Supporting Information

Structural insights on fragment binding mode conservation

Malgorzata N. Drwal[∞], Guillaume Bret [∞], Carlos Perez[§], Célien Jacquemard[∞], Jérémy Desaphy[#], Esther Kellenberger[∞]*

CONTENT

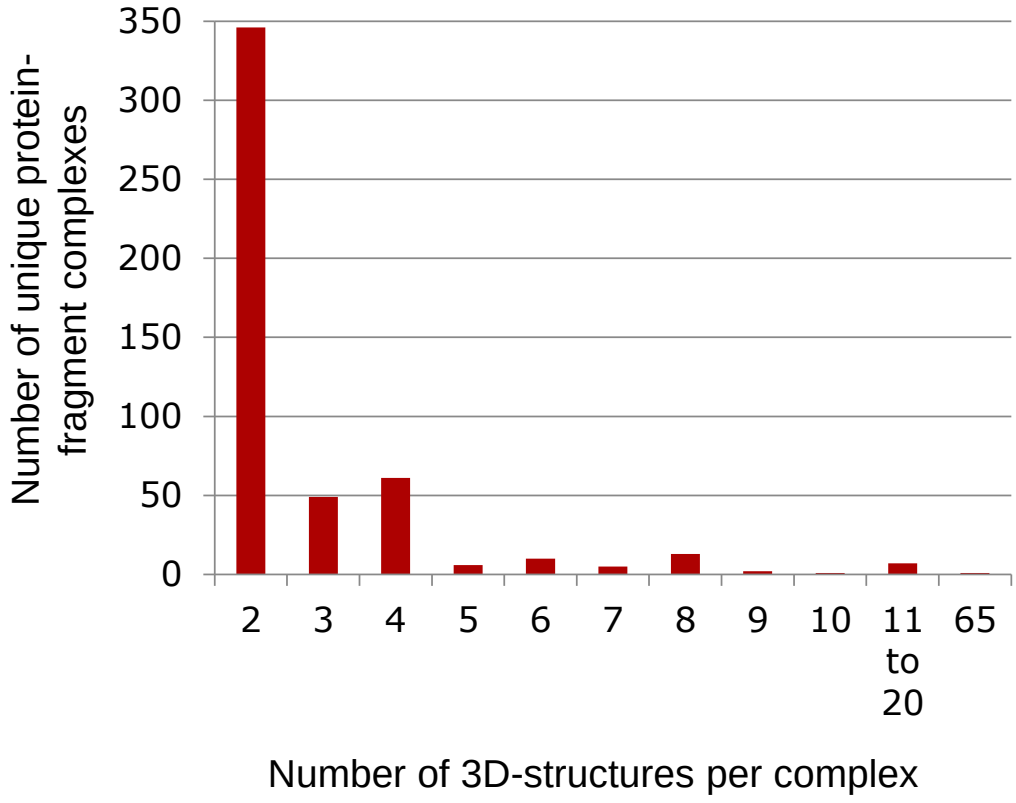
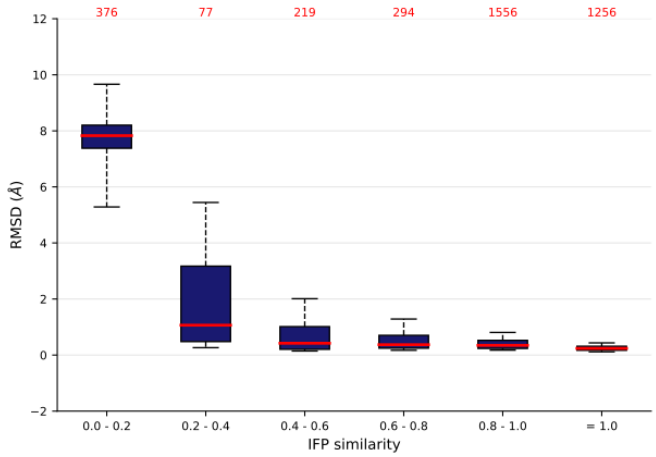


Figure S1. Number of structures available for fragment-protein complexes.

A



B

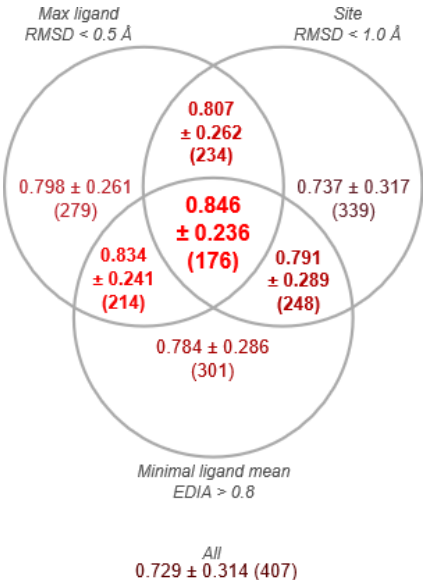


Figure S2. Effect of conformational change and structure quality on the binding mode conservation of the same fragment within the same binding pocket.

(A) Relationship between the conservation of fragments coordinates (RMSD) and the conservation of intermolecular interactions (IFP similarity). RMSD is computed on all the fragment non-hydrogen atoms after structural alignment of the protein to a reference structure. At the top of each boxplot is written in red the number of comparisons in the corresponding IFP similarity interval.

(B) The minimal binding mode similarity (minimal IFP similarity) observed for each of the 407 studied complexes is considered. The figure shows the average minimal IFP similarity before and after applying filters for ligand conformational variability (maximal RMSD after optimal superposition of the fragment poses < 0.5 Å), structural change in the backbone of the binding site (mean site RMSD < 1.0 Å) and coordinates coverage by the electron density (minimal mean EDIA > 0.8). The number of complexes retained in the different filtering scenarios is given in below the average minimal IFP similarity (parenthesis). Note that 94 complexes were discarded from the analysis because the electron density map was not available.

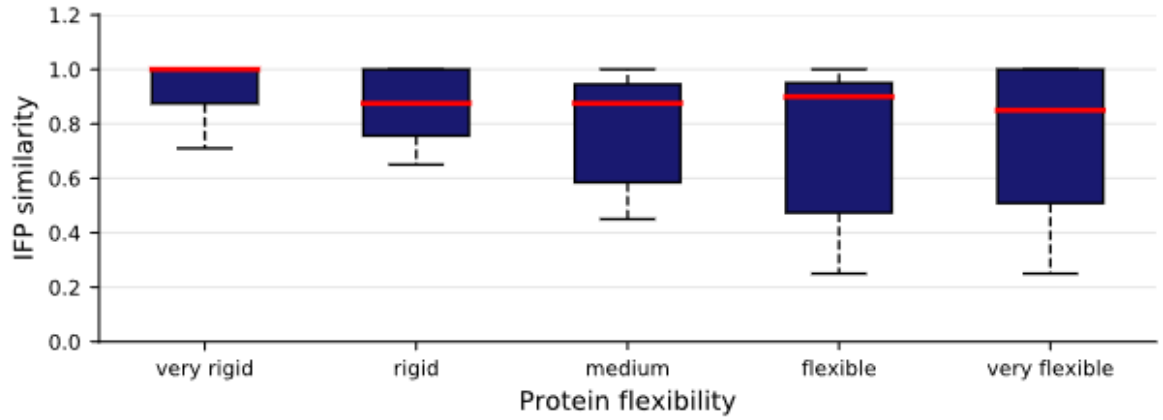


Figure S3. Effect of protein flexibility on binding mode conservation of fragment-ligand substructure pairs.

