Supplementary Information for

Comprehensive Analysis of Hot Loops at Protein-Protein Interfaces as Targets for Macrocyclic Inhibitors

Jason Gavenonis*, Brad Sheneman*, Timothy R. Siegert*, Matt Eshelman and Joshua Kritzer
*These authors contributed equally to this work

Department of Chemistry
Tufts University
Medford, Massachusetts 02155

	Page
SUPPLEMENTARY NOTE	
Defining loop structures by PDBeMotif	2
SUPPLEMENTARY RESULTS	6
Supplementary Figure 1. Workflow for using LoopFinder to identify hot loops at PPI interfaces.	6
Supplementary Table 1. The full analysis of hot spots from the total interface loop data set of 25,005 structures.	7
Supplementary Table 2. Identical analysis of hot spots conducted for the data set of hot loops only.	8
Supplementary Figure 2. Relative contributions of hot loops compared to total interface energy.	9
Supplementary Figure 3. Selected hot loops containing unusual turn motifs	10
Supplementary Figure 4. Amino acid abundance relative to surface propensity for the hot loop set.	11
Supplementary Figure 5. Comparison of amino acid abundance between hot loops and all interface loops.	12
Supplementary Figure 6. Loop-mediated PPIs by functional class.	13
Supplementary Figure 7. An alignment of MSL1 across species shows that the region identified as a hot loop is highly conserved across species.	14
Supplementary Figure 8. A second hot loop in the MOF complex.	14
Supplementary Table 3. Comparison to experimental alanine scanning for hGH-hGHbp complexes.	15
Supplementary Table 4. Checking reproducibility using the online server Robetta.	16
Supplementary Table 5. Comparison of LoopFinder results to HippDB results.	18
Supplementary Table 6. Analysis of all interface loops and the hot loop set with respect to protein function.	24
Supplementary Information References	28
Supplementary Data Set 1 (caption) The entire set of hot loops generated by LoopFinder.	28
Supplementary Data Set 2. (caption) The subset of 364 hot loops that do not contain two or more consecutive hot spots.	29

SUPPLEMENTARY NOTE

Defining loop structures using PDBeMotif

In order to characterize the types of loops identified as hot loops, the list of 1,407 hot loops were subjected to analysis using the PDBeMotif program.[3] For this process, the PDB file for each identified hot loop was truncated with a short Python script to include only the residues in the hot loop and the three amino acids directly N-terminal and C-terminal to the identified hot loop. This results in an instant analysis of backbone torsional angles and hydrogen bonding patters that exist only within the loop in question. Below is a list of each type of loop as defined by PDBeMotif as well as the requirements used to identify each type of loop.[3] All comments are reproduced from the PDBeMotif definition file. The "nest" and "niche" types of loops were categorized for the purpose of this paper as non-canonical loops due the fact that they lack any organized structural elements such as secondary structure or hydrogen bonding interactions within the loop. In addition to giving analysis of loop structure, PDBeMotif also gives a read out on secondary structure. In this case, if the loop in question was part of an α-helix, they were defined as such in the analysis, making up the final category of structures identified by LoopFinder.

αβ-motif:
A motif of 5 consecutive residues and two H-bonds in which:
-H-bond between CO of residue (i) and NH of residue (i+4)
-H-bond between CO of residue (i) and NH of residue (i+3)
-φ angles of residue (i+1), (i+2) and (i+3) are negative

asx-motif:
A motif of 5 consecutive residues and two H-bonds in which:
-residue (i) is aspartate or asparagine (Asx)
-side-chain O of residue (i) is H-bonded to the main-chain NH of residue (i+2) or (i+3)
-main-chain CO of residue (i) is H-bonded to the main-chain NH of residue (i+3) or (i+4)

Comments: A common feature of the C- and N-termini of α-helices.

