ligand Q80 are 2(14), 3(12), and 5(8). The lack of overlap with stronger hot spots also contributes to the relatively weak binding.

31 LM	1QLWB 1QLWB			0(21) 5(7)	3.74 0.12	Alcaligenes esterase enzyme forms a dimer. *The loop formed by residues 31–42 of another chain is located close to the active site.* In the enzyme−ligand structure the Cα positions of residues 36–39 have moved 3–4 Å away from the active-site binding pocket. This movement of the residues implies that the loop interacts with the natural substrate and is involved in substrate recognition or binding. In the unbound structure 1QLWB there is only a single chain, and hence the site is not well formed. Probes bind at the entrance of a deep and narrow pocket, and only the small hot spot 5(7) reaches the bottom. In the bound structure 2WKWB a larger hot spot, 0(15), overlaps with the ligand binding site.
		2WKWB	W22	0(15), 7(4)	0.12	With reaction product analog W22.
32 LO	1HAGE 1EOJA 1H8IH 3D49H			0(19), 2(14), 5(7) 0(27) 0(24) 0(20)	0.31 0.27 0.13 0.14	Thrombin. In the unbound structure 1HAGE several loops are far from the pocket, making it too open. However, the main hot spot still overlaps with the ligand. The loops are closing down toward the site in the bound structure 1GHYH, resulting in a better defined pocket and a triad of very short ionic hydrogen bonds that stabilize the complex (21). We see the same changes in a number of unbound structures, indicating that the binding can be well described by a conformational selection model.
		1GHYH	121	0(22)	0.12	Inhibitor, Ki = 8 nM. A novel, multi-centered short hydrogen-bonding array between the enzyme and the inhibitor at the active site involving ionization of the inhibitor and of His57 supplies a significant, but not enormous affinity component (21).

33 LC	1OK8A 1TG8A 4UTAB			3(15) 0(17) 3(16)	4.60 3.74 0.33	Dengue 2 virus envelope protein. Crystals were grown in both the presence and the absence of the detergent n-octyl- -D-glucoside (BOG). The key difference between the two structures is a local rearrangement of the ‘‘kl’’ -hairpin (residues 268–280) and the concomitant opening up of a hydrophobic pocket, occupied by a molecule of BOG. Detergent binding marks the pocket as a potential site for small-molecule fusion inhibitors (22). The cryptic site is between two domains. After domain split we find 3(15), but not close to the site. The alternative unbound structure 1TG8A is similar to bound structure 1OKEB, and after domain split its main hot spot 0(17) is closer but not overlapping with the ligand binding site. The overlap is better in 4UTAB, but it is not the strongest hot spot.
		1OKEB	BOG	0(17)	0.15	The ligand, B-octylglucoside is a detergent, weak binder, but helps to open the site. *The finding is similar to that of beta-lactamase.*
34 LM	3DXNA			2(16)	4.79	Calcium-dependent protein kinase CDPK activation domain. In the unbound form, the C-terminal CDPK activation domain (CAD) resembles a calmodulin protein with an unexpected long helix in the N terminus that inhibits the kinase domain. Loop of residues 31 to 34 is missing in 3DXNA, mapping with domain split helps. No other unbound structure with intact loop has been found. Calcium or magnesium binding triggers the reorganization of the CAD into a highly intricate fold, leading to its relocation around the base of the kinase domain to a site remote from the substrate binding site. This large conformational change constitutes a distinct mechanism in calcium signal-transduction pathways (23).
		3HZTA	J60	0(17), 1(16), 3(11)	0.12	Allosteric inhibitor. Calcium or magnesium binding triggers a large conformational change.

35 LO	1K5HC 1ONNB 1K5HB 1K5HA 1ONNA			0(14), 3(11), 4(9) 0(21), 2(16) 0(19), 1(18) 0(16) 0(16)	4.80 4.91 1.77 2.42 1.91	Reductoisomerase. The loop of residues 208 to 215 is extremely flexible. In the unbound structure 1K5HC the loop moves outward, making the pocket very open. FTMap finds only the weak hot spots 0(14) and 3(11) in the neighborhood of the cryptic site, and 0(14) is 4.80 Å away from the location of the bound ligand. However, the loop moves toward the cryptic site in several unbound structures, resulting in somewhat stronger hot spots closer to the ligand binding site. In particular, in chain B of the unbound structure 1K5H part of the loop is unstructured and residues 212 through 215 are not visible in the X-ray structure. Nevertheless, the remaining part of the loop is sufficient to close on the cryptic site, resulting in two strong hot spots 0(19) and 1(18), the first being only 1.77 Å from the bound ligand.
		2EGHB	FOM	0(15), 3(11)	0.12	In the bound structure loop 208 to 215 closes down on the small ligand fosmidomycin (FOM) (24), a malaria drug/inhibitor, and the hot spot shifts toward the ligand. Mapping after domain split yields 0(15) and 3(11). Ki = 38 nM.
36 LO	1HKAA 1G4CA 2F63A 3KUGA			1(18), 3(9), 4(9) 0(28) 0(21) 0(17)	0.26 0.23 0.27 0.56	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase. The *enzyme is closing down on the product 6-carboxypterin.* In the bound structure 3IP0A loops 43-53 and 83-92 are closing down on the site, forming the ligand binding pocket (25). In the unbound structure the same loops move outward, resulting in a more open site. While it still has the second strongest hot spot 1(18), the strongest one, 0(19) is at the APC binding site. In the alternative unbound structure 1G4CA the loop 83-92 is still closing down on the pocket, and although loop 43-53 is very open, the cryptic site has the strongest hot spot 0(28).
		IP0A 3IP0A	HHS	0(23) 2(13)	2.20 0.11	Product 6-carboxypterin. In the ligand-bound structure R88 divides the pocket. It favorably contributes to ligand binding, but moves the strongest hot spot away from the ligand position.

37 LM	2BU8A 3CRKA 3CRLA 4MPCA			0(16) 0(28) 0(34) 0(19)	1.42 0.19 0.09 0.19	Human pyruvate dehydrogenase kinase. Upon binding by the inhibitor TF1, helix alpha-2 shifts by a hinge motion around the CR position of Met25. This induces a 2-2.5 Å shift of the Cα position of Phe31, and conformational change of its side chain from a lid position to open an induced pocket. The Phe31 side chain is located further in the binding site, alongside the bound inhibitor. *The loop of residues 34-37 is found to be very flexible in all structures determined to date (26).* This may be necessary to facilitate the hinge movement of the helix. *Residues 32 and 33 are actually missing in the bound structure 2BU2A.* In the alternative unbound structure 3CRKA the loop is not broken, but the site is wider. It has a strong hot spot at the site when mapped without domain split.
		2BU2A	TF1	0(23)	0.16	Inhibitor. Mapped without domain split.
38 LC	2OHGA		NHL	0(22), 1(21), 2(18)	0.30	Glutamate Racemase (GluR). GluR is composed of two domains of α/β protein that are related by pseudo-2-fold symmetry and the active site is located at the domain interface. The inhibitor, γ-2-naphthylmethyl-D-glutamate, located in the hydrophobic pocket. In the unbound structure 2OHGA the loop of residues 41 to 45 protrudes into the pocket, but the hot spot persists. The same loop is missing in the bound structure 2OHVA, resulting in a larger pocket (27).
		2OHVA	NHL	0(25), 1(18)	0.38	Inhibitor, IC50: 350 nM, Ki: 16 nM (2).

39 SC	2BLSB 4JXSA 3GR2A			0(19), 1(18), 2(17) 0(16) 0(16)	4.88 0.24 3.78	β-lactamase. In the unbound structure 2BLSB residues 289-293 form a small helix protruding into the cryptic site. Although the pocket has some strong hot spots, the subpocket that would bind the ligand is almost completely closed. The same fragment forms a loop in the bound structure 3GQZA, opening the subpocket, requiring induced fit. (28). 4JXSA has the hot spot closer to the inhibitor binding site because residues 285-290 are missing in the X-ray structure. A similar system will be discussed in more detail (see TEM β-lactamase, 89).
		3GQZA	GF7	0(24)	0.18	The small inhibitor fills up only a fraction of the pocket. Although this newly opened subpocket has the strongest hot spot, Ki = 6,700,000 nM (28) because of the need for induced fit.
40 S	1RHBA 3EUXB			0(17), 1(16), 5(6) 0(34)	1.56 1.79	Ribonuclease A. In 1RHBA H119 protrudes into the ligand binding site, which becomes shallower but is still open. 1(16) overlaps with NDP. In the alternative unbound structure 3EUXB the His119 side chain has two alternative conformers, one of them turns out of the pocket as in the bound structure 2W5KB. Thus, here the side chain appears to compete with the ligand (29).
		2W5KB	NDP	2(16)	0.28	Inhibitor, Ki = 12,000 nM (2).
41 LO	1ADEA 1ADIB			1(16), 2(13), 4(9) 0(20), 1(17)	0.31 0.41	Adenylosuccinate synthetase. Several loops around the site change conformation. In the unbound structure 1ADEA the loops fling out, resulting in a fairly open site. The loops tighten around the site upon ligand binding. In the alternative unbound structure 1ADIB the loops are closing more on the site, making it better defined (30). This is *fairly typical induced fit mechanism with flexible loops, also observed in other enzymes.*
		1CIBA	IMP	0(16)	0.29	Inosinic acid (IMP) co-factor.

42 I	1KS9A 1YJQA			0(31) 0(24)	0.34 0.38	Ketopantoate reductase. Two domains, with the active site located in a large and deep crevice between the domains. The crevice binds one pantoate and one $NADP^+$ molecule. In the unbound structure the main hot spot 0(31) is close to the pantoate binding site, but goes deeper into the crevice. The analysis of the conformational changes between unbound and bound structures reveals that C-terminal domain motions occur around three different rotation axes. In each case the rotation axis has a component perpendicular to the active site cleft, indicating that a degree of shear-like domain motion occurs (31). No special hinge bending domain closure is identified, but the binding crevice is still substantially narrowed. Domain split results in large hot spot deep in the crevice, not overlapping with the ligand.
		2OFPA	PAF	2(17), 4(9)	0.22	Reaction product pantoate. Very small ligand in a very deep and large crevice. The strongest hot spots 0(26) overlaps $NADP^+$, 1(22) is between the two ligands, and only 2(17) and 4(9) overlap with pantoate. Kd: 160,000 nM (2).

43 S	1H09A			5(7), 9(3)	0.32	Lysozyme (Cpl-1 endolysin). Interchain binding, with the given single chain structures the site is open. In addition, the side chain of Y127 is placed in the center of the substrate-binding cleft, making polar interactions with Y125 and Y153, via their hydroxyl groups, and Y127 completely blocks access of the substrate to the active site. Interaction of the active site with the peptidoglycan (PG) chain facilitates repositioning of the tyrosine side chain to a hydrophobic pocket created by F130 and P129 and places it in a suitable conformation for interacting with the substrate. It is not merely that the repositioning of Y127 makes room for binding of PG, but it also would lower the energy barrier for the hydrolytic reaction by its interaction with the substrate (32).
		2IXUA	MU2	0(29)	0.31	Muramyl dipeptide N-Acetylmuramyl-L-alanyl-D-isoglutamine, which was identified as the minimum structural entity essential for the immunoadjuvant activities characteristic of bacterial cell walls.
44 I	3BL9B			0(20), 2(14), 6(3)	0.09	Scavenger mRNA-decapping enzyme DcpS. Site is formed only when both chains A and B are considered. Mapping the dimer yields 0(20) overlapping the ligand. Mapping 3BL9B only places 1(26) at 3.9 Å from the ligand in the superimposed bound structure. It is likely that this was considered cryptic site due to disregarding the inter-chain binding
		3BL7A	DD1	2(19)	0.11	C5-substituted quinazoline D156844, Inhibitor, IC50: 7.62 nM (2). Again, we mapped the AB dimer. However, mapping chain A yields 0(26) overlapping with the ligand.