Pairs of sites are classified as very rigid, rigid, medium, flexible or very flexible if the average Cα RMSD computed on site residues is in the range 0-0.5Å, 0-5-1 Å, 1-1.5 Å, 1.5-2 Å or >2 Å, respectively.

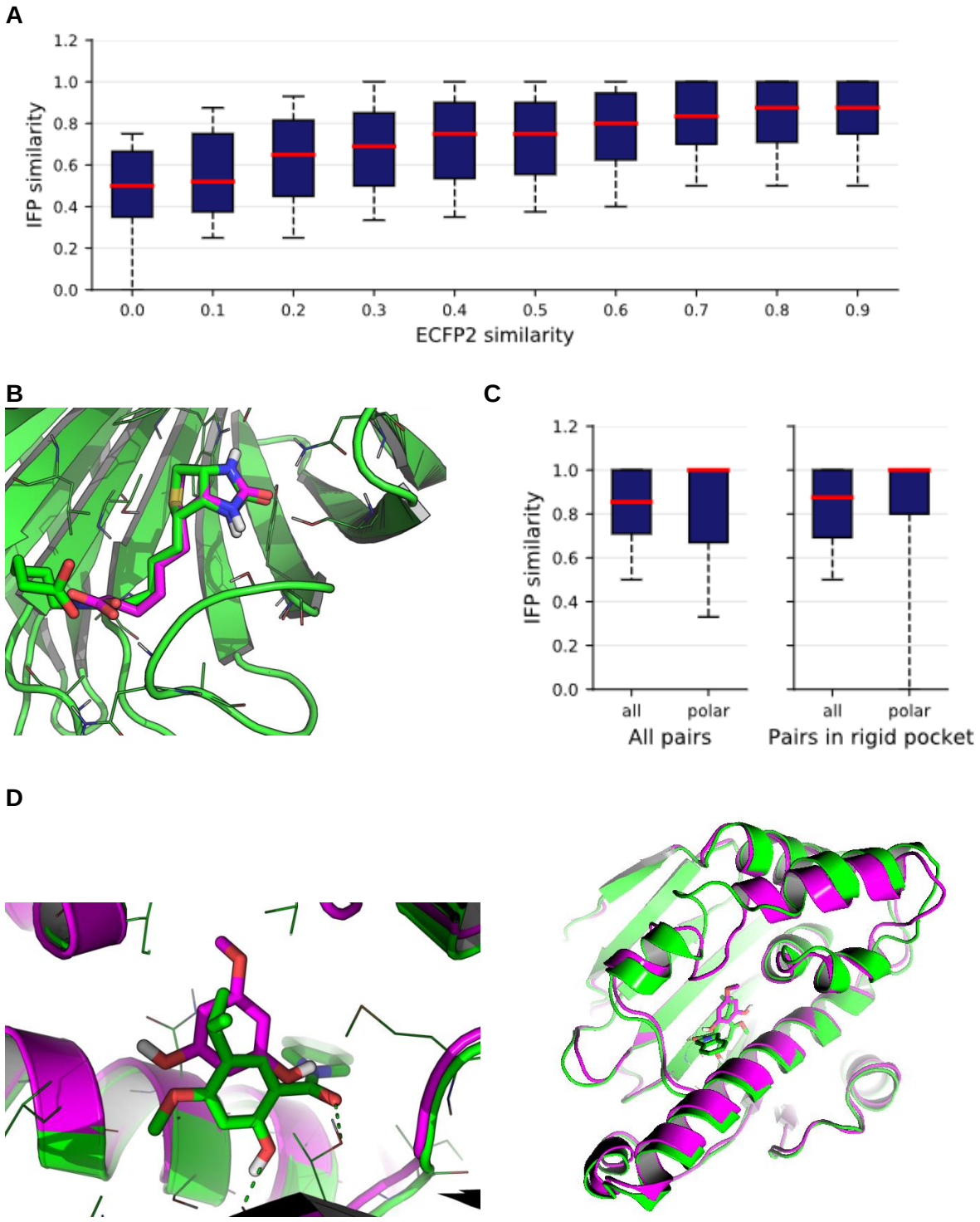
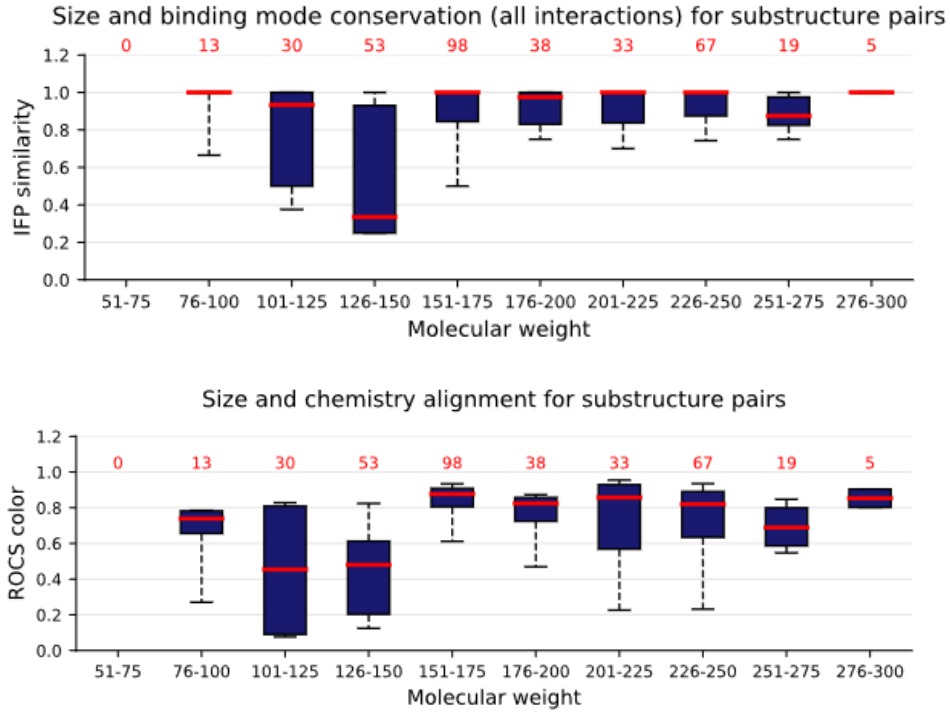


Figure S4. Dependence of fragment-ligand binding mode conservation on chemical similarity.