asx-turn:
A motif of three consecutive residues and one H-bond in which:
-residue (i) is Aspartate or Asparagine (Asx)
-the side-chain O of residue (i) is H-bonded to the main-chain NH of residue (i+2)
Sub-categories:
Type I:
residue (i): $-140^\circ < \chi 1 - 120^\circ < -20^\circ$ $\quad -90^\circ < \psi + 120^\circ < 40^\circ$

residue (i+1): $-140^\circ < \varphi < -20^\circ$ $-90^\circ < \psi < 40^\circ$

Type I':

Left-handed form of Type I

Type II:

residue (i): $-140^\circ < \chi 1 - 120^\circ < -20^\circ$ $80^\circ < \psi + 120^\circ < 180^\circ$

residue (i+1): $20^\circ < \varphi < 40^\circ$ $-40^\circ < \psi < 90^\circ$

Type II':

Left-handed form of Type II

Comments: about half of the asx-turns are found at the N-termini of α-helices

beta-bulge:

A motif of three residues within a β-sheet in which the main chains of two consecutive residues are H-bonded to that of the third, and in which the dihedral angles are as follows:

residue (i): $-140^\circ < \varphi < -20^\circ$ $-90^\circ < \psi < 40^\circ$

residue (i+1): $-180^\circ < \varphi < -25^\circ$ or $120^\circ < \varphi < 180^\circ$

$40^\circ < \psi < 180^\circ$ or $-180^\circ < \psi < -120^\circ$

Comments: classic β-bulges are mostly found at the edges of the sheets

beta-bulge-loop(5 or 6):

A motif of three residues within a β-sheet consisting of two H-bonds in which:

-the main-chain NH of residue (i) is H-bonded to the main-chain CO of residue (i+4) (Type 5) or residues (i+5) (Type 6)

-the main-chain CO of residue (i) is H-bonded to the main-chain NH of residue (i+3) (Type 5) or residue (i+4) (Type 6)

Type 1 β-bulge loops have an RL nest at residues i+2 and i+3

Type 2 β-bulge loops have an RL nest at residues i+3 and i+4

Comments: β-bulge loops often occur at the loop ends of β-hairpins

Beta-turn:

A motif of four consecutive residues that may contain one H-bond, which, if present, is between the main-chain CO of the first residue and the main-chain NH of the fourth. It is characterized by the dihedral angles of the second and third residues, which are the basis for the sub-categorization:

Sub-categories:

Type I:

residue (i+1): $-140^\circ < \varphi < -20^\circ$ $-90^\circ < \psi < 40^\circ$

residue (i+2): $-140^\circ < \varphi < -20^\circ$ $-90^\circ < \psi < 40^\circ$

Type I':

Left-handed form of Type I

Type II:

residue (i+1): $-140^\circ < \varphi < -20^\circ$ $80^\circ < \psi < 180^\circ$

residue (i+2): $20^\circ < \varphi < 140^\circ$ $-40^\circ < \psi < 90^\circ$

Type II':

Left-handed form of Type II

Comments: The most common of the small protein motifs. The service presents hydrogen bonded motifs only at the distance between the first and last residues C α atoms must be less than 7 Å. Also the dihedral angles were restricted to not fall into the helical region for the first and the last residues or for the middle two residues.

Gamma-turn:

A motif of three consecutive residues i, i+1, i+2 and one H-bond in which:

-the main-chain O of residue (i) is H-bonded to the main-chain NH of residue (i+2)

Sub-categories:

Classic:

residue (i+1): $35^\circ < \varphi < 115^\circ$ $-104^\circ < \psi < -24^\circ$

Inverse:
residue (i+1): -115˚ < φ < -35˚ 24˚ < ψ < 104˚
Comments:

Nest:
A motif of two consecutive residues with dihedral angles as follows:
Sub-categories:
Type RL:
residue (i): -140˚ < φ < -20˚ -90˚ < ψ < 40˚
residue (i+1): 20˚ < φ < 140˚ -40˚ < ψ < 90˚
Type LR:
In LR nests the values for (i) and (i+1) are interchanged
Comments: Nest should not have any prolines. The nest can form a binding site for an anionic group ('the egg'). Nests frequently occur as parts of other motifs such as schellmann loops.