45 LC	1BP5A 1BP5D 1BTJB			2(18), 4(10) 0(17) 1(16)	2.43 2.51 2.64	Serotransferrin enzyme. Two domains, large scale rearrangement, the very small ligand is buried. Y95 protrudes into the site. Mapping with FTFlex helps, without it only the hot spot 1(14) is found in the site. In the alternative unbound structure 1BP5D Y95 is slightly pulled back, but it is sufficient to get the strongest hot spot, and the side chain is turned out of the way in the bound structure 1RYOA.
		1RYOA	OXL	0(25)	4.00	Oxalate ion.
46 S	2Q8FA			N/A		Pyruvate dehydrogenase kinase isoform 1. No probes cluster in unbound structure, because H149 protrudes into the site. It turns outward in 2Q8HA. Tiny ligand, deeply buried site. Neither domain split nor FTFlex help to find a hot spot near the cryptic site.
		2Q8HA	TF4	2(15)	0.11	Dichloro-acetic acid, Kd: 1000000 nM (33)
47 LO	2IYTA			0(22), 1(16)	0.31	Shikimate kinase. Very large motion of the 112-124 loop, but strong hot spots are still present in the site. Domain split increases CS0 to 22 from 17. FTFlex also helps, it yields the same increase. The loop closes down on co-substrates in the bound structure (34).
		2IYQA	ADP/ SKM	0(22)	0.21	Co-substrates ADP and shikimate. Mapped without domain split.
48 LO	1EX6A 4F4JB			0(18), 1(16), 2(12) 0(21), 2(16)	3.18 1.84	Guanylate kinase. Very large motion of the 32-53 loop. In the unbound structure 1EX6A the site is too open. Nevertheless, hot spots 0(18) and 1(16) are in the site. The hot spots are even stronger in the alternative unbound structure 4F4JB. The loop is closing down on the substrate binding site in the bound structure, resulting in 0(21), although somewhat far from the cryptic site.
		1GKYA	5GP	0(21)	1.65	Substrate guanosine-5'-monophosphate.

49 LM	1RRGA 1R8SA 4C0AD			0(25) 1(17) 0(16)	1.24 1.72 1.83	ADP-Ribosylation Factor 1 (ARF). *Loop 71-73 is missing in the unbound structure 1RRG.* However, the main hot spot is still in the site. The unbound structure is crystallized as a homodimer. The formation of the cryptic site is complicated by the fact that, in the bound structure, the complex also includes another protein, the guanine nucleotide exchange factor ARNO. The inhibitor brefeldin A (BFA) is buried in a cavity at the ARF–ARNO interface, with one-third of its surface in contact with ARNO and two-thirds with ARF. It displays close-packed hydrophobic interactions with both proteins. The cavity is closed by Tyr 81 in ARF and the 198–208 loop in ARNO (35)
		1S9DA	AFB	0(14), 1(12)	0.21	Uncompetitive Inhibitor brefeldin A stabilizes the ARF–ARNO complex. The hot spots 0(14) and 1(12) are in the ARF pocket, even when mapping without ARNO.
50 LO	1ECJD 1ECJB			0(24), 1(19) 2(17)	0.81 1.37	Glutamine phosphoribosylpyrophosphate amidotransferase enzyme. In the unbound structure 1ECJD, the loop 329-345 is far from the binding site. Nevertheless, mapping after domain split we find strong hot spots at the ligand binding location. The substrate analog PCP binds in a deep grove between two domains. The binding results in moving the loop toward the ligand binding site and the formation of a 20 Å channel connecting the active site for glutamine hydrolysis in one domain with the PCP site in a second domain. This solvent-inaccessible channel permits transfer of the NH3 intermediate between the two active sites. Tunneling of NH3 may be a common mechanism for glutamine amidotransferase-catalyzed nitrogen transfer and for coordination of catalysis at two distinct active sites in complex enzymes. *Thus, substrate induced conformational change is essential for enzyme function.* 1ECJB has almost the same structure as 1ECJD.
		1ECCB	PCP	0(19)	0.23	PCP substrate analog.

51 LO	2CM2A			0(20)	0.51	PTP1B. Loop 179 to 187 is far from the binding site in the unbound structure. The site is very open, but has a strong hot spot 0(20). In the ligand-bound structures 2H4K, 1PTY, and 2QBP this loop forms one side of the high affinity pTyr binding site in the structure 1PTY.
		2H4KA	509	0(17), 7(6)	0.18	Monocyclic thiophene inhibitor. Ki = 1300-3200 nM (2). The loop 179-187 moves, and becomes very similar to the loop in 1PTY, forming the high affinity pTyr binding site. Although the boundstructure binds a low affinity inhibitor in 2H4KA, essentially the same structure binds an inhibitor with Ki = 4 nM in the PDB structure 2QBP. In view of such high affinity binding it is suspected that the loop motion is spontaneous rather than induced by the ligand, but proving this hypothesis requires further analysis.
52 LC	1NUWA 5FBPB 1FBHA 4FBPC			10(3) 0(20) 1(17) 1(16)	0.25 0.27 0.09 0.15	Fructose 1,6-bisphosphatase enzyme. Large protein, many cavities. In the absence of AMP, the enzyme crystallizes in the R-state conformation, with loop 52-72 associated with the active site (36). In addition, the side chains E1020 and R1140 protrude into the site in 1NUWA, but not in the alternative unbound structure 5FBPB. In structures without AMP, three metal-binding sites are occupied by Zn^{2+} and two of three metal sites by Mg^{2+}. It is assumed that the association of AMP with FBPase disorders the loop, the consequence of which is a weaker hot spot, and the release of cations from two of three metal binding sites. In the resulting bound structure the enzyme is in the T-state conformation with a disordered loop of residues 52-72.
		1EYJB	AMP	3(12)	0.21	Adenosine monophosphate (36).

53A LM	3CJ0A 1NB7B 1GX6A			1(16) 0(17) 2(16)	0.12 0.32 0.20	Hepatitis C virus polymerase, complexed with an inhibitor. In the unboundenzyme 3CJ0A, *the binding site for POO is occupied by a small α-helix at the tip of the N-terminal loop that connects the fingers and thumb domains.* Nevertheless, the site includes a fairly strong hot spot 1(16) after domain split. Similar hot spot is found a number of other unbound structures. *This loop becomes partially disordered in the inhibitor-bound structures with no electron density present for residues 22–35.* Another consequence of the displacement of α–helix and the resulting weaker interaction between the thumb and fingers is a slight opening of the polymerase. POO inhibits the enzyme by preventing formation of intramolecular contacts between two domains and consequently precluding their coordinated movements during RNA synthesis (37).
		2BRLA	POO	2(17)	0.36	Inhibitor POO: IC50 = 18 nM (37). Mapped with domain split.
53B LC	3CJ0A 1NB7_B			1(15) 0(17)	0.23 0.32	Hepatitis C virus polymerase (same as 53A), but different inhibitor, binding at a different location. The inhibitor 79Z binds near the polymerase active site in a very enclosed, elongated, predominantly hydrophobic pocket between the primer grip motif (residues 364 –369) and the central sheet (strands 214–219, 319–325, and 310–316) in the core of the palm domain. In the unbound structure loop 364 –369 slightly moves toward the ligand binding site and pushes several side chains in. Nevertheless, mapping without domain split finds 1(15) there. In the bound structure the loop moves slightly out to accommodate the ligand and mapping yields a stronger hot spot. Residue N316 contacts the inhibitor (38).
		3FQKB	79Z	0(18), 5(8)	0.11	Inhibitor 79Z IC50 = 81 nM.

54 SC	1UK2A 2GT7A 2GZ9A 3EA8A			1(16), 3(12) 0(23), 2(16) 0(19) 0(17), 1(16)	2.81 0.68 0.71 0.07	SARS-CoV main protease. In the unbound structure 1UK2A, residues 46-50 form a short helix that reduces the pocket. In the bound structure 2GZ7 the same fragment is a loop, opening the site. In the alternative unbound structure 2GT7A residues 45-49 are missing, resulting in a hot spot. In 2GZ9A and 3EA8A the helix 46-50 is present, but located further from the site. This case is somewhat similar to beta-lactamase. However, the conformational change needed for ligand binding is much smaller, and several unbound structures (e.g., 2GT7A, 2GZ9A, and 3EA8A) already have a hot spot close to the cryptic site, resulting in higher affinity.
		2GZ7A	D3F	0(21)	0.37	Nonpeptide Inhibitor. Without domain split. IC50 = 300 nM (39).
55 LC	2WGBA 2V57D			0(25) 0(19)	1.01 1.91	TetR-like transcriptional regulator. The unboundprotein reveals a structurally asymmetric homodimer exhibiting local unfolding and a blocked drug-binding site, emphasizing the significant conformational plasticity of the protein necessary for DNA and multidrug recognition. Crystallographic and calorimetric studies confirm the intrinsic flexibility of the homodimer. In particular, loop 104-115 closes the pocket in the unbound structure, but the hot spot persists. The loop moves out of the way in the bound structure.
		2V57A	PRL	0(22), 1(19)	0.13	Proflavin drug, Kd = 79 nM.
56 LO	1HOOB 1ADEA 1ADIB			0(18) 0(16) 0(20), 1(17)	2.48 0.34 0.41	Adenylosuccinate synthetase with the inhibitor hadacidin. In the unbound structure 1HOOB 1(18) overlaps with GDP, 2(17) overlaps with IMP, both close to HDA. Upon ligand binding several loops (e.g., 42-53 and 294-306) are rearranged such that the pocket is much more open in unbound than in bound structure. In particular, loop 42–53 undergoes a 9-Å conformational change.
		1CIBA	HDA	0(16)	0.29	With domain split. Hadacidin is a small antibiotics.

57 F	1FA9A			N/A		Human liver glycogen phosphorylase. Large protein, major loop rearrangement, close to the dimer interface. However, the *unbound structure 1FA9A in the PDB has only one chain.* The bound structure 1L5SB also includes ligands NBG (1-N-acetyl-beta-D-glucosamine) at the hot spots 0(21) and 4(8); and PLP (vitamin B6 Phosphate) at 1(17) and 5(5). The URC (uric acid) site is a very narrow cavity on the monomer surface. It contains hot spot neither in unbound nor in bound structures.
		1L5SB	URC	N/A		URC: Kd = 550000 nM
58 LO	1W50A 1FKNA 1XN2C 4DVFA 4DVFB 4TRYC			1(17) 0(26) 0(16) 0(23) 0(18) 0(22)	2.85 0.21 0.39 0.30 0.10 0.04	Beta-secretase 1 (BACE-1) protease. Several loops, primarily 71-74, move such that in the unbound structure the pocket is too open. Domain split increases hot spot 1(13) to 1(17). The loops close down upon ligand binding, resulting in a better defined pocket and stronger hot spots.
		3IXJC	586	0(18), 1(18)	0.16	Selective beta-secretase 1 Inhibitor 586, IC50 = 0.32 nM (40). Bound structure mapped without domain split.
59 S	3B7DE			2(8), 3(6),4(6), 5(5)	3.2	Glutamate receptor 2. Ligand CX6, a potentiator, binds in the dimer interface, and hence we mapped the dimer (41). The potentiator binding site is adjacent to the “hinge” in the ligand-binding core “clamshell” that undergoes conformational rearrangement after glutamate binding. Diffuse hot spots covering the site in both unbound and bound structures. The side chain of Ser 108 getting into the site in the unbound structure seems to be the main difference.
		2AL4F	CX6	0(13), 3(8), 4(8), 7(6)	3.0	Potentiator CX6. Mapping the F-D dimer. Diffuse hot spots.