(A) Relationship between structural similarity and binding mode similarity across all pairs. The chemical similarity is expressed as a proportion of fragment ECFP2 fingerprint bits common to both fragments and ligands. Each box represents ECFP2 similarities of a given interval starting at the X-value (e.g. the first box represents fragment-ligand pairs with structural similarities ≥ 0 and < 0.1). **(B)** The fragment SHM (magenta, PDB: 1LCW) and the ligand BH7 (green, PDB: 1LCZ), binding to *Streptomyces avidinii* streptavidin, show very high structural similarity and high binding mode conservation. **(C)** Binding mode conservation of fragment-ligand pairs with high chemical similarity (ECFP2 similarity ≥ 0.7). Overall

binding mode similarity computed considering all interactions (all) or polar interactions (polar) in all pairs (left) or only pairs in rigid binding site (RMSD Cα < 1Å, right). In (A) and (C), the binding mode similarity is expressed as proportion of common interactions to all fragment interactions (Pr, see Methods section). The median for each substructure pair is shown. **(D)** Example of a fragment-ligand pair with high structural similarity but low binding mode conservation. Fragment 2A7 (PDB: 2YE5, pink) and ligand XJG (PDB: 2XJG, green) bound to human heat shock protein 90. Although the molecules share a common substructure and thus the structural similarity is high (0.82), the binding mode is not conserved. The close view (left) focuses on the common substructure. The general view (right), taken from the opposite side of the protein, shows conformational changes of the protein in the two complexes. Comparing the two complexes reveals large conformational changes of a loop in the binding site (Cα RMSD = 1.9Å).

A



B

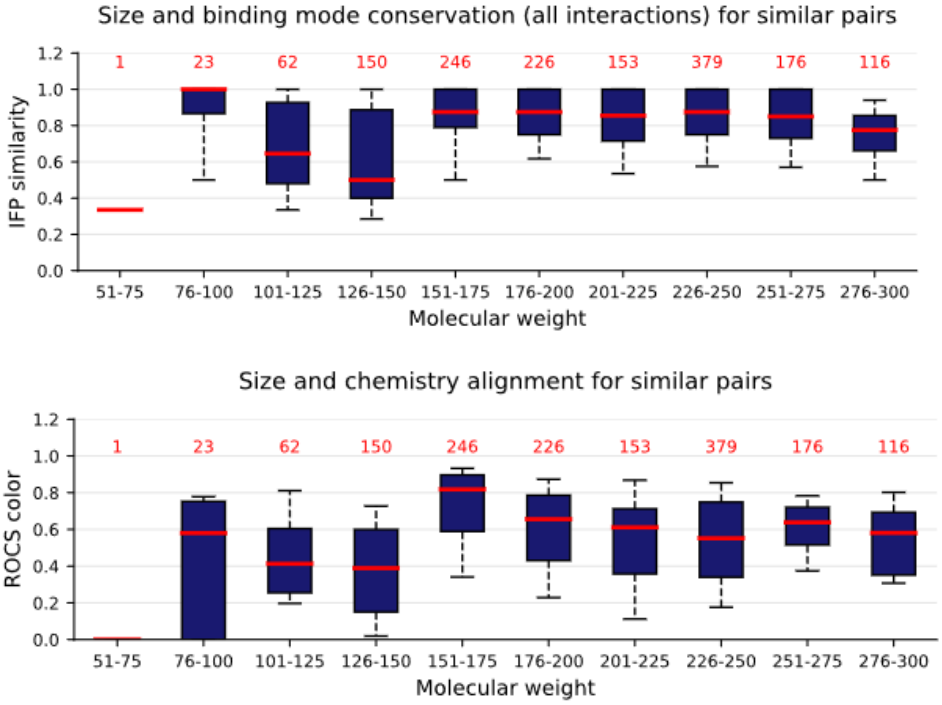
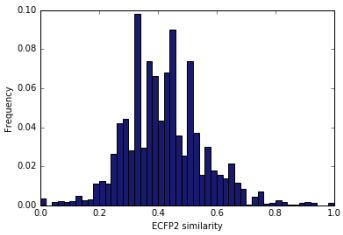


Figure S5. Fragment size and conservation of binding mode and binding pose.

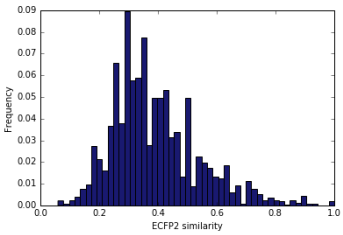
(A) Histograms are generated for the 359 substructures fragment-ligand pairs.

(B) Histograms are generated for the1553 similar fragment-ligand pairs (ECFP2 similarity ≥ 0.7).

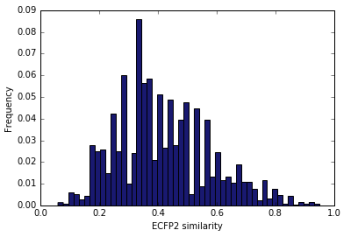
Number of distinct pairs is shown in red.



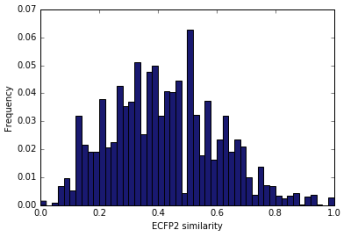
BACE1 (P56817)



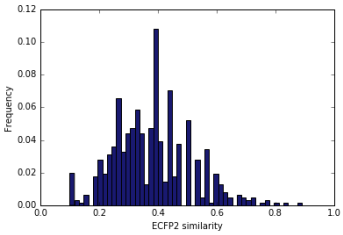
CDK2 (P24941)



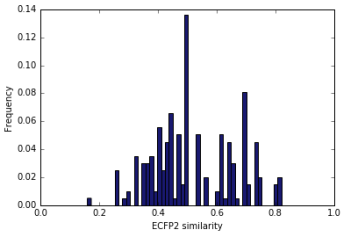
HSP90 (P07900)



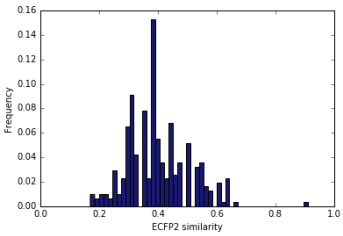
CAH2 (P00918)



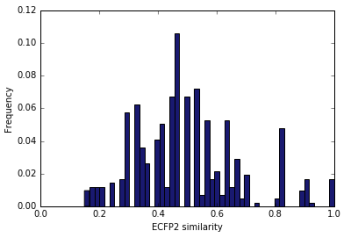
PIM1 (P11309)