Niche-3l:
A motif of 3 consecutive residues with dihedral angles as follows:
residue (i-1): no limitations
residue (i): -120˚ < φ < -60˚ -50˚ < ψ < 30˚
residue (i+1): -100˚ < φ < -50˚ 110˚ < ψ < 170˚
Comments: This version of the database contains motifs with the amended definition. The criteria used was that the distance between i-1 and i+1 oxygen atoms must be within 4.5Å, both C-O vectors must look towards each other and the residues must not fall into the helical region.

Niche-3r:
A motif of 3 consecutive residues with dihedral angles as follows:
residue (i-1): no limitations
residue (i): -120˚ < φ < -60˚ -50˚ < ψ < 30˚
residue (i+1): -100˚ < φ < -50˚ 110˚ < ψ < 170˚
Comments: This version of the database contains motifs with the amended definition. The criteria used was that the distance between i-1 and i+1 oxygen atoms must be within 4.5Å, both C-O vectors must look towards each other and the residues must not fall into the helical region.

Niche-4l:
A motif of 4 consecutive residues with dihedral angles as follows:
residue (i-1): no limitations
residue (i): -160˚ < φ < -30˚ 90˚ < ψ < 180˚
residue (i+1): 50˚ < φ < 140˚ -40˚ < ψ < 50˚
residue (i+2): -160˚ < φ < -30˚ 90˚ < ψ < 180˚
Comments: According to these definitions, virtually all niche4s incorporate a niche3. This version of the database contains motifs with the amended definition. The criteria used was that the distance between i-1 and i+1 oxygen atoms must be within 4.5Å, both C-O vectors must look towards each other and the residues must not fall into the helical region.

Niche-4r:
A motif of 4 consecutive residues with dihedral angles as follows:
residue (i-1): no limitations
residue (i): -160˚ < φ < -30˚ 90˚ < ψ < 180˚
residue (i+1): -140˚ < φ < -50˚ -50˚ < ψ < 40˚
residue (i+2): -160˚ < φ < -30˚ 90˚ < ψ < 180˚
Comments: According to these definitions, virtually all niche4s incorporate a niche3. This version of the database contains motifs with the amended definition. The criteria used was that

the distance between i-1 and i+1 oxygen atoms must be within 4.5Å, both C-O vectors must look towards each other and the residues must not fall into the helical region.

Schellman-loop-6:

A motif of six consecutive residues (common type) or seven consecutive residues (wide type) that contains two H-bonds in which:

-the main-chain CO of residue (i) is H-bonded to the main-chain NH of residue (i+5) (common type) or residue (i+6) (wide type)

-the main-chain CO of residue (i+1) is H-bonded to the main-chain NH of residue (i+4) (common type) or residue (i+5) (wide type)

Comments: The common Schellman Loop contains an RL Nest at positions (i+3) and (i+4), whereas the wide type contains one at positions (i+4) and (i+5). Schellman Loops are a common feature of the C-termini of α-helices.

ST-motif:

A motif of 5 consecutive residues and two H-bonds in which:

-residue (i) is Serine (S) or Threonine (T)

-side-chain O of residue (i) is H-bonded to the main-chain NH of residue (i+2) or (i+3)

-main-chain CO of residue (i) is H-bonded to the main-chain NH of residue (i+3) or (i+4)

Comments: the ST-motif is analogous to the Asx-motif, with the oxygen of the hydroxyl side-chain replacing that of the acid or amide.

ST-staple:

A motif of four or five consecutive residues and one H-bond in which:

-residue (i) is Serine (S) or Threonine (T)

-the side-chain OH of residue (i) is H-bonded to the main-chain CO of residue (i-3) or (i-4)

-φ angles of residues (i-1), (i-2) and (i-3) are negative

Comments: This is a common feature in the middle part of α-helices. The Ser or Thr side-chain can be regarded as stapling together two adjacent turns of the α-helices.