60A 60B I	1PKLB 3HQOC 3HQQE 3HQQV			4(7), 5(7) 0(26) 2(20) 1(20)	4.1 0.16 2.75 2.75	Pryuvate kinase enzyme. Functions as tetramer, each subunit with substantial hinge motion between domains. The transition between inactive T-state and active R-state is accompanied by a simple symmetrical rigid body rocking motion of the A- and C-domain cores in each of the four subunits. In R-state eight essential salt bridge locks that form across the C-C interface provide tetramer rigidity with a coupled 7-fold increase in rate. The conformational changes coupled with effector binding correlate with loss of flexibility and increase in thermal stability (42). Thus, flexibility is crucial to T- to R transition and enzyme function. Ligands bind in domain interface that changes conformation. Although the hot spots in the unbound structure 1PKLB are weak, pocket is more visible when mapping the alternative unbound structure 3HQOC with domain split. Strong hot spots are also fund in various chains of the structure 3HQ*O.*
		3HQPP	ATP/ FDP/ OXL	0(20)	0.24	Mapped after domain split.
61 LC	1K3FB 1ZL2E 2OECC 2PGAE 3FWPE 4E1VI			0(27) 0(26) 0(17) 0(22) 0(20) 0(23)	0.42 0.58 0.16 0.19 0.18 0.41	Uridine phosphorylase (UP). It is suggested (43) that the ‘induced-fit’ movement in UP involves an active-site loop containing residues 225–230 that acts as a lid over the pyrimidine-binding site upon ligand binding. The loop is too open in unbound structure and moves inward in bound. Mapping of different unbound structures shows that the hot spot persists without any ligand. However, the loop 225–230 is somewhat far from the ligand in all unbound structures, thus it is possible that the binding site is formed by some combination of conformational selection and induced fit.
		1U1DF	181	0(23), 1(18)	0.19	Inhibitor. Ki = 353 nM (human, not E. coli) (43).

62 LC	1FVRA 1FVRA			2(17), 3(15) 0(21)	0.75 2.22	Angiopoietin-1 receptor. In the unbound structure 1FVRA loop 981-995 protrudes into the binding site and would clash with the bound ligand. Residues 996-999 are missing. Hot spots 2(17) and 3(15) are at the site, but the two strongest hot spots are at another missing loop.
		2OO8X	RAJ	0(16), 2(11), 3(10), 4(9)	0.22	In 2OO8X residues 987-996 are missing. Inhibitor, IC50 = 1 nM (2). Hot spots cover most of the large ligand.
63 S	3HQDA 4A28B			4(9), 5(7) 0(26)	0.04 0.31	Kinesin-like mitotic motor protein KIF11. In the unbound structure 3HQDA E116 and W127 protrude into the cryptic site; the main hot spot with 0(18) and 7(4) is at the ADP binding site. *In the alternative unbound structure 4A28B E116 has a slightly different rotamer, resulting in a much stronger hot spot, demonstrating the large impact of the single side chain on the properties of the binding site.* The KIF11 protein binds the small molecule monastrol and Mg^{2+}·ADP. The bound structure reveals that monastrol confers inhibition by “induced-fitting” onto the protein some 12 Å away from the catalytic center of the enzyme, resulting in the creation of a previously non-existing binding pocket (44).
		1Q0BB	NAT	0(15), 3(14)	0.27	NAT (Monastrol) IC50=7000 nM (2). Hot spot 1(15) is at the at ADP site.

64 SC	2ZB1A 1YWRA			N/A 0(23), 2(16)	0.19	P38 MAP kinase. Helix 253-261 closes down the site in the unbound structure 2ZB1, whereas loop 195-200 is too far from the ligand. No hot spot there even after domain split. 1YWRA has an active site inhibitor at a different location, but no ligand around the cryptic site. Nevertheless, the binding of the inhibitor at the active site affects the cryptic site region. The helix 253-261 now occupies a position that is between those in 2ZB1A and 2NPQA, whereas loop 195-200 moves closer to the ligand than in the two other structures, resulting in a strong hot spot.
		2NPQA	BOG	0(21), 6(6)	0.12	The entire helix 253-261 moves outward to accommodate the ligand BOG (octyl glucoside), which is a weakly binding solubilization agent (Kd = 3080 nM) (2). The site is a known lipid binding site (45). *Based on the 1YWR structure, the site can also be formed by the binding of a ligand some 30 Å from the cryptic site.*
65 I	3PEOG 3PEOG			0(22) 1(18)	0.15 0.25	Nicotinic acetylcholine receptor. Inter-chain site. Mapping chain G of 3PEO yield two strong hot spots 0(22) and 1(18) that are located at the agonist binding sites on the two sides of the protein. Upon ligand binding at the subunit interface, the extracellular domain of the receptor undergoes conformational changes, and agonist binding allosterically triggers opening of the ion channel. In particular, loop C wraps around the agonist lobeline, resulting in strong binding (46). The hot spot is conserved in the process.
		2BYSJ 2BYSJ	LOB LOB	0(24) 1(18), 4(12)	0.18 0.21	Agonist lobeline, Kd = 0.3 nM (2). As in the case of the unbound structure, mapping yields hot spots at LOB binding sites on both sides of the protein.

66 LM	2BRKA			0(23)	0.17	RNA-directed RNA polymerase (47). In the unbound structure 2BRKA loop 495-497 protrudes into the site, L497 clashes with ligand, and the *C-end segment 532-558 is missing.* Domain split helps to remove irrelevant regions, since otherwise many probes bind in an inter-domain crevice. Strong hot spot 0(23) is still found after the domain split, so even the unformed pocket has strong binding capacity for the small probes.
		2GIRB	NN3	2(15), 3(12)	0.12	In the bound structures the clashing loop moves out, and after domain split the surface pocket is found. The hot spots cover the two arms of the inhibitor NN3. IC50: 140 nM (2).
67 S	2H4EB			1(13)	0.40	Transthyretin (TTR). In the unbound structure 2H4EB, K15 and L17 protrude into the site residues 1-9 are missing, and 0(15) is at residue 10 rather than at the cryptic site.
		3CFNB	2AN	0(20)	0.24	Surface pocket, barely formed. The TTR-specific ligand binds due to stacking of the naphthalene ring between the side-chain of K15 and L17 (48). Kd: 4300 nM (2).
68 LO	2CGAB 2CGAB 1OXGA 1ACBE 1CGJE 1GL0E 1GL1A			0(25), 1(21) 0(26) 0(18) 0(26) 0(17) 0(22)	2.17 0.14 0.11 0.30 0.09 0.25 0.23	Gamma-chymotrypsin. Loop 191-195 moves outward in the unbound structures, making the pocket too open. In the bound structure the loop closes down on the inhibitor. In several alternative unbound structures the conformation of the loop is similar to its conformation in the bound structure 1OXGA, i.e., the loop remains fairly open.
		1AFQC	0FG	1(24)	0.20	Dipeptide inhibitor, Ki = 610 nM (49).
69 LO	1XCGB 1LB1D 2RGNC 2RGNF			0(22) 0(20) 0(20) 0(21)	0.17 0.31 0.26 0.24	Transforming protein RhoA. In 1XCGB loop 25-42 moves outward, makes pocket too open, but hot spot persists in several unbound structures. The loop is closing down on GDP in the bound structure.
		1OW3B	GDP	0(23)	0.23	Cofactor, strong binding.

70 S	1FXXA 3HL8A	 BBP	N/A 5(8), 7(5)	 	Exodeoxyribonuclease I (Exol). In the unbound structure 1FXXA, W245 protrudes into the surface site, which is very weak. FTFlex does not help. With domain split. Inhibitor, K_i = 32,000 nM (2).
71 SO	4AKEB 4AKEB 4AKEA 4AKEA 4X8MA 1 ANKB	 ANP	0(22) 4(12) 0(17) 2(16) 0(20) 0(19)	4.18 0.23 0.68 1.31 0.67 0.15	Adenylate kinase. Large conformational change. In 4AKEB, segment 115-163 turns away, and hence the pocket is not formed. The largest hot spot, 0(22), overlaps with AMP, 4.18 Å from the ANP site, and only 4(12) overlaps with the cryptic site. However, the strongest hot spot is closer to the ANP binding site in alternative unbound structures such as 4AKEA and 4X8MA, although they are similar to the 4AKEB structure. With bound nucleotides the 115-163 lid becomes closed, and the active site is buried (50). However, after domain split mapping places there the strongest hot spot, 0(19). ATP analog inhibitor. With domain split.
72 LC	3NNUA 3HL7A	 I46	N/A 6(6)	 0.23	I46 binds at an allosteric site of the p38 MAP kinase. In 3NNU loops 251-253 and 197-198 approach the binding site of I46, with several side chains protruding into the pocket (51). After domain split. Even with domain split, most probes are at the ATP binding site, forming 0(22), the site that binds the inhibitor I47. The next strongest hot spot, 1(21), is at the EGFR dimerization interface. I46 is a weak binder.

73 I	1A8IA			N/A		Glycogen phosphorylase is a dimer and binds two 1-substituted-3,4-dihydro-2-quinolone inhibitors at the dimer interface. The unbound structure 1A8IA is only a monomer in the PDB. However, mapping the dimer does not really help because of the two symmetric binding sites in the dimer interface (see below).
		2IEGB	FRY	3(8)	0.32	For the bound structure 2IEGB we have mapped the dimer that has many cavities. However, due to the two bound inhibitors the hot spots are not well defined, in spite of the fairly good inhibitor binding (IC50 = 135 nM (2)). In addition, the stronger hot spots 0(15) and 1(12) are at the PLR binding sites.
74 I	1SWXA 3RWVA			0(19) 0(26)	0.18 0.35	Glycolipid transfer protein (52). Inter-chain binding, in the given single chain pocket is very open. The main hot spot 0(19) overlaps with the lactosylceramide (LAT) binding site in the unbound structure 1SWX. The hot spot is also present in the alternative unbound structure 3RWVA.
		2EUMA	LAT	7(6)	0.32	Parts of two large helices (42 to 62 and 142 to 160) move away from the site in 2EUMA, opening more the pocket that becomes fairly flat. In addition, the bound structure 2EUMA binds 4 more ligands close to LAT: lipids SPH and OCA, decane D10, and octanoic acid OCT. These sites attract the probes, resulting in a weaker hot spot at the LAT binding site.
75 I	2QLRC			N/A		Kynurenine/alpha-aminoadipate aminotransferase. 0(21), 1(19), and 6(6) bind at LLP (L-peptide) binding site; AKG *binds inter-chain between A and B;* 2QLRC and 3DC1A have different quaternary structures; 2QLR is a homotetramer, whereas 3DC1 is a homodimer; the interfaces are not the same.
		3DC1A	AKG	7(5)	2.50	2-oxoglutaric acid. Binding causes large conformational change of the N-terminal fraction, residues 15-33, that is able to adapt to different substrate sizes (53).

76 SC	2F6VA			9(3)	0.30	Protein tyrosine phosphatase 1B (PTP1B). Extra helix in unbound structure (54). *Truncated in the bound and a number of lower resolution unbound structures* (see text).
		1T49A	892	N/A		Inhibitor, IC50 = 22000 nM. The inhibitor binds in a groove formed by helices α3 and α6 that positions the catalytically important WPD loop. Distance from the active site Cys is ~20 Å. A very narrow hydrophobic pocket is formed by Leu192, Phe196 and Phe280. The strength of binding is defined by the interactions with the single side chain of F280 (54), and FTMap does not place any probes in the pocket.
77 S	1G24D 1G24B 1UZIA 1UZIB 2C8FG			0(16), 1(16) 0(17) 0(22) 0(18) 0(21)	0.23 0.38 0.30 0.14 0.19	Rho Adp-Ribosylating Clostridium Botulinum C3 exoenzyme. In the unbound structure F183 protrudes into the side, in the bound it moves in to form a better-defined pocket.
		1GZFC	NIR	1(16), 4(9)	0.36	Structure also contains a NAD and an ADP molecule (55). 0(17) is at the ADP site.
78 I	1SU4A 2O9JA 1XP5A 2ZBGA 3N5KA			1(15), 2(14) 0(23) 0(18) 0(22) 0(19)	4.90 0.15 0.21 0.25 0.26	Sarcoplasmic/endoplasmic reticulum calcium ATPase (56). Large hinge motion, In the unbound structure the interdomain region is very open. Unbound and bound structures do not even align well. Cryptic site is between two domains. Domain split helps to find hot spots 1(15) and 2(14), but at 4.90 Å from the cryptic site. The alternative unbound structures 2O9JA, 1XP5A, 2ZBGA, and 3N5KA are more similar to the bound structure than 1SU4A. These structures were also mapped after domain split.
		3FGOB	ACP	1(18)	0.26	Adenylate ester binds adenosine-5'-[beta, gamma-methylene] triphosphate (ACP), an ATP analog. Mapped with domain split.