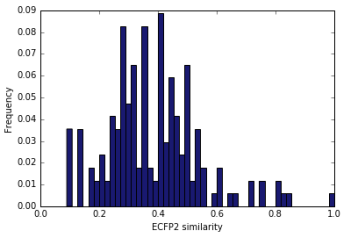
ESR1 (P03372)



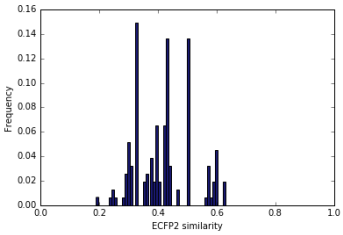
PDE10A (Q9Y233)



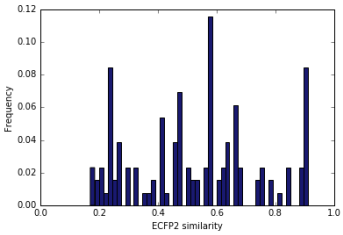
TNKS2 (Q9H2K2)



BRD4 (O60885)



LTA4H (P09960)

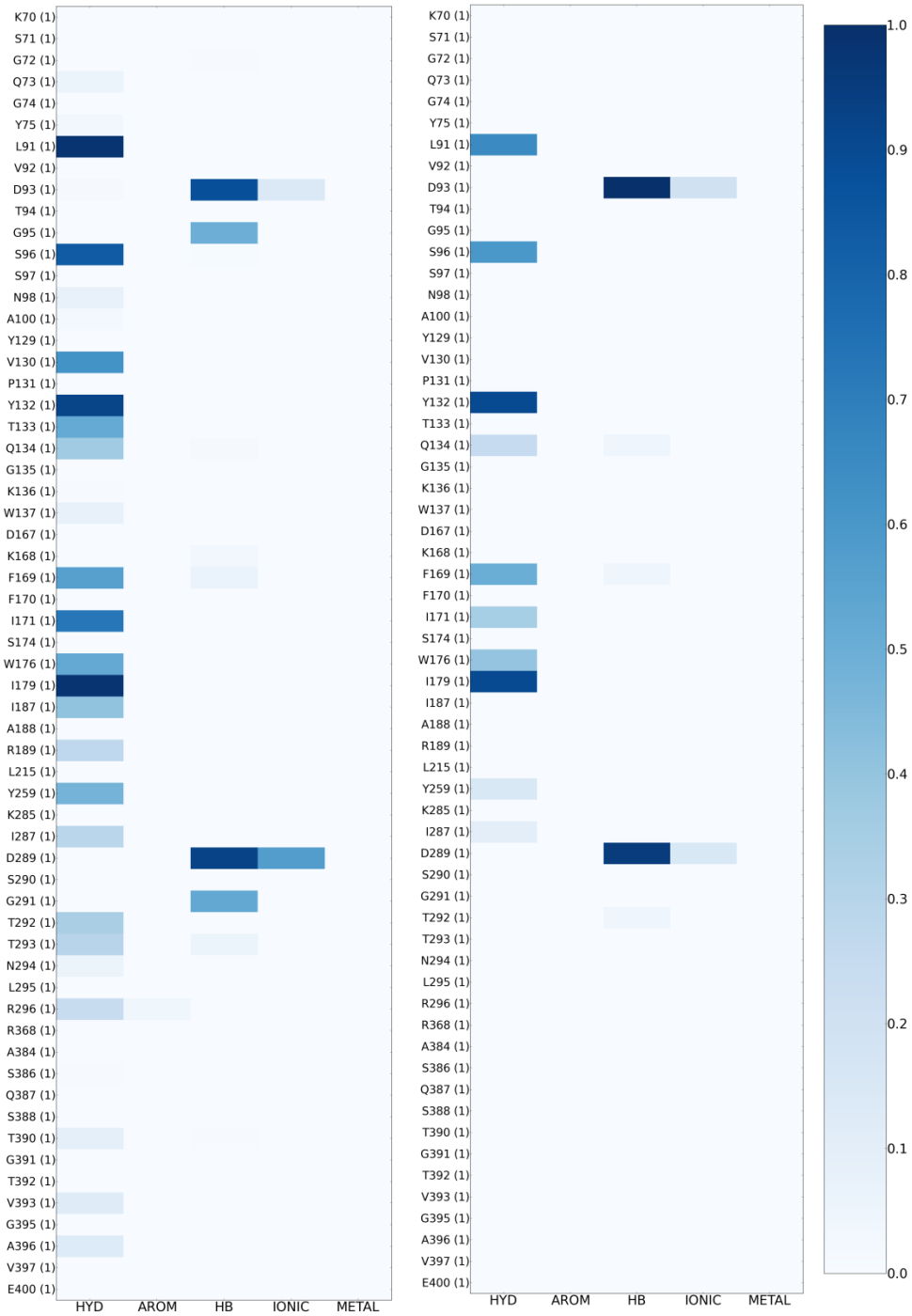


PYGM (P00489)

Figure S6. Chemical similarity between fragment and ligand sets for the fragment- and ligand-rich targets.

All pairwise ECFP2 similarities between fragments and ligands of the given targets are shown. The similarity is expressed as the proportion of common features against all fragment features.

A


K70 (1)
S71 (1)
G72 (1)
Q73 (1)
G74 (1)
Y75 (1)
L91 (1)
V92 (1)
D93 (1)
T94 (1)
G95 (1)
S96 (1)
S97 (1)
N98 (1)
A100 (1)
Y129 (1)
V130 (1)
P131 (1)
Y132 (1)
T133 (1)
Q134 (1)
G135 (1)
K136 (1)
W137 (1)
D167 (1)
K168 (1)
F169 (1)
F170 (1)
I171 (1)
S174 (1)
W176 (1)
I179 (1)
I187 (1)
A188 (1)
R189 (1)
L215 (1)
Y259 (1)
K285 (1)
I287 (1)
D289 (1)
S290 (1)
G291 (1)
T292 (1)
T293 (1)
N294 (1)
L295 (1)
R296 (1)
R368 (1)
A384 (1)
S386 (1)
Q387 (1)
S388 (1)
T390 (1)
G391 (1)
T392 (1)
V393 (1)
G395 (1)
A396 (1)
V397 (1)
E400 (1)
HYD
AROM
HB
IONIC
METAL
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