ST-turn:

A motif of three consecutive residues and one H-bond in which:

-residue (i) is Serine (S) or Threonine (T)

-the side-chain O of residue (i) is H-bonded to the main-chain NH of residue (i+2)

Sub-categories:

Type I:

residue (i+1): $-140^\circ < \chi1 - 120^\circ < -20^\circ$ $-90^\circ < \psi + 120^\circ < 40^\circ$

residue (i+2): $-140^\circ < \varphi < -20^\circ$ $-90^\circ < \psi < 40^\circ$

Type I':

Left-handed form of Type I

Type II:

residue (i+1): $-140^\circ < \chi1 - 120^\circ < -20^\circ$ $80^\circ < \psi + 120^\circ < 180^\circ$

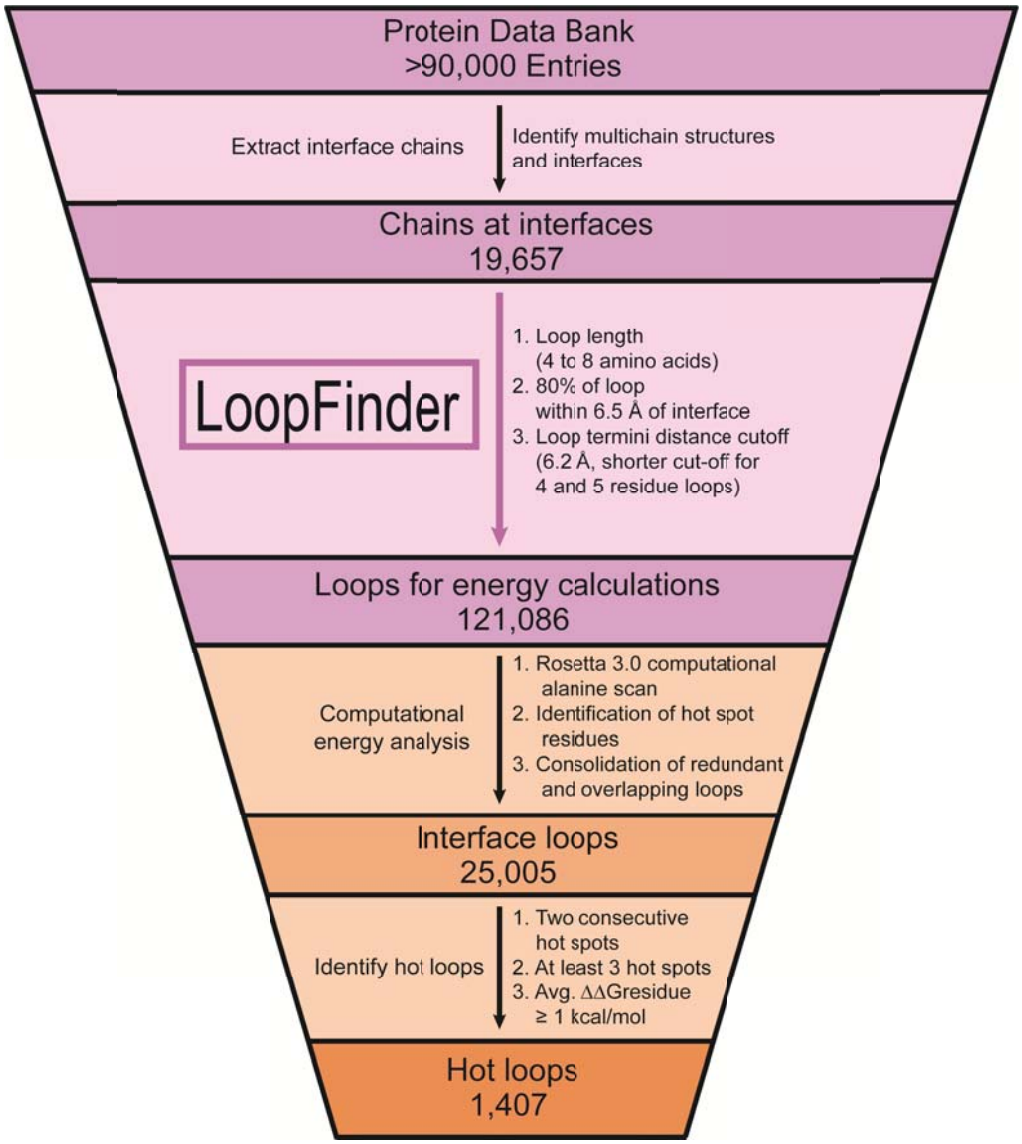
residue (i+2): $20^\circ < \varphi < 140^\circ$ $-40^\circ < \psi < 90^\circ$

Type II':

Left-handed form of Type II

Comments: The ST-turn is analogous to the Asx-turn, with the oxygen of the hydroxyl side-chain replacing that of the acid or amide.