79 U	2AIRH 4E2FD 3MPUD			6(2) 0(27) 2(16)	0.30 0.33 0.27	Aspartate transcarbamoylase. Binding of CTP is at the flexible N-terminal end, which is disordered. Pocket does not exist in the unbound structure 2AIRH. 4E2FD is mapped with domain split. In both 4E2FD and 3MPUD residues 1-9 are missing, which makes the amino ends more ordered resulting in a pocket with a hot spot.
		1ZA1D	CTP	2(16)	0.33	CTP: Cytidine-5'-triphosphate. The conformational change *is induced upon the binding of carbamoyl phosphate*. Large surface pocket is formed. Hot spot 2(16) is placed there after domain split. The induced fit dramatically alters the electrostatics of the active site, creating a binding pocket for aspartate (57).
80 S	1E2XA 1H9TB			6(4) 1(16)	3.6 3.02	Fatty acid metabolism regulator protein (FadR) (58). The pocket is essentially closed in the unbound structure, with the side chains M168, and Y172 protruding into the MYR binding site. There is only a small hot spot. The stronger hot spot 2(16) is at the coenzyme A (COA) rather than the myristic acid (MYR) site at the entrance of the pocket. In the bound structure 1H9GA two helices slightly turn, moving the side chains out of the site. The alternative unbound structure 1H9TB shows similar movement.
		1H9GA	MYR	0(21)	2.45	Myristic acid (MYR).

81 I	1MY0B 1FTOB 1FTOA 4U2RA4 U2RC			3(11) 0(30) 0(26) 0(35) 0(34)	0.20 0.17 0.20 0.17 0.23	Glutamate receptor 2. Interdomain binding. In the unbound structure hinge motion of the domains pushes loops 59-67 and 136-142 into the pocket. Ligand does not fit, but small probes do, and only 3(11) overlaps with the cryptic site. Hot spot 0(18) is at the entrance of the pocket between the two domains, at 5.4Å from the position of the bound ligand. In the bound 1N0TD structure the domains are moved apart by hinge motion, expanding the pocket. The two domains also move away from each other in the alternative unbound structures 1FTOB, 1FTOA, 4U2RA, and 4U2RC, creating strong hot spots.
		1N0TD	AT1	0(24), 1(16)	0.21	Antagonist (S)-ATPO *stabilizes the open form of the receptor* (59). IC50 = 12200 nM, weak binder.
82 LM	1ZAHB 3DFND 2OT0B 2OT0D 3DFPD			0(16) 0(25) 0(16) 0(28) 0(24)	0.96 2.19 2.56 2.16 2.14	Fructose-1,6-bisphosphate aldolase enzyme. Small changes make the cavity in bound structure tighter. In particular, residues 301-306 form a short helix supporting the bound ligand, but the same segment is a loop in the unbound structure 1ZAHB. *The same segment is in a very different conformation in the two other structures, with a large part not visible in 1ZAHB.* The flexible C-terminal loop 341-361 in the alternative unbound structure 3DFND further supports the pocket, resulting in stronger hot spots.
		2OT1D	N3P	1(17)	0.21	Weak Inhibitor, Ki = 100000 nM (2). 0(23) at 3 Å from N3P.

83 S	4HB2C 2OT0D 3DFNC 3DFPD 4E2FD			9(5) 0(28) 3(19) 3(24) 0(27)	0.33 2.30 3.10 1.49 0.33	Exportin-1 subunit of chromosomal region maintenance 1 protein (CRM1), which also includes the GTP-binding nuclear protein Ran and Ran-specific GTPase-activating protein 1 (60). Very large protein, multiple domains. In the unbound structure 4HB2C, helices 521-541 and 570-585 that define the binding site for the polyketide natural product leptomycin B (LMB) are too close to each other, and move slightly apart in the bound structure 4HATC. Conformational differences between empty and inhibitor-bound grooves result from both slight helix reorientation and *rearrangements of a few side chains, including R543, K545, K548, F572, E582, and F583.*
		4HATC	LMB	2(17), 3(9)	0.11	The ligand is the polyketide natural product leptomycin B. Mapped with domain split.
84 LC	1BSQA 5IO5A 1BSYA 3NPOA			4(9) 0(24) 0(17) 0(16)	4.90 0.34 4.11 4.78	Bovine Beta-lactoglobulin (61). Loop 84 to 90 in the unbound structure 1BSQA protrudes into the RTL binding site, and it has only a weak hot spot at the entrance of the pocket. This is also the case in some other unbound structures such as 1BSYA and 3NPOA. *However, in the unbound structure 5IO5A the loop opens the pocket.* This results in a strong hot spot at the cryptic site, and suggests that loop motion can be spontaneous rather than induced.
		1GX8A 1GX8A	RTL RTL	2(12), 3(9), 5(8) 1(14)	0.29 1.80	Retinol, Kd = 430 nM (2). Binding in the central cavity of the beta-barrel. hot spots 2(12), 3(9) and 5(8) are deep in the pocket and cover the entire ligand fairly well. The hot spot 1(14) is at the entrance of the pocket.

85 SM	2GPOA			4(10)	2.20	Estrogen-related receptor gamma (ERRγ). In the unbound structure 2GPOA the C-terminal activation helix H12 protrudes into the CHD site. This is frequently seen in nuclear receptors. Nevertheless, 4(10) is only 2.2 Å away from the position of the bound CHD (cholic acid). The primary ligand, hydroxytamoxifen (OHT), binds at the hot spot 0(17), correctly found by FTMap. When hydroxytamoxifen is bound in 1S9QB, the activation helix is disordered and CHD can bind in the cryptic site. The presence of CHD in the crystal was unexpected. One possible explanation is that CHD has been co-purified from the E. coli expression host, since bacteria present in the human colon are known to be involved in the metabolism of bile acids that escape the enterohepatic circulation. Cholic acid together with parts of H12 from a neighboring molecule covers the hydrophobic coactivator groove and, thus, stabilizes the crystal packing. However, there is no physiological relevance for the observed association of cholic acid with the ERR (62).
		1S9QB	CHD	4(8)	1.9	*Helix H12 is missing in bound structure.* 0(15), 1(15), 3(11), and 5(7) overlap with the OHT site.
86 SO	3FDLA 3R85A 4Z9VA 5C3GA			0(23), 2(15) 0(20) 0(16) 0(16)	0.23 0.07 0.52 0.09	Bcl-xL(63). In 3FDLA the helix 101-111 seems to be further from the ligand binding site than in the bound structure 2YXJA. Nevertheless, mapping 3FDLA yields 0(18) without and 0(23) with domain split. Thus, the site has strong hot spots in the unbound structure. The alternative unbound structure 3R85A is similar to 2YXJA and has 0(20) without domain split.
		2YXJA	N3C	0(29), 1(15), 2(10) 3(10)	0.23	Inhibitor ABT-737. Ki = 0.5 nM. Very strong hot spots covering the ligand.

87 LC	1B6BA 1B6BA 1B6BB 1B6BB			0(24) 1(16), 2(14) 0(17) 1(16)	0.41 0.13 0.32 0.15	Serotonin N-acetyltransferase (64). In the unbound structure 1B6BA loop 52-63 protrudes into the binding site. Nevertheless, the strong hot spots persist. The loop is moved out in the ligand-bound structure.
		1KUVA	CA5	0(19)	0.11	Potent bisubstrate analog inhibitor. Ki = 22 nM.
88 LO	1KZ7D 1KZGB 5FI1B			1(16) 1(16) 1(16)	0.85 0.67 0.71	Rho-related small GTP-binding protein Cdc42. Loops 31-38 and 58-63 in the unbound structure 1KZ7D are too open. Nevertheless, we have a strong hot spot, but the pocket becomes better defined when the tiny ligand binds in 1GRNA. It was shown that aluminum fluoride to act as a transition state analog in the hydrolytic pathway of low molecular weight GTP-binding proteins (65).
		1GRNA	AF3	0(20)	3.5	Aluminum fluoride, very close to the phosphate of bound GDP. 0(20) is overlapping with the phosphate site, but its center is only 3.5 Å from the Al.

89 SC	1JWPA 1ZG6A 4IBXE 3GMWA			0(17) 0(24) 0(32) 0(17)	1.09 0.84 0.60 0.60	TEM β-lactamase, a textbook example of cryptic sites (66). The hot spot 0(17) in the unbound structure is at the substrate binding active site, but one of the inhibitors bound to 1PZO is only 1.09 Å from this site. Thus, while the inhibitors do not overlap with the hot spot, β-lactamase still satisfies the condition that there is a strong hot spot near the cryptic site. It is well known that the cryptic site opens because helices 11 and 12 move apart. Specifically, the α-carbon atoms of residues 218 – 224 (helix 11) move 3 – 7 Å and those of residue 274 – 285 (helix 12) move 1 – 3 Å from unbound structure positions. The flexibility of the region can be seen in some of the unbound structures. In particular, in chain E of 4IBX helix 11 is completely unfolded, although still is closer to helix 12 than in 1PZO. However, 4IBX is an engineered protein, specifically designed to have a different stabilized conformation, and 1ZG6 and 3GMW also have mutations that change their conformation.
		1PZOA	CBT	0(16)	0.13	Inhibitor, Ki = 480000 nM.
90 LC	3GXDB 2NT1D2 NT1D			7(5) 1(19) 3(16)	0.21 2.84 1.95	Acid-beta-glucosidase (67). In the unbound structure 3GXDB, loop 342-356 moves outward from the site, which makes it too open. The loop is somewhat closer in the alternative unbound structure 2NT1D, and it is substantially closer in the inhibitor-bound structure 2WCGA. All structures were mapped with domain split.
		2WCGA	MT5	1(17)	0.18	Inhibitor. Ki = 420 nM, C50 = 4000 nM (2).

91 LM	1BNCB 2V5AB 2W6PB 2W6ZB 3G8DA 3JZFA			1(16), 4(11) 0(19) 0(20) 0(18) 1(19) 0(23)	4.36 1.00 0.88 0.67 1.26 0.72	Biotin carboxylase (68). Loop 159-169 is missing in the unbound structure 1BNCB, which makes the site completely open. Chain A of 2V5A has the bound inhibitor LZL, but chain B of 2V5A does not. Nevertheless, its conformation is similar to that of chain A, the loop is present, resulting in a strong hot spot. The site is even better defined in 2V5AA, but the change in loop conformation appears to be at least partially *spontaneous.*
		2V5AA	LZL	0(23), 2(11), 3(11)	0.35	Hot spots cover most parts of the inhibitor. IC50 = 150 nM (2).
92 LC	1Z92A 1M47A 1M4CA 1NBPA 3QB1D 3QB1G			0(23) 0(20) 0(21) 0(21) 0(17) 0(17)	0.19 0.75 0.24 0.30 2.07 1.94	Interleukin-2. Loop 30-35 of 1Z92 protrudes into the site. Although 1Z92 does not have a bound inhibitor, it has been co-crystallized with IL2R-α, so it is not entirely an unbound structure. Although 1Z92A has a large hot spot that overlaps with the long ligand, it has no hot spot in the flexible nonpolar pocket close to the (left) cryptic region. In contrast, the unbound structures 1M47A, 1M4CA, and 1NBPA all have hot spots in the nonpolar pocket.
		1PY2A	FRH	1(17), 2(14)	0.30	IC50 = 60 nM (69).
93 LM	1JBUH 1JBUH 1FAKH 1KLJH 2PUQH			1(16) 0(18) 0(30) 0(16) 0(28)	0.47 3.50 0.05 0.67 0.08	Coagulation Factor VII Zymogen (EGF2/Protease). In the unbound structure several loops are missing around the cryptic site, e.g., 184-189 and 220-222. These loops are well formed in the bound structure 1WUNH and result in a better defined pocket.
		1WUNH	P5B	0(24)	0.04	Inhibitor IC50 = 62 nM (70)

LC Loops protruding into the site in the unbound structure, and making it closed to ligand: 21
LO Loops too open in the unbound structure, making the pocket not well formed: 18
S Side chains protruding into the site in the unbound structure: 18
LM Loops missing in the unbound structure, leading to loss of the pocket: 11
I Inter-domain cryptic site, affected by hinge or other motion of the two domains: 11
U Unstructured regions in the unbound structure, in most cases N or C termini: 3
SC Secondary structure elements too close in the unbound structure, closing on the pocket: 5
SO Secondary structure elements too far in the unbound structure, making the pocket too open: 2
SM Secondary structure elements missing in the unbound or bound structure: 2
CT Very large conformational transition from unbound to bound structures (calmodulin): 1
F FTMap fails: pocket is too weak to bind probes: 1

Table S1. Mapping results for the unbound structures in the CryptoSite set (line 1 for each protein), for selected unbound structures in the extended set, and for the bound structure given in the CryptoSite set. The PDB ID includes the chain, e.g., 3CHEA for protein 1 means chain A of the protein with the PDB ID 3CHE. Lig denotes the ligand code as specified in the PDB. The column “Hot spots (CSs)” lists the rank of the consensus sites, starting from CS0 for the strongest consensus site with the highest number of probe clusters. For each consensus site the number of probe clusters is given in parenthesis. The column “Dist.” shows the shortest distance, in Å, between any ligand atom and any probe atom in the consensus cluster.