B

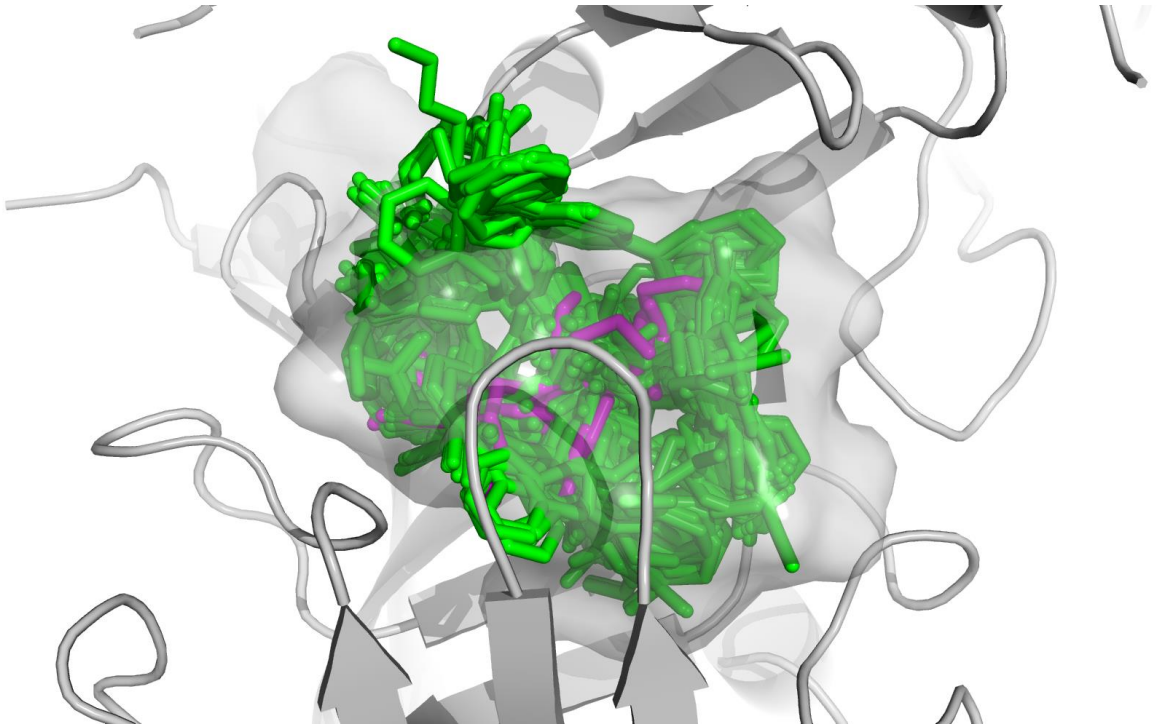


Figure S7. Coverage of human beta secretase pocket.

(A) Drug-like ligand (left) and fragment (right) interaction heatmaps. Different types of interactions (hydrophobic: HYD, aromatic: AROM, hydrogen bonding: HB, ionic: IONIC and metal: METAL) are displayed on the X-axis. The binding site residues (one-letter code, residue number, chain) are displayed on the y-axis. The color intensity describes the frequency of the observed interaction in all complexes for this set (e.g. fragments). **(B)** Overlay of all fragments and drug-like ligands in the reference 3D-structure. Drug-like ligands (green) and fragments (magenta) are shown as sticks, the binding site as surface and the residues of the binding site as lines.

Description of Figure S8 to S10:

The binding modes of additives within the same protein site are highly variable, even if the additives structure is well support by the electronic density (Figure S8). Only 20.6 % of all rigid additives have a conserved binding mode (min. Tanimoto ≥ 0.6). As a consequence, when examining the binding mode conservation of additives in substructure pairs with drug-like ligands, the overall conservation level is also very low (median sim. < 0.3, Figure S9). However, the *apo* additives, which have been mostly crystallized only once in the same protein site and thus have a single binding mode, show higher binding mode conservation than other additives (median sim. = 0.50), especially when considering polar interactions (median sim. = 0.67). Lastly, all the additives crystallized in a protein site do not explore all the subsites targeted by the drug-like ligands of that site. On average, additives cover less than 20% of the interactions observed with drug-like ligands (Figure S10). Nevertheless, provided three or more *apo* additives and drug-like ligands have been crystallized into the same site, almost all the interactions made by the *apo* additives were also found in complexes with drug-like ligands

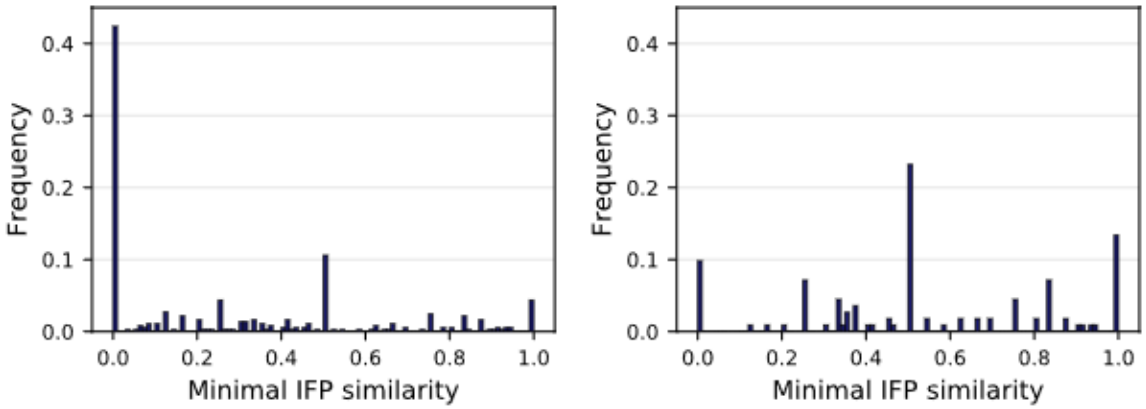


Figure S8. Binding mode conservation of the same additive bound to multiple structures of the same protein site.

The histograms show the frequency of the observed minimal binding mode similarity values. Left: all additives (269 complexes), right: additives which display mean EDIA > 0.8 (112 complexes).