SUPPLEMENTARY RESULTS



Supplementary Figure 1. Workflow for using LoopFinder to identify hot loops at PPI interfaces. LoopFinder was used to analyze all heterogeneous protein-protein complexes in the PDB, identifying just over 121,000 loops at protein-protein interfaces with the given structural parameters. These loop regions were subjected to computational alanine scan mutagenesis using Rosetta to calculate the relative contributions of each residue to the PPI.[14,15] The resulting data were consolidated into a set of 25,005 interface loops, representing a non-redundant set of all loops that mediate PPIs, each with complete computational alanine scan data. The interface loop set was further sorted by key criteria in order to identify those loops containing large proportions of the overall binding energy of the PPI. These comprise the set of 1,407 hot loops.

Supplementary Table 1. The full analysis of hot spots from the total interface loop data set of 25,005 structures. The avg. ΔΔG value for each amino acid, in alphabetical order, is shown in the first column. The total number of each amino acid present in the interface loop data set is shown in the second column followed directly by the percent abundance of each amino acid: $(\frac{\text{number of each amino acid (column 2)}}{\text{total number of amino acids in all loops}} \times 100)$. The number that each amino acid appears as a hot spot, and the percent of each amin o acid that occurs as a hot spot $(\frac{\text{Number of each amino acid that are hot spots}}{\text{number of each amino acid (column 2)}} \times 100)$. In order to identify which amino acids are favored as hot spots, the fold enrichment was calculated:

$$(\frac{\text{number of each amino acid that are hot spots/total number of hot spots for all residues}}{\text{number of each amino acid (column 2)/total number of all amino acids in all loops}} \times 100).$$

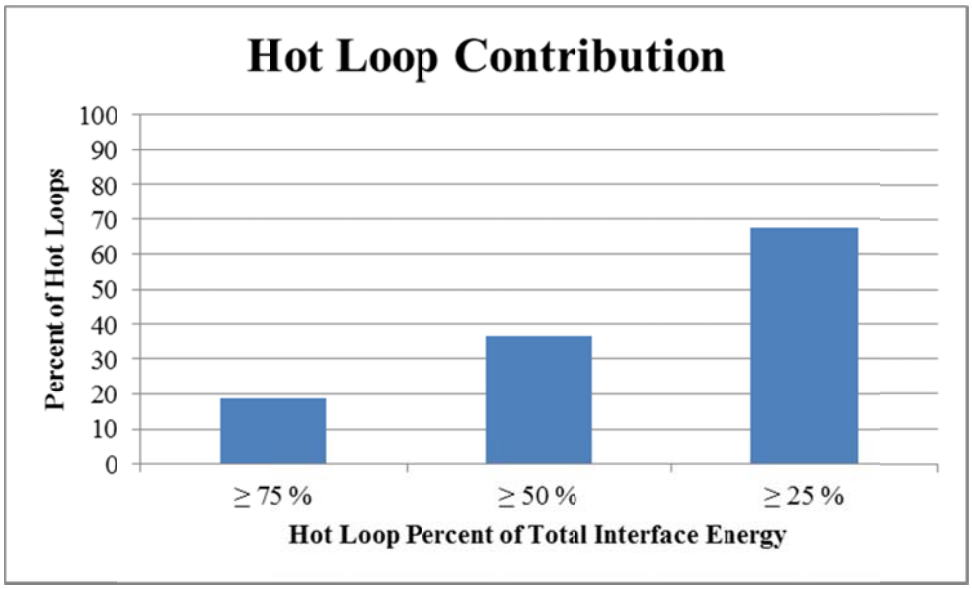
Amino Acid Analysis of Hot Spots in All Loops