Table S2. RMSD values and distances from local alignment based on the residues that are in the 9 Å neighborhood of the ligand in the bound structure. Residues missing in any structure are disregarded.

No	Unbound_Chain_Bound_Chain_Ligand	A	B	C	D	E	F	G
1	3CHE_A_2IUZ_B_D1H (S)	0.24	2.00	1.52	2.64	1.39	2.60	0.52
2	2AKA_A_1YV3_A_BIT (S)	0.86	1.63	1.31	2.48	1.00	1.18	0.17
3	2GFC_A_2JDS_A_L20	1.01	1.02	1.02	2.61	0.66	2.73	0.52
4	1ALB_A_1LIC_A_HDS (S)	0.50	1.00	0.79	2.92	0.21	2.58	0.21
5	1NEP_A_2HKA_C_C3S	1.59	2.60	2.15	3.47	0.38	0.59	0.38
6	3MN9_A_3EKS_A_CY9	1.57	3.08	2.48	2.87	0.77	3.06	0.17
7	1ALV_A_1NX3_A_ISA (S)	0.62	1.65	1.28	3.02	0.45	2.07	0.45
8	2YQC_A_2YQS_A_UD1	1.77	2.16	1.96	2.64	1.42	1.42	1.42
9	2QFO_B_2WI7_A_2KL	1.95	2.51	2.22	2.83	2.06	2.23	1.71
10	1RTC_A_1BR6_A_PT1 (S)	0.36	1.28	0.96	2.57	0.27	2.82	0.23
11	1RDW_X_1J6Z_A_RHO	1.01	3.16	2.42	2.54	0.15	3.27	0.09
12	1TQO_A_1TR5_A_THP	1.08	3.30	2.56	2.84	2.33	2.33	2.33
13	3PUW_E_1FQC_A_GLO	2.71	3.19	2.99	2.44	2.55	2.79	0.34
14	3L7U_C_2HVD_C_ADP	0.69	1.49	1.21	2.48	2.29	2.29	2.29
15	1R1W_A_3F82_A_353	2.04	3.83	3.03	2.81	0.59	0.62	0.37
16	1G4E_B_1G67_B_POP_TZP	0.43	2.01	1.38	2.67	1.91	2.55	1.46
17	3F74_C_3BQM_C_BQM	2.00	2.51	2.30	2.75	0.39	1.94	0.15
18	1MY1_C_1FTL_A_DNQ	1.42	1.53	1.48	2.65	2.14	2.60	0.58
19	2WGQ_B_1D6Y_B_HY1	0.27	0.91	0.69	2.84	1.04	1.40	1.01
20	1DUB_D_1EY3_F_DAK (S)	0.39	1.59	1.09	2.91	0.97	2.49	0.97
21	1PZT_A_1PZY_D_UDP	3.37	3.45	3.40	2.78	1.09	2.10	0.28
22	1XMG_B_1XVC_A_5BR	0.29	0.95	0.72	3.54	0.51	0.98	0.21
23	1IMF_A_1IMB_B_LIP	5.27	4.68	5.25	2.60	2.17	2.17	2.17
24	2AX9_A_2PIQ_A_RB1 (S)	0.38	1.38	1.00	2.95	0.49	3.04	0.38
25	1EXM_A_1HA3_B_MAU	2.25	3.62	2.99	2.57	0.36	0.66	0.06
26	3KQA_B_3LTH_A_UD1	4.26	6.11	5.28	2.59	0.88	2.19	0.51
27	2BF3_A_3DHH_E_BML	0.87	2.35	1.83	3.29	0.58	3.28	0.58
28	1CLL_A_1CTR_A_TFP	9.10	24.68	11.77	3.07	0.57	2.89	0.09
29	3H5R_A_3H9J_D_APC(S)	0.65	1.25	0.96	2.55	1.55	2.63	1.04
30	1NI6_D_3HOK_B_Q80	1.73	3.24	2.60	2.83	1.35	1.35	0.55
31	1QLW_B_2WKW_B_W22	0.28	0.96	0.70	2.75	2.42	2.42	2.42
32	1HAG_E_1GHY_H_121	2.76	4.65	3.75	1.88	0.77	2.70	0.43
33	1OK8_A_1OKE_B_BOG	3.10	4.00	3.60	2.68	0.25	3.71	0.14
34	3DXN_A_3HZT_A_J60	1.60	2.81	2.25	2.76	0.61	0.61	0.61
35	1K5H_C_2EGH_B_FOM	3.52	6.46	5.11	2.49	1.23	1.69	1.23
36	1HKA_A_3IP0_A_HHS	3.05	5.67	4.60	1.62	1.27	1.91	0.85
37	2BU8_A_2BU2_A_TF1	0.85	1.27	1.09	2.63	0.71	2.79	0.29
38	2OHG_A_2OHV_A_NHL	1.36	1.46	1.40	2.46	0.70	0.70	0.70
39	2BLS_B_3GQZ_A_GF7	2.62	4.69	3.75	2.65	0.53	2.34	0.10
40	1RHB_A_2W5K_B_NDP (S)	0.43	1.68	1.24	2.65	1.84	4.08	0.29
41	1ADE_A_1CIB_A_IMP	2.25	2.74	2.49	2.64	1.12	3.39	1.12
42	1KS9_A_2OFP_A_PAF	1.12	1.64	1.39	2.70	2.18	2.91	2.18
43	1H09_A_2IXU_A_MU2 (S)	0.21	2.00	1.43	2.85	2.04	2.04	2.00
44	3BL9_B_3BL7_A_DD1	6.86	7.03	6.96	2.68	0.57	0.57	0.53
45	1BP5_A_1RYO_A_OXL	3.72	4.27	4.00	2.60	2.40	2.46	1.88
46	2Q8F_A_2Q8H_A_TF4 (S)	0.65	1.25	1.02	2.80	1.73	1.73	1.42

47	2IYT_A_2IYQ_A_ADP_SKM_TRS	5.30	8.04	6.63	2.53	0.67	0.67	0.67
48	1EX6_A_1GKY_A_5GP	3.61	4.57	4.12	2.61	0.66	1.72	0.46
49	1RRG_A_1S9D_A_AFB	2.56	3.93	3.38	2.86	0.52	0.76	0.33
50	1ECJ_D_1ECC_B_PCP	6.76	8.96	7.93	2.37	0.48	0.53	0.48
51	2CM2_A_2H4K_A_509	1.78	3.30	2.61	2.71	2.02	2.46	0.42
52	1NUW_A_1EYJ_B_AMP	1.12	2.30	1.79	2.58	0.14	2.18	0.14
53A	3CJ0_A_2BRL_A_POO	0.46	0.97	0.76	2.57	0.48	0.89	0.17
53B	3CJ0_A_3FQK_B_79Z	0.61	1.48	1.13	2.63	0.62	1.15	0.18
54	1UK2_A_2GZ7_A_D3F	0.70	2.05	1.56	2.71	0.74	1.56	0.47
55	2WGB_A_2V57_A_PRL	3.64	5.37	4.67	2.84	0.46	0.48	0.27
56	1HOO_B_1CIB_A_HDA	2.84	3.82	3.36	2.71	2.07	2.88	0.65
57	1FA9_A_1L5S_B_URC	5.33	6.35	5.88	2.96	2.46	3.28	2.37
58	1W50_A_3IXJ_C_586	2.07	3.27	2.73	2.46	0.69	2.52	0.58
59	3B7D_E_2AL4_F_CX6	0.32	0.90	0.68	2.79	1.58	2.91	1.16
60A	1PKL_B_3HQP_P_ATP_OXL	3.37	4.50	3.97	2.55	0.62	1.79	0.54
60B	1PKL_B_3HQP_P_FDP	1.96	2.05	2.00	2.12	2.25	2.25	1.11
61	1K3F_B_1U1D_F_181	2.64	5.53	4.37	2.60	0.54	3.11	0.54
62	1FVR_A_2OO8_X_RAJ	2.24	3.64	3.02	2.93	0.45	0.45	0.44
63	3HQD_A_1Q0B_B_NAT	0.98	2.93	2.23	2.74	0.40	3.09	0.40
64	2ZB1_A_2NPQ_A_BOG	1.35	2.39	1.97	2.45	0.68	1.91	0.27
65	3PEO_G_2BYS_J_LOB	2.19	3.34	2.86	2.41	0.20	2.04	0.14
66	2BRK_A_2GIR_B_NN3	0.79	1.64	1.35	2.63	0.45	2.81	0.24
67	2H4E_B_3CFN_B_2AN	0.19	1.49	0.97	3.33	2.07	2.56	0.67
68	2CGA_B_1AFQ_C_0FG	1.92	2.94	2.41	2.54	2.33	2.70	1.67
69	1XCG_B_1OW3_B_GDP	2.57	4.01	3.30	2.73	0.99	1.42	0.50
70	1FXX_A_3HL8_A_BBP	0.38	1.46	1.08	2.25	0.57	2.57	0.21
71	4AKE_B_1ANK_B_ANP	3.76	4.48	4.12	1.88	1.72	1.82	1.04
72	3NNU_A_3HL7_A_I46	1.27	2.78	2.18	2.84	0.53	2.87	0.12
73	1A8I_A_2IEG_B_FRY	0.50	1.09	0.90	2.55	0.16	2.08	0.16
74	1SWX_A_2EUM_A_LAT	0.55	0.75	0.66	2.51	1.78	2.62	1.78
75	2QLR_C_3DC1_A_AKG	0.35	0.47	0.41	2.57	0.53	3.43	0.33
76	2F6V_A_1T49_A_892	1.02	2.65	2.00	2.55	0.18	1.84	0.02
77	1G24_D_1GZF_C_NIR	0.39	1.52	1.12	2.95	1.92	2.75	0.96
78	1SU4_A_3FGO_B_ACP	16.95	17.73	17.33	2.67	0.64	2.77	0.11
79	2AIR_H_1ZA1_D_CTP	6.73	8.42	7.63	2.53	0.49	2.12	0.24
80	1E2X_A_1H9G_A_MYR	0.86	2.09	1.62	3.13	0.82	0.82	0.38
81	1MY0_B_1N0T_D_AT1	1.71	1.71	1.71	2.65	1.65	2.47	0.40
82	1ZAH_B_2OT1_D_N3P	0.39	1.50	1.03	2.74	2.39	3.21	0.16
83	4HB2_C_4HAT_C_LMB	0.82	1.27	1.07	1.63	1.47	1.47	0.46
84	1BSQ_A_1GX8_A_RTL	0.42	1.60	1.19	1.48	0.68	1.44	0.27
85	2GPO_A_1S9Q_B_CHD	0.23	1.24	0.93	3.64	0.45	3.01	0.11
86	3FDL_A_2YXJ_A_N3C	0.62	1.88	1.43	3.07	0.60	2.27	0.20
87	1B6B_A_1KUV_A_CA5	2.14	1.71	1.97	2.47	0.64	0.64	0.44
88	1KZ7_D_1GRN_A_AF3	3.08	5.78	4.55	2.58	1.60	3.44	0.74
89	1JWP_A_1PZO_A_CBT	2.34	3.12	2.72	2.60	0.68	0.84	0.24
90	3GXD_B_2WCG_A_MT5	1.22	2.35	1.94	2.42	1.56	2.33	0.82
91	1BNC_B_2V5A_A_LZL	2.38	2.85	2.65	2.73	2.15	2.55	1.39
92	1Z92_A_1PY2_A_FRH	1.12	2.42	1.91	2.81	0.65	1.39	0.23
93	1JBU_H_1WUN_H_P5B	4.65	6.78	5.71	2.74	0.61	2.24	0.61