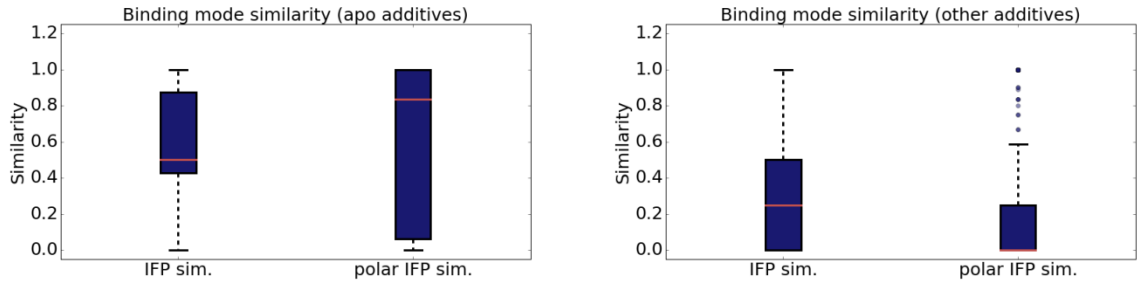
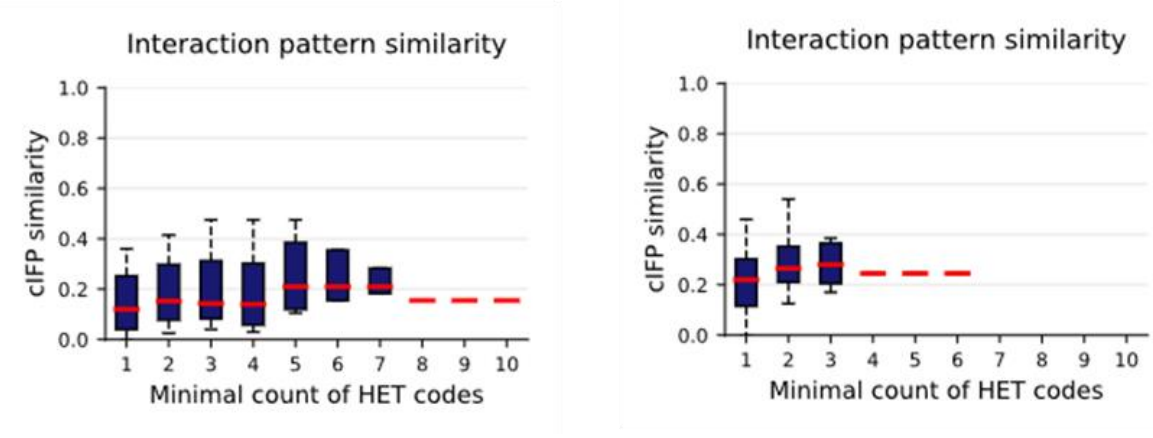


Figure S9. Binding mode conservation of additive-ligand substructure pairs.

Overall binding mode similarity considering *apo* additives (left) and other additives (right). The binding mode similarity is expressed as proportion of common interactions to all fragment interactions for all (IFP sim.) and polar (polar IFP sim.) interactions (Pr, see Methods section). The median for each substructure pair is shown. The *apo* additive set consists in only 26 substructure pairs with the molecules acetic acid (ACY), dodecane (D12), 1,2-ethanediol (EDO), ethanol (EOH), glycerol (GOL) and acetohydroxamic acid (HAE). Only GOL and HAE have been crystallized multiple times within the same pocket and they express low and intermediate binding mode conservation, respectively (GOL & P00918: min. Tanimoto = 0.415; GOL & P9WIL5: min. Tanimoto = 0.625; HAE & P39900: min. Tanimoto = 0.5).

A



B

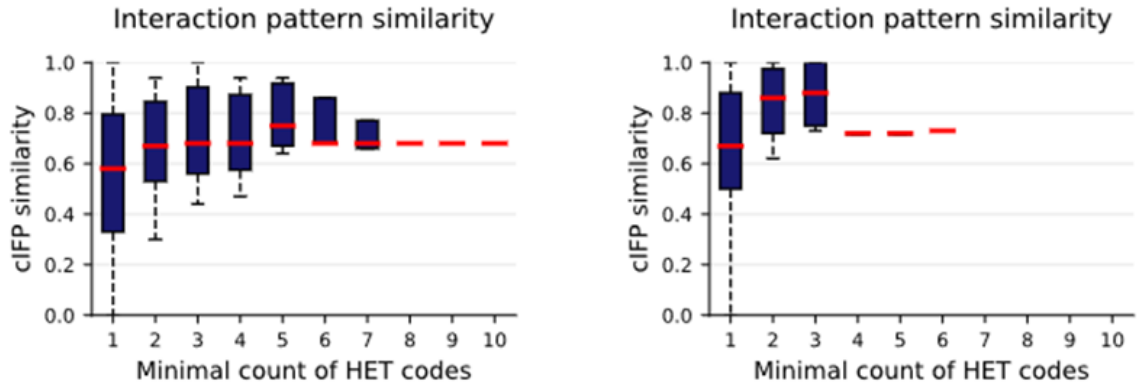


Figure S10. Dependence of additive-drug-like ligand interaction pattern similarity on the number of different molecules considered for all additives (left) and *apo* additives (right).

In each step, all targets with at least X (1, 2, 3, *etc.*) HET codes of both fragments and ligands are considered. **(A).** Tanimoto similarity. **(B).** Proportion similarity

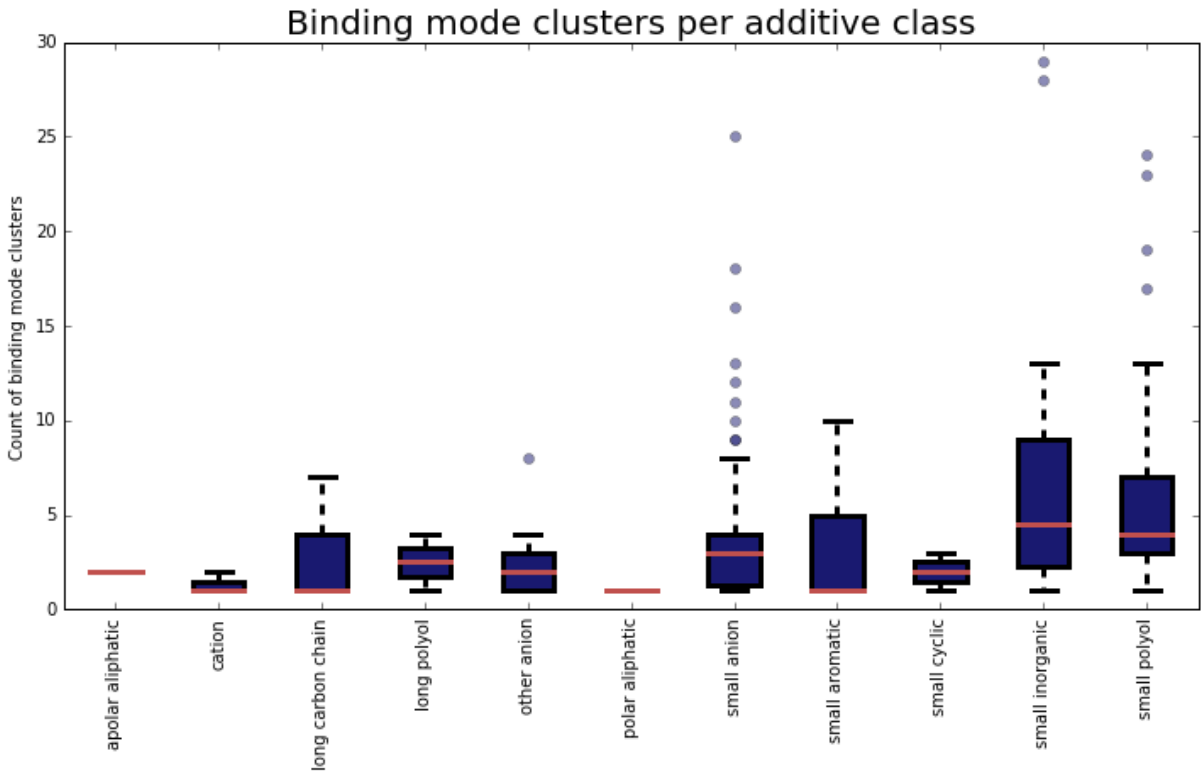


Figure S11. Variability of additive binding modes in multiple structures of the same protein site.