Residue	Average $\Delta\Delta G_{residue}$	Total # of amino acids	Percent Abundance	Contributes ≥ 1 kcal/mol		Fold Enrichment in Hot Spots	Contributes ≥ 2 kcal/mol		Fold Enrichment in Hot Spots
				(total)	(%)		(total)	(%)	
Ala	0.01	12341	7.95	98	0.79	0.13	80	0.65	0.36
Arg	0.07	8184	5.27	696	8.50	1.36	184	2.25	1.26
Asn	0.08	7560	4.87	352	4.66	0.74	98	1.30	0.72
Asp	0.21	9169	5.91	881	9.61	1.54	235	2.56	1.43
Cys	-0.03	2291	1.48	27	1.18	0.19	13	0.57	0.32
Gln	-0.02	5378	3.47	165	3.07	0.49	39	0.73	0.41
Glu	0.15	9199	5.93	730	7.94	1.27	217	2.36	1.32
Gly	0.00	19786	12.75	0	0.00	0.00	0	0.00	0.00
His	0.27	3924	2.53	619	15.77	2.52	163	4.15	2.32
Ile	0.14	6502	4.19	592	9.10	1.46	116	1.78	1.00
Leu	0.15	12020	7.75	1048	8.72	1.39	148	1.23	0.69
Lys	0.03	7403	4.77	309	4.17	0.67	64	0.86	0.48
Met	0.01	2955	1.90	117	3.96	0.63	31	1.05	0.59
Phe	0.28	6275	4.04	1201	19.14	3.06	341	5.43	3.04
Pro	0.09	7245	4.67	405	5.59	0.89	230	3.17	1.77
Ser	0.09	10503	6.77	398	3.79	0.61	117	1.11	0.62
Thr	0.09	8639	5.57	372	4.31	0.69	123	1.42	0.80
Trp	0.30	1980	1.28	423	21.36	3.42	221	11.16	6.24
Tyr	0.17	5714	3.68	904	15.82	2.53	223	3.90	2.18
Val	0.12	8092	5.22	367	4.54	0.73	134	1.66	0.93

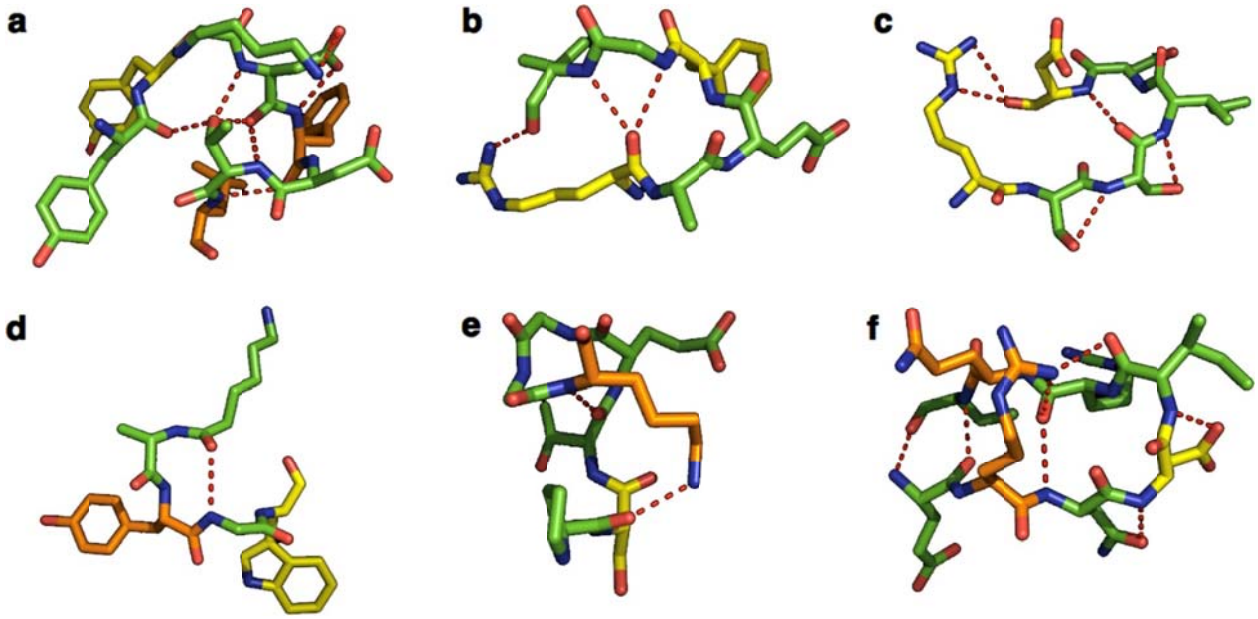
Supplementary Table 2. Identical analysis of hot spots conducted for the data set of hot loops only. Upon comparison to Supplementary Table 1, conclusions can be made as to which amino acid residues are most likely to be contained within a hot loop.