Table S2. RMSD values and distances from local alignment based on the 9 Å neighborhood of the ligand. The notation used in the second column is PDB ID unbound_Chain_PDB ID bound_Chain_Ligand code. S in parenthesis identifies the proteins with side chain conformational changes only. The additional columns are as follows. **A.** Local backbone RMSD between CryptoSite unbound and bound structures. **B.** Local side chain RMSD between CryptoSite unbound and bound structures. **C.** Local all atom RMSD between CryptoSite unbound and bound structures. **D.** Minimum distance between any atom of in the bound structure and any atom of the bound ligand in the CryptoSite set. **E.** Minimal distance between any atom of the protein and any atom of the superimposed ligand in the CryptoSite unbound structure. **F.** Largest value of the minimal distance between any atom of the protein and any atom of the superimposed ligand in the unbound structures of the extended CryptoSite set. **G.** Smallest value of the minimal distance between any atom of the protein and any atom of the superimposed ligand in the unbound structures of the extended CryptoSite set.

Supplementary Examples

Cryptic site in Interleukin 2

Interleukin-2 (IL-2), which is one of the best-studied examples of a protein with a cryptic site (71, 72). At its interface with the receptor IL-2Rα, the unbound structure of IL-2 (PDB ID 1M47) has two hot spots (Fig. S2A). The stronger, CS0, contains 20 probe clusters and is located in a rigid and largely polar pocket, whereas CS3, with 10 probe clusters, is in an adaptive hydrophobic region (71, 73). Fig. S2A also shows an inhibitor of the IL-2/IL-2Rα interaction, extracted from the bound structure 1PY2 (yellow sticks), superimposed on the mapped ligand-free IL-2 structure, illustrating that the inhibitor would clash with the protein in the unbound conformation and thus that the ligand biding site is cryptic. In Fig. S2B show the mapping of the structure 1Z92, considered unbound in the CryptoSite set (item 92 in Table S2). 1Z92 does not have a bound inhibitor, but it has been co-crystallized with IL2Rα, so it is not entirely an unbound structure. The loop 30-35 of 1Z92 (cyan) approaches the ligand on the left end of the binding site and Lys35 would clash with the ligand. 1Z92A has one hot spot CS0 (green, 23 probe clusters) at the same location as CS0(20) in 1M47, but it has no second hot spot in the flexible nonpolar pocket close to the (left) cryptic region. In the bound structure 1PY2 (magenta) residues 30 and 32-35 participate in helices, and move away from the ligand binding site. Although the major change is in the 30-35 loop region, the backbones of 1Z92 and 1PY2 deviate at several locations, and due to the backbone difference several side chains of 1Z92 protrude into the ligand binding site. In particular, the C-end of the large helix 54 to 72 slightly turns, and the Leu72 side chain protrudes into the site. Although it is difficult to assess how much the binding of the inhibitor contributes to the conformational change between free and bound states, all known IL-2 structures with a fully open cryptic site have inhibitors either bound to both hot spots (1PY2, 1M48, 1PW6, 1M49, and 1QVN), or covalently bound to the hot spot at the cryptic site (1M4A), emphasizing the potential role of these hot spots in ligand binding and the opening of the cryptic site.

Cryptic site in TEM β-lactamase

While the hot spots that the ligand utilizes in the bound structure of IL-2 already exist in the unbound state, for other proteins this is not always the case, as demonstrated by the well-known example of TEM β-lactamase. In the unbound structure of this protein (PDB ID 1JWP), the main hot spot CS0(30) lies in the large, pre-existing substrate binding site of the enzyme (Fig. S2C). However, in bound structures such as 1PZO the position of helix 11 (shown in dark blue in Fig. S2C) substantially shifts, opening a substantial crevice adjacent to the substrate binding site (Fig. S2D). This elongated cryptic site was discovered serendipitously, when crystals

revealed two small inhibitor molecules bound between helices 11 and 12 (66). Mapping this inhibitor-bound structure without the inhibitors shows eight hot spots lining the substrate binding and cryptic sites (Fig. S2D). The strongest hot spot found for the bound structure 1PZO, CS0(16), occupies the same location as the main hotspot seen in the mapped unbound structure, at a distance of 1.1 Å from one of the inhibitors occupying the cryptic site. Thus, β-lactamase satisfies the condition that there is a strong hot spot in the unbound structure near the cryptic site. However, in this instance the ligands that occupy the cryptic site in the bound structure do not overlap with this nearby hot spot. It is nevertheless likely that ligand binding plays some role in the opening of the cryptic site in TEM β-lactamase, as no open cryptic site is seen in any unbound structure of this protein, yet the site can open even under the influence of very weakly binding ligands. For example, an open cryptic site was first observed in the structure of SHV β-lactamase (PDB ID 2FH4), which shares 68% sequence identity with TEM, and binds two molecules of the nonionic detergent Cymal-6 in the cryptic site (66).

Application to Cyclin Dependent Kinase 2 (CDK2)

We have been collaborating with the Wells group of UCSF, who have been interested in finding allosteric kinase inhibitors. In a particular application we focused on the potential druggability of the PIF pocket in various kinases. The PIF pocket is a hydrophobic groove, found on the N-terminal domain of phosphoinositide-dependent kinase 1 (PDK1), which uses this pocket to recruit the C-terminal hydrophobic motif (HM) on other members of the AGC kinase family and thereby regulate their activity through phosphorylation (75). Both activators (75) and inhibitors (76, 77) have been developed for the PIF pocket, with many in the low micromolar affinity range. We have mapped a number of kinases of interest, and found that the strongest hot spot in the PIF pocket was predicted for the cyclin dependent kinase 2 (CDK2). Occupying this pocket would disrupt the interaction between CDK2 and its endogenous activator Cyclin. However, the PIF pocket is cryptic, as its depends on the conformation of the kinase. Fig. S3A shows mapping result for the inactive conformation without ATP (PDB ID 1HCL). In this conformation the PIF pocket includes only the weak hot spot CS4(11), shown as magenta sticks. However, mapping the CDK2 structure from the phosphorylated CDK2-Cyclyin A-substrate peptide complex (PPDB ID 1QMZ) shows that the PIF pocket is more open and includes the second strongest hot spot, CS1(17) after the ATP binding site, which has the hot spot CS0(19) (see Fig. S3B). Based on this result the PIF pocket was deemed druggable (78). Dr. Justin Rettenmaier in the Wells lab at UCSF explored the site using a tethering technique (79) to identify ligands that would bind in that region. This work is not yet published and hence only a brief summary relevant to this paper is given here. Tethering employs a library of disulfide-containing small-molecule fragments (150-300 Da) in order to find ligands that bind at the site of interest (79). Two surface exposed cysteines of CDK2 were mutated to alanine (C118A and C177A) so they would not trap ligands at off-target sites. Based on the location of the hot spot, Lys56 was mutated to Cys (K56C) to enable a tethering screen for allosteric CDK2 ligands. A library of disulfide-containing fragments was

screened for binding to CDK2 using intact protein mass spectrometry. The relative potency of the top 12 confirmed hits was measured by determining the concentration of β-mercaptoethanol required to displace half of the compound from the protein (βME50), where the largest βME50 corresponds to the most potent compound. Next, it was determined whether any of the 12 hits modulated the ability of Cyclin A2 to stimulate CDK2 using a radioactive peptide phosphorylation assay, which revealed that two compounds reduced phosphorylation of the Histone H1 peptide by CDK2-Cyclin by 42% and 81%, respectively. Additional results of this work will be given in a future publication.

Supplementary References

1. Schuttelkopf AW, et al. (2006) Screening-based discovery and structural dissection of a novel family 18 chitinase inhibitor. J Biol Chem 281(37):27278-27285.
2. Benson ML, et al. (2008) Binding MOAD, a high-quality protein-ligand database. Nucleic Acids Res 36(Database issue):D674-678.
3. Allingham JS, Smith R, & Rayment I (2005) The structural basis of blebbistatin inhibition and specificity for myosin II. Nat Struct Mol Biol 12(4):378-379.
4. Sha RS, Kane CD, Xu ZH, Banaszak LJ, & Bernlohr DA (1993) Modulation of Ligand-Binding Affinity of the Adipocyte Lipid-Binding Protein by Selective Mutation - Analysis Invitro and Insitu. J Biol Chem 268(11):7885-7892.
5. Ko DC, Binkley J, Sidow A, & Scott MP (2003) The integrity of a cholesterol-binding pocket in Niemann-Pick C2 protein is necessary to control lysosome cholesterol levels. Proc Natl Acad Sci U S A 100(5):2518-2525.
6. Wang R, Fang X, Lu Y, & Wang S (2004) The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. J Med Chem 47(12):2977-2980.
7. Maruyama D, et al. (2007) Crystal structure of uridine-diphospho-N-acetylglucosamine pyrophosphorylase from Candida albicans and catalytic reaction mechanism. J Biol Chem 282(23):17221-17230.
8. Cotton FA, Hazen EE, Jr., & Legg MJ (1979) Staphylococcal nuclease: proposed mechanism of action based on structure of enzyme-thymidine 3',5'-bisphosphate-calcium ion complex at 1.5-A resolution. Proc Natl Acad Sci U S A 76(6):2551-2555.
9. Miller DM, 3rd, Olson JS, Pflugrath JW, & Quiocho FA (1983) Rates of ligand binding to periplasmic proteins involved in bacterial transport and chemotaxis. J Biol Chem 258(22):13665-13672.
10. Chen Y, et al. (2003) Nucleotide binding to nucleoside diphosphate kinases: X-ray structure of human NDPK-A in complex with ADP and comparison to protein kinases. J Mol Biol 332(4):915-926.