Number of clusters is given per type of additive. The HET codes are: polar aliphatic (148), cation (TRS, SPD, SPM), small anion (CO3, PO4, TFA, POP, SO4, ACT, BCT, ACY, 2PL, GLV, NO3), long carbon chain (MYR, FOH), small cyclic (TMH, MES), long polyol (PG4, 1PE), other anion (NLP, MLI, DXP, CIT, FLC, 2OG, OGA, SIN, 2HG), apolar aliphatic (NBN), small aromatic (1PB, BEZ, PNZ, PMP, IMD), small polyol (GOL, MPD, IPA, EDO, PGE, MRD), small inorganic ('FMT, DMS, ACN, HAE).

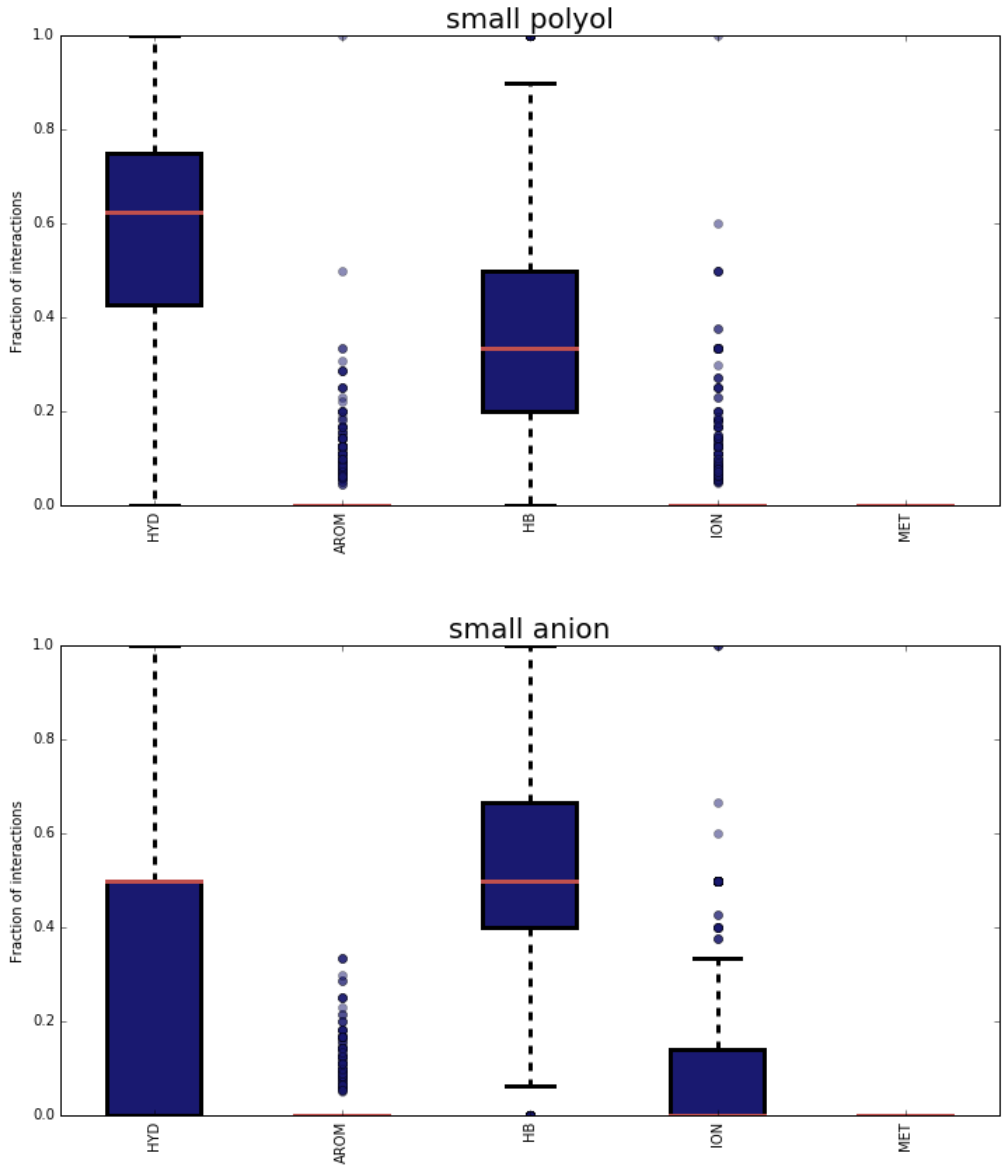


Figure S12. Nature of interaction made by additives.

Proportion of hydrophobic contacts (HYD), pi-pi and pi cation interactions (AROM), H-bonds (HB), ionic bonds (ION) and interaction with metal cation (MET) are shown for additives in the small polyol (upper panel) and small anion (lower panel) classes.

Table S1. 3D-structure quality evaluated using the electron density support for individual atoms (EDIA).

structure quality	Top			Median			Mean		
	Bad	Medium	Good	Bad	Medium	Good	Bad	Medium	Good
protein binding site	0.0	2.5	97.5	0.1	2.4	97.5	0.1	4.6	95.3
drug-like ligand	0.0	13.1	86.9	0.4	12.8	86.9	0.1	16.7	83.2
fragment	0.0	18.6	81.4	0.2	17.2	82.5	0.2	20.4	79.4
small additives	0.0	41.6	58.4	0.1	34.9	65.0	0.1	40.8	59.1
apo additives	0.0	24.6	75.4	0.0	22.8	77.2	0.0	27.2	72.8

Distribution of binding site, drug-like ligand, fragment, small and apo crystallization additive in three categories: bad ($0 \leq EDIA < 0.4$, atom shows substantial inconsistencies with the electron density fit), medium ($0.4 \leq EDIA < 0.8$, atom shows minor inconsistencies with the electron density fit) and good ($0.8 \leq EDIA \leq 1.2$, atom is well covered with the electron density). For each structure are considered the most populated category (top), the category defined by the median EDIA (Median), and the category defined by the mean EDIA (Mean).