Amino Acid Analysis of Hot Spots in All Loops

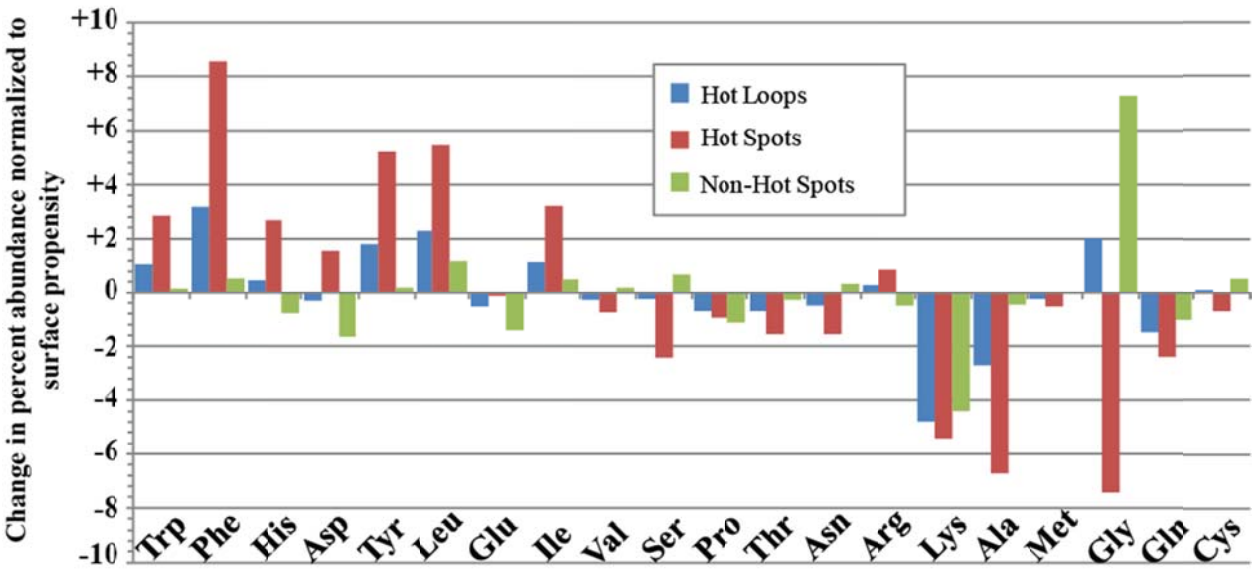
Residue	Average $\Delta\Delta G_{residue}$	Total # of amino acids	Percent Abundance	Contributes ≥ 1 kcal/mol		Fold Enrichment in Hot Spots	Contributes ≥ 2 kcal/mol		Fold Enrichment in Hot Spots
				(total)	(%)		(total)	(%)	
Ala	0.30	468	5.39	36	7.69	0.22	32	6.84	0.54
Arg	0.83	545	6.28	234	42.94	1.23	67	12.29	0.97
Asn	0.56	418	4.82	99	23.68	0.68	40	9.57	0.76
Asp	0.97	597	6.88	284	47.57	1.37	80	13.40	1.06
Cys	0.37	97	1.12	11	11.34	0.33	7	7.22	0.6
Gln	0.53	246	2.84	61	24.80	0.71	21	8.54	0.68
Glu	0.86	613	7.06	262	42.74	1.23	87	14.19	1.12
Gly	0.00	860	9.91	0	0.00	0.00	0	0.00	0.00
His	1.07	319	3.68	181	56.74	1.63	60	18.81	1.49
Ile	0.95	368	4.24	165	44.84	1.29	49	13.32	1.05
Leu	0.83	659	7.59	294	44.61	1.28	65	9.86	0.78
Lys	0.50	348	4.01	99	28.45	0.82	27	7.76	0.61
Met	0.55	119	1.37	29	24.37	0.70	13	10.92	0.86
Phe	1.29	484	5.58	318	65.70	1.89	113	23.35	1.85
Pro	1.27	459	5.29	183	39.87	1.15	125	27.23	2.16
Ser	0.59	594	6.85	155	26.09	0.75	54	9.09	0.72
Thr	0.75	470	5.42	140	29.79	0.86	52	11.06	0.88
Trp	1.55	186	2.14	116	62.37	1.79	78	41.94	3.32
Tyr	1.11	408	4.70	233	57.11	1.64	70	17.16	1.36
Val	0.74	419	4.83	120	28.64	0.82	56	13.37	1.06