11. Bahnson BJ, Anderson VE, & Petsko GA (2002) Structural mechanism of enoyl-CoA hydratase: three atoms from a single water are added in either an E1cb stepwise or concerted fashion. Biochemistry 41(8):2621-2629.
12. Ramasamy V, Ramakrishnan B, Boeggeman E, & Qasba PK (2003) The role of tryptophan 314 in the conformational changes of beta1,4-galactosyltransferase-I. J Mol Biol 331(5):1065-1076.
13. Sazinsky MH & Lippard SJ (2005) Product bound structures of the soluble methane monooxygenase hydroxylase from Methylococcus capsulatus (Bath): Protein motion in the alpha-subunit. J Am Chem Soc 127(16):5814-5825.
14. Estebanez-Perpina E, et al. (2007) A surface on the androgen receptor that allosterically regulates coactivator binding (vol 104, pg 16074, 2007). Proc Natl Acad Sci U S A 104(47):18874-18874.
15. Lai J, Ghaemi Z, & Luthey-Schulten Z (2017) The Conformational Change in Elongation Factor Tu Involves Separation of Its Domains. Biochemistry 56(45):5972-5979.
16. Vogeley L, Palm GJ, Mesters JR, & Hilgenfeld R (2001) Conformational change of elongation factor Tu (EF-Tu) induced by antibiotic binding. Crystal structure of the complex between EF-Tu.GDP and aurodox. J Biol Chem 276(20):17149-17155.
17. Han HJ, et al. (2010) The Fungal Product Terreic Acid Is a Covalent Inhibitor of the Bacterial Cell Wall Biosynthetic Enzyme UDP-N-Acetylglucosamine 1-Carboxyvinyltransferase (MurA). Biochemistry 49(19):4276-4282.
18. Cook WJ, Walter LJ, & Walter MR (1994) Drug-Binding by Calmodulin - Crystal-Structure of a Calmodulin Trifluoperazine Complex. Biochemistry 33(51):15259-15265.
19. Regni CA, et al. (2009) How the MccB bacterial ancestor of ubiquitin E1 initiates biosynthesis of the microcin C7 antibiotic. Embo J 28(13):1953-1964.
20. Rahman MN, et al. (2009) X-ray Crystal Structure of Human Heme Oxygenase-1 with (2R,4S)-2-[2-(4-Chlorophenyl)ethyl]-2-[(1H-imidazol-1-yl)methyl]-4[((5-trifluoromethylpyridin-2-yl)thio)methyl]-1,3-dioxolane: A Novel, Inducible Binding Mode. J Med Chem 52(15):4946-4950.
21. Katz BA, et al. (2001) A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site. J Mol Biol 307(5):1451-1486.
22. Modis Y, Ogata S, Clements D, & Harrison SC (2003) A ligand-binding pocket in the dengue virus envelope glycoprotein. Proc Natl Acad Sci U S A 100(12):6986-6991.
23. Wernimont AK, et al. (2010) Structures of apicomplexan calcium-dependent protein kinases reveal mechanism of activation by calcium. Nat Struct Mol Biol 17(5):596-U595.
24. Yajima S, et al. (2007) Structure of 1-deoxy-D-xylulose 5-phosphate reductoisomerase in a quaternary complex with a magnesium ion, NADPH and the antimalarial drug fosmidomycin. Acta Crystallographica Section F-Structural Biology and Crystallization Communications 63:466-470.

25. Blaszczyk J, Li Y, Shi GB, Yan HG, & Ji XH (2003) Dynamic roles of arginine residues 82 and 92 of Escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: Crystallographic studies. Biochemistry 42(6):1573-1580.
26. Knoechel TR, et al. (2006) Regulatory roles of the N-terminal domain based on crystal structures of human pyruvate dehydrogenase kinase 2 containing physiological and synthetic ligands. Biochemistry 45(2):402-415.
27. Kim KH, et al. (2007) Structural basis for glutamate racemase inhibition. J Mol Biol 372(2):434-443.
28. Teotico DG, et al. (2009) Docking for fragment inhibitors of AmpC beta-lactamase. Proc Natl Acad Sci U S A 106(18):7455-7460.
29. Holloway DE, Chavali GB, Leonidas DD, Baker MD, & Acharya KR (2009) Influence of Naturally-Occurring 5 '-Pyrophosphate-Linked Substituents on the Binding of Adenylic Inhibitors to Ribonuclease A: An X-Ray Crystallographic Study. Biopolymers 91(12):995-1008.
30. Hou Z, Cashel M, Fromm HJ, & Honzatko RB (1999) Effectors of the stringent response target the active site of Escherichia coli adenylosuccinate synthetase. J Biol Chem 274(25):17505-17510.
31. Ciulli A, Chirgadze DY, Smith AG, Blundell TL, & Abell C (2007) Crystal structure of Escherichia coli ketopantoate reductase in a ternary complex with NADP+ and pantoate bound: substrate recognition, conformational change, and cooperativity. J Biol Chem 282(11):8487-8497.
32. Perez-Dorado I, et al. (2007) Elucidation of the molecular recognition of bacterial cell wall by modular pneumococcal phage endolysin CPL-1. J Biol Chem 282(34):24990-24999.
33. Kato M, Li J, Chuang JL, & Chuang DT (2007) Distinct structural mechanisms for inhibition of pyruvate dehydrogenase kinase isoforms by AZD7545, dichloroacetate, and radicicol. Structure 15(8):992-1004.
34. Hartmann MD, Bourenkov GP, Oberschall A, Strizhov N, & Bartunik HD (2006) Mechanism of phosphoryl transfer catalyzed by shikimate kinase from Mycobacterium tuberculosis. J Mol Biol 364(3):411-423.
35. Renault L, Guibert B, & Cherfils J (2003) Structural snapshots of the mechanism and inhibition of a guanine nucleotide exchange factor. Nature 426(6966):525-530.
36. Choe JY, Fromm HJ, & Honzatko RB (2000) Crystal structures of fructose 1,6-bisphosphatase: mechanism of catalysis and allosteric inhibition revealed in product complexes. Biochemistry 39(29):8565-8574.
37. Di Marco S, et al. (2005) Interdomain communication in hepatitis C virus polymerase abolished by small molecule inhibitors bound to a novel allosteric site. J Biol Chem 280(33):29765-29770.
38. Hang JQ, et al. (2009) Slow binding inhibition and mechanism of resistance of non-nucleoside polymerase inhibitors of hepatitis C virus. J Biol Chem 284(23):15517-15529.

39. Lu IL, et al. (2006) Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. J Med Chem 49(17):5154-5161.
40. Bjorklund C, et al. (2010) Design and synthesis of potent and selective BACE-1 inhibitors. J Med Chem 53(4):1458-1464.
41. Jin R, et al. (2005) Mechanism of positive allosteric modulators acting on AMPA receptors. The Journal of neuroscience : the official journal of the Society for Neuroscience 25(39):9027-9036.
42. Morgan HP, et al. (2010) Allosteric mechanism of pyruvate kinase from Leishmania mexicana uses a rock and lock model. J Biol Chem 285(17):12892-12898.
43. Bu W, Settembre EC, el Kouni MH, & Ealick SE (2005) Structural basis for inhibition of Escherichia coli uridine phosphorylase by 5-substituted acyclouridines. Acta Crystallogr D Biol Crystallogr 61(Pt 7):863-872.
44. Yan Y, et al. (2004) Inhibition of a mitotic motor protein: where, how, and conformational consequences. J Mol Biol 335(2):547-554.
45. Diskin R, Engelberg D, & Livnah O (2008) A novel lipid binding site formed by the MAP kinase insert in p38 alpha. J Mol Biol 375(1):70-79.
46. Hansen SB, et al. (2005) Structures of Aplysia AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations. Embo J 24(20):3635-3646.
47. Le Pogam S, et al. (2006) Selection and characterization of replicon variants dually resistant to thumb- and palm-binding nonnucleoside polymerase inhibitors of the hepatitis C virus. J Virol 80(12):6146-6154.
48. Lima LM, et al. (2010) Identification of a novel ligand binding motif in the transthyretin channel. Bioorg Med Chem 18(1):100-110.
49. Kashima A, et al. (1998) X-ray crystal structure of a dipeptide-chymotrypsin complex in an inhibitory interaction. Eur J Biochem 255(1):12-23.
50. Berry MB, et al. (1994) The closed conformation of a highly flexible protein: the structure of E. coli adenylate kinase with bound AMP and AMPPNP. Proteins 19(3):183-198.
51. Xing L, et al. (2009) Structural bioinformatics-based prediction of exceptional selectivity of p38 MAP kinase inhibitor PH-797804. Biochemistry 48(27):6402-6411.
52. Malinina L, et al. (2006) The liganding of glycolipid transfer protein is controlled by glycolipid acyl structure. PLoS biology 4(11):e362.
53. Han Q, Cai T, Tagle DA, Robinson H, & Li J (2008) Substrate specificity and structure of human aminoadipate aminotransferase/kynurenine aminotransferase II. Biosci Rep 28(4):205-215.
54. Wiesmann C, et al. (2004) Allosteric inhibition of protein tyrosine phosphatase 1B. Nat Struct Mol Biol 11(8):730-737.

55. Menetrey J, et al. (2002) NAD binding induces conformational changes in Rho ADP-ribosylating clostridium botulinum C3 exoenzyme. J Biol Chem 277(34):30950-30957.
56. Laursen M, et al. (2009) Cyclopiazonic Acid Is Complexed to a Divalent Metal Ion When Bound to the Sarcoplasmic Reticulum Ca2+-ATPase. J Biol Chem 284(20):13513-13518.
57. Wang J, Stieglitz KA, Cardia JP, & Kantrowitz ER (2005) Structural basis for ordered substrate binding and cooperativity in aspartate transcarbamoylase. Proc Natl Acad Sci U S A 102(25):8881-8886.
58. van Aalten DMF, Dirusso CC, & Knudsen J (2001) The structural basis of acyl coenzyme A-dependent regulation of the transcription factor FadR. Embo J 20(8):2041-2050.
59. Hogner A, et al. (2003) Competitive antagonism of AMPA receptors by ligands of different classes: Crystal structure of ATPO bound to the GluR2 ligand-binding core, in comparison with DNQX. J Med Chem 46(2):214-221.
60. Sun QX, et al. (2013) Nuclear export inhibition through covalent conjugation and hydrolysis of Leptomycin B by CRM1. Proc Natl Acad Sci U S A 110(4):1303-1308.
61. Kontopidis G, Holt C, & Sawyer L (2002) The ligand-binding site of bovine beta-lactoglobulin: Evidence for a function? J Mol Biol 318(4):1043-1055.
62. Greschik H, Flaig R, Renaud JP, & Moras D (2004) Structural basis for the deactivation of the estrogen-related receptor gamma by diethylstilbestrol or 4-hydroxytamoxifen and determinants of selectivity. J Biol Chem 279(32):33639-33646.
63. Lee EF, et al. (2007) Crystal structure of ABT-737 complexed with Bcl-x(L): implications for selectivity of antagonists of the Bcl-2 family. Cell Death Differ 14(9):1711-1713.
64. Wolf E, De Angelis J, Khalil EM, Cole PA, & Burley SK (2002) X-ray crystallographic studies of serotonin N-acetyltransferase catalysis and inhibition. J Mol Biol 317(2):215-224.
65. Nassar N, Hoffman GR, Manor D, Clardy JC, & Cerione RA (1998) Structures of Cdc42 bound to the active and catalytically compromised forms of Cdc42GAP. Nat Struct Biol 5(12):1047-1052.
66. Horn JR & Shoichet BK (2004) Allosteric inhibition through core disruption. J Mol Biol 336(5):1283-1291.
67. Brumshtein B, et al. (2009) 6-Amino-6-deoxy-5,6-di-N-(N '-octyliminomethylidene)nojirimycin: Synthesis, Biological Evaluation, and Crystal Structure in Complex with Acid beta-Glucosidase. Chembiochem 10(9):1480-1485.
68. Miller JR, et al. (2009) A class of selective antibacterials derived from a protein kinase inhibitor pharmacophore. Proc Natl Acad Sci U S A 106(6):1737-1742.
69. Thanos CD, Randal M, & Wells JA (2003) Potent small-molecule binding to a dynamic hot spot on IL-2. J Am Chem Soc 125(50):15280-15281.
70. Kadono S, et al. (2005) Structure-based design of P3 moieties in the peptide mimetic factor VIIa inhibitor. Biochem Biophys Res Commun 327(2):589-596.

71. Oleinikovas V, Saladino G, Cossins BP, & Gervasio FL (2016) Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. J Am Chem Soc 138(43):14257-14263.
72. Bowman GR & Geissler PL (2012) Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. Proc Natl Acad Sci U S A 109(29):11681-11686.
73. Eyrisch S & Helms V (2007) Transient pockets on protein surfaces involved in protein-protein interaction. J Med Chem 50(15):3457-3464.
74. Kozakov D, et al. (2011) Structural conservation of druggable hot spots in protein-protein interfaces. Proc Natl Acad Sci U S A 108(33):13528-13533.
75. Hindie V, et al. (2009) Structure and allosteric effects of low-molecular-weight activators on the protein kinase PDK1. Nat Chem Biol 5(10):758-764.
76. Bobkova EV, et al. (2010) Discovery of PDK1 kinase inhibitors with a novel mechanism of action by ultrahigh throughput screening. J Biol Chem 285(24):18838-18846.
77. Rettenmaier TJ, et al. (2014) A small-molecule mimic of a peptide docking motif inhibits the protein kinase PDK1. Proc Natl Acad Sci U S A 111(52):18590-18595.
78. Kozakov D, et al. (2015) New Frontiers in Druggability. J Med Chem 58(23):9063-9088.
79. Erlanson DA, Wells JA, & Braisted AC (2004) Tethering: fragment-based drug discovery. Annu Rev Biophys Biomol Struct 33:199-223.