Table S2. EDIA score of the complexes described in the figures

Figure	PDB ID	Ligand HET code	Protein Uniprot ID	ligand mean EDIA	binding site mean EDIA
Figure 2B	2HD1	IBM (biounit 1)	O76083	0.99 ± 0.12	0.93 ± 0.21
Figure 2B	2HD1	IBM (biounit 2)	O76083	0.98 ± 0.14	0.96 ± 0.22
Figure 2C	1DZ4	CAM	P00183	0.98 ± 0.10	0.99 ± 0.09
Figure 2C	4JX1	CAM	P00183	0.77 ± 0.27	0.94 ± 0.16
Figure 3B left	1PXJ	CK2	P24941	0.92 ± 0.15	0.87 ± 0.20
Figure 3B right	2C5O	CK2 (biounit 2)	P24941	0.61 ± 0.21	0.91 ± 0.24
Figure 3B	2C5N	CK8 (biounit 1)	P24941	0.83 ± 0.27	0.93 ± 0.23
Figure S5B	1LCZ	BH7	P22629	NO MAP	NO MAP
Figure S5B	1LCW	SHM	P22629	NO MAP	NO MAP
Figure 6 left	1JIZ	CGS	P39900	0.83 ± 0.17	0.86 ± 0.27
Figure 6 left	1OS2	HAE	P39900	0.83 ± 0.27	0.99 ± 0.14
Figure 6 left	1OS2	HAE	P39900	0.86 ± 0.26	0.99 ± 0.14
Figure 6 right	3FGD	BYA	P00800	0.92 ± 0.07	0.99 ± 0.06
Figure 6 right	3N21	PGO	P00800	0.83 ± 0.18	0.99 ± 0.10
Figure 6 right	2A7G	ACY	P00800	0.87 ± 0.13	1.01 ± 0.07
Figure 7 left	3B2X	MLI	P04825	0.79 ± 0.18	1.02 ± 0.12
Figure 7 left	3B34	MLI	P04825	0.62 ± 0.18	1.04 ± 0.12
Figure 7 left	3B3B	MLI	P04825	0.82 ± 0.18	1.00 ± 0.15
Figure 7 left	4XN8	MLI	P04825	0.70 ± 0.23	0.97 ± 0.22
Figure 7 left	4XN8	MLI	P04825	0.7.0 ± 0.23	0.99 ± 0.16
Figure 7 left	4XNA	MLI	P04825	0.71 ± 0.2	0.91 ± 0.16
Figure 7 left	4XNB	MLI	P04825	0.72 ± 0.25	1.00 ± 0.15
Figure 7 left	4XND	MLI	P04825	0.73 ± 0.24	1.03 ± 0.11
Figure 7 left	4XND	MLI	P04825	0.73 ± 0.24	1.00 ± 0.17
Figure 7 right	4Q4E	BB2	P04825	0.94 ± 0.17	1.00 ± 0.15
Figure 7 right	3B2P	GOL	P04825	0.82 ± 0.17	1 ± 0.15
Figure 7 right	3B2X	GOL	P04825	0.78 ± 0.20	1.04 ± 0.09
Figure 7 right	3B34	GOL	P04825	0.73 ± 0.22	1.05 ± 0.08
Figure 7 right	3KED	GOL	P04825	0.78 ± 0.15	0.93 ± 0.2.0
Figure 7 right	3QJX	GOL	P04825	0.67 ± 0.23	1.02 ± 0.05
Figure 7 right	4Q4E	GOL	P04825	0.75 ± 0.20	1.00 ± 0.14
Figure 7 right	4XN7	GOL	P04825	0.76 ± 0.17	0.95 ± 0.12
Figure 7 right	4XNB	GOL	P04825	0.75 ± 0.21	1.01 ± 0.13
Figure 7 right	4XND	GOL	P04825	0.82 ± 0.17	0.99 ± 0.17
Figure 7 right	4XO3	GOL	P04825	0.87 ± 0.08	0.98 ± 0.15
Figure 8 left	4CCB	OFG	Q9UM73	NO MAP	NO MAP
Figure 8 right	2YEK	EAM	P25440	1.02 ± 0.11	1.00 ± 0.14
Figure 8 right	2YEK	EAM	P25440	1.02 ± 0.11	1.01 ± 0.12
Figure 8 right	2YEK	EAM	P25440	0.98 ± 0.15	1.00 ± 0.14
Figure S5D	2YE5	2A7	P07900	0.65 ± 0.23	1.01 ± 0.09
Figure S5D	2XJG	XJG	P07900	0.92 ± 0.14	0.93 ± 0.16

Table S3. Properties of ligand- and fragment-rich targets.

Target	Fragment set diversity[1]	Ligand set diversity[1]	Max. (median) sim.[2] fragments and ligands	Residues in pocket	Av. pocket volume (Å^3)	Pocket volume variation (Å^3)	Av. Cα RMSD to ref. (Å)	Av. polar points (%)[3]
BACE1	0.781	0.765	1 (0.41)	60	739.0	985.5	0.977	62.461
CDK2	0.787	0.765	1 (0.36)	48	539.6	495.4	1.147	48.175
HSP90	0.798	0.777	0.95 (0.40)	48	540.8	706.8	1.255	47.457
CAH2	0.805	0.74	1 (0.40)	38	326.7	324.0	0.246	62.445
PIM1	0.799	0.8	0.89 (0.38)	42	510.3	475.9	0.878	37.633
ESR1	0.707	0.672	0.82 (0.50)	48	416.9	377.1	1.354	28.727
PDE10A	0.798	0.729	0.91 (0.38)	46	555.3	632.4	0.465	56.076
TNKS2	0.705	0.712	1 (0.47)	45	455.1	300.4	0.927	57.773
BRD4	0.821	0.775	1 (0.35)	25	388.3	364.5	0.441	42.745
LTA4H	0.783	0.645	0.63 (0.42)	58	575.4	564.5	0.176	65.276
PYGM	0.629	0.683	0.91 (0.53)	55	486.7	216.0	0.402	71.938

[1] defined as the average diversity (1 – Tanimoto) of all molecules using ECFP2 fingerprints; [2] maximal Tanimoto similarity between molecules of the two sets using ECFP2 fingerprints; [3] average percentage of polar pocket points calculated for each protein cavity using IChem

Table S4. Unique polar interactions of targets with many fragments and ligands

Target	Unique/all fragment hydrogen bonds	Unique/all fragment aromatic bonds	Percentage polar interaction points (pocket)
BACE1	0.167	0	62.461
CDK2	0	0	48.175
HSP90	0.222	0.25	47.457
CAH2	0.364	0.4	62.445
PIM1	0	0	37.633
ESR1	0.2	0	28.727
PDE10A	0.333	0	56.076
TNKS2	0	0.167	57.773
BRD4	0	1	42.745
LTA4H	0.333	0	65.276
PYGM	0.385	-	71.938

Polar pockets (> 50% of polar interaction points) are highlighted in green.