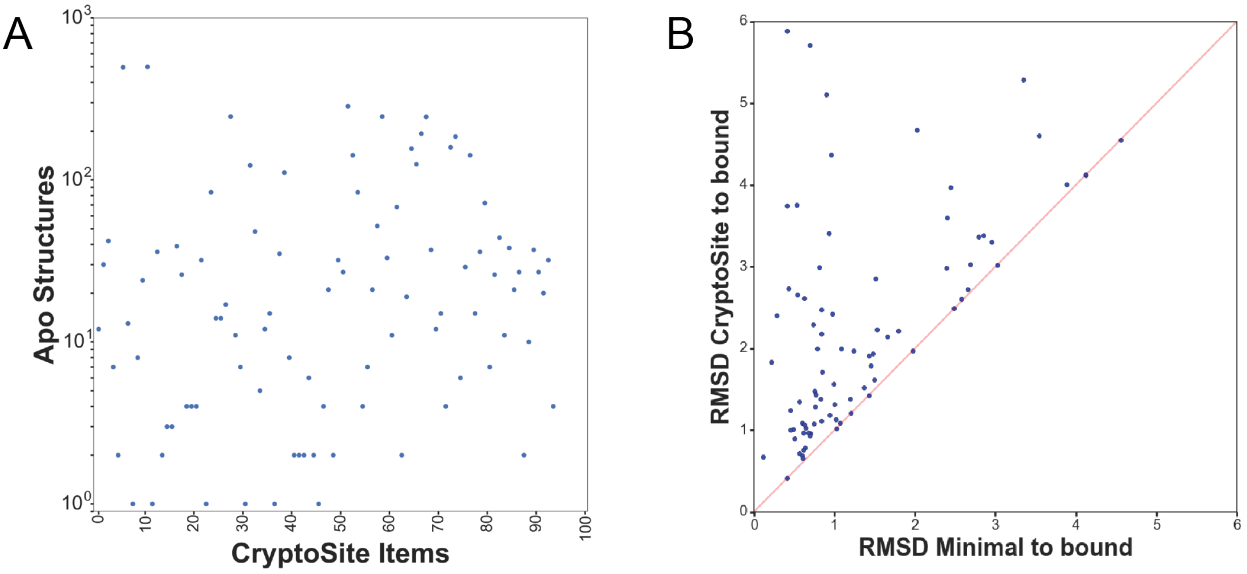

Figure S1. Characterization of the extended CryptoSite set. **A**. Number of unbound structures for each of the 93 protein in the extended CryptoSite set (logarithmic scale). **B.** Local RMSD values calculated for the pairs in the CryptoSite set, versus the smallest RMSD between the bound structure and any unbound structure of the same protein in the extended set.

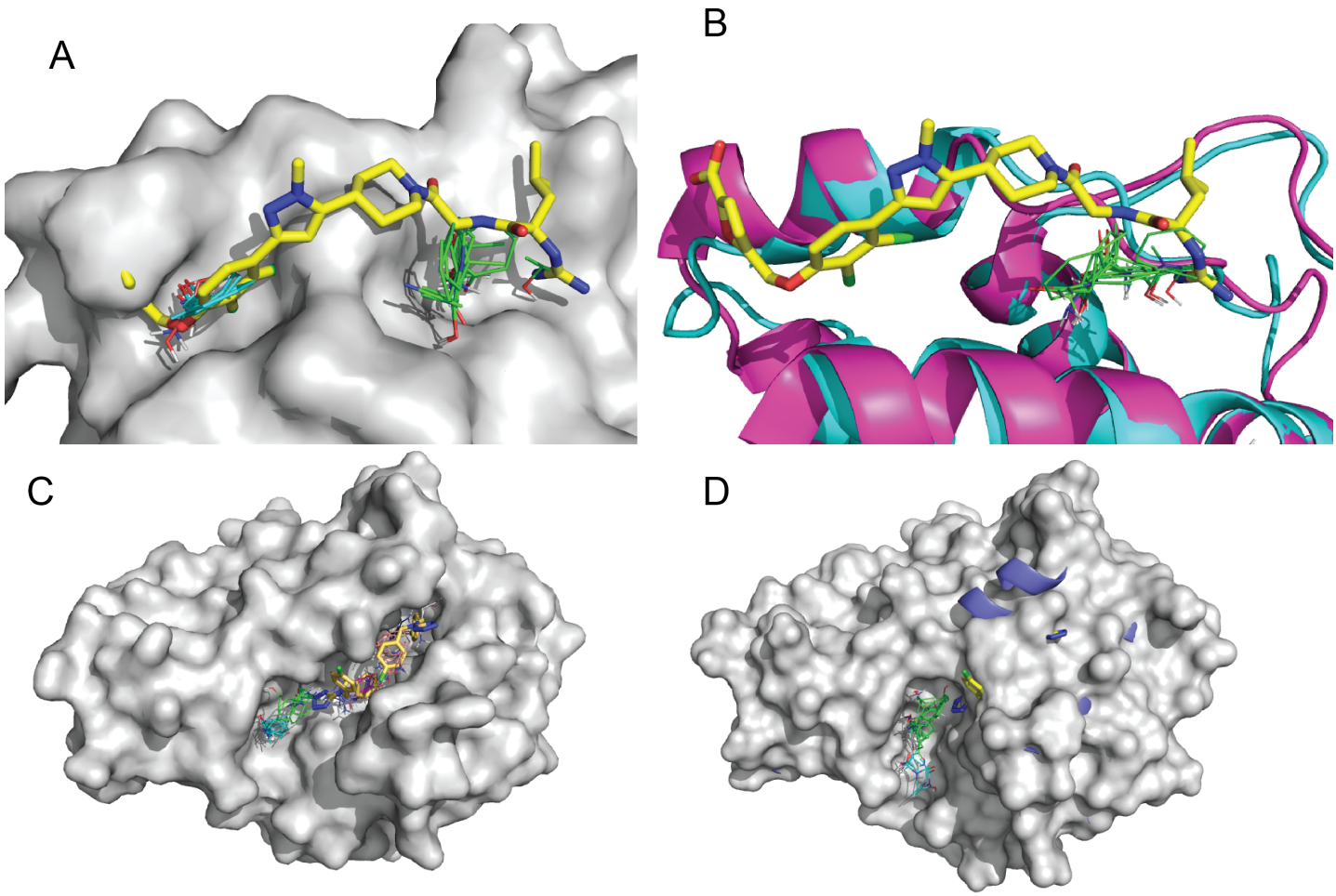

Figure S2. Examples of hot spots near cryptic sites**. A**. Mapping of the structure of unbound IL-2 1M47. The hot spots are CS0 (green, 20 probe clusters), and CS3 (cyan, 10 probe clusters), The inhibitor from the bound structure 1PY2 (yellow sticks) is superimposed to show the steric clashes that would occur. **B.** Mapping of the IL-2 structure 1Z92, considered unbound in the CryptoSite set (Item 92 in Table S2). 1Z92 does not have a bound inhibitor, but it has been co-crystallized with IL2R-α, so it is not entirely an unbound structure. The loop 30-35 of 1Z92 (cyan) protrudes into the left end of the site and also would clash with the ligand. 1Z92A has one hot spot CS0 (green, 23 probe clusters) at the same location as CS0(20) in 1M47, but it has no second hot spot in the flexible nonpolar pocket close to the (left) cryptic region. In the bound structure 1PY2 (magenta) residues 30 and 32-35 participate in helices, and the fragment moves out of the ligand binding site. **C.** Mapping of the unbound TEM β-lactamase structure 1JWP. The two hot spots are CS0 (green, 30) and CS4 (cyan, 8). The large hot spot, CS0, is only 1.1 Å from one of the inhibitors (yellow sticks). **D.** Mapping of the bound TEM β-lactamase structure 1PZO, which binds two small inhibitors shown as yellow sticks. The ligands were removed prior to mapping. The hot spots, with the numbers indicating the numbers of probe clusters, are as follows: CS0 (green, 16), CS1 (cyan, 15), CS2 (magenta, 13), CS3 (orange, 13), CS4(blue, 8), CS5 (lilac, 8), CS7 (white, 6) (see Supplementary Examples).

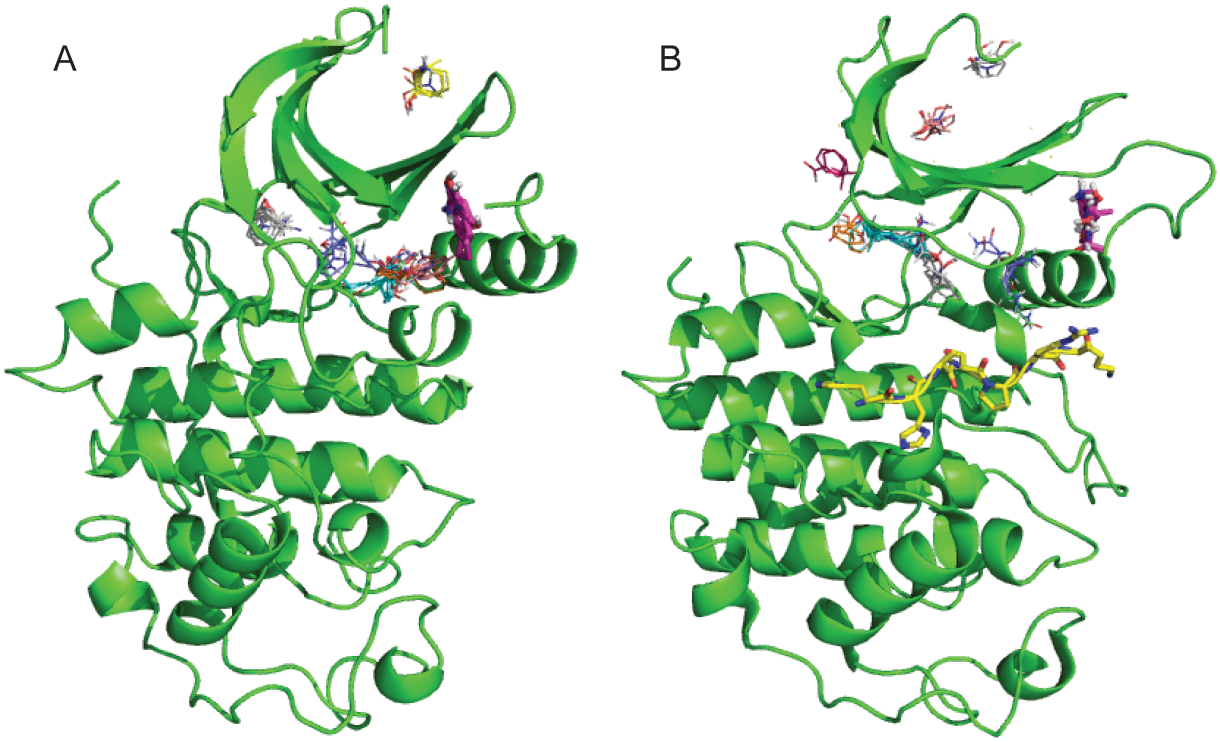

Figure S3. Hot spots of Cyclin Dependent Kinase 2 (CDK2). **A.** Mapping of the inactive CDK2 conformation without ATP (PDB ID 1HCL). **B.** Mapping of the CDK2 structure from the phosphorylated CDK2-Cyclyin A-substrate peptide complex (PPDB ID 1QMZ). The latter shows a strong hot spot 1(19) in the PIF pocket of CDK2 (see Supplementary Examples).