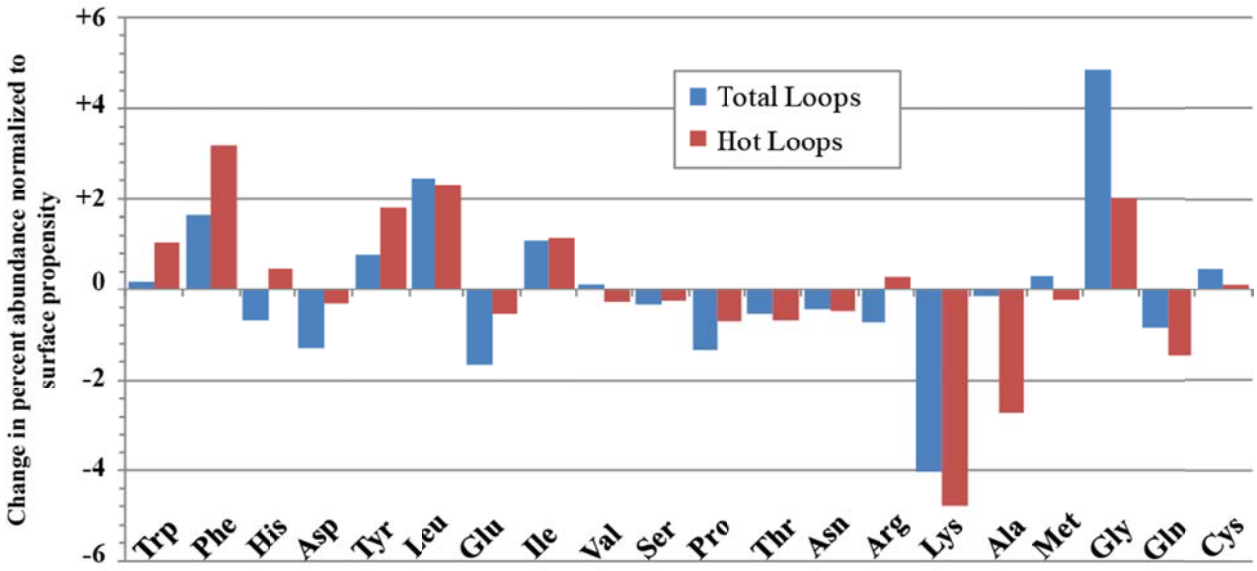
Supplementary Figure 2. Relative contributions of hot loops compared to total interface energy. The contribution of hot loops to the interface energy calculated for the entire chain was analyzed. The resulting contributions show that, on the whole, the hot loops identified by LoopFinder comprise a significant percentage of the total interface energy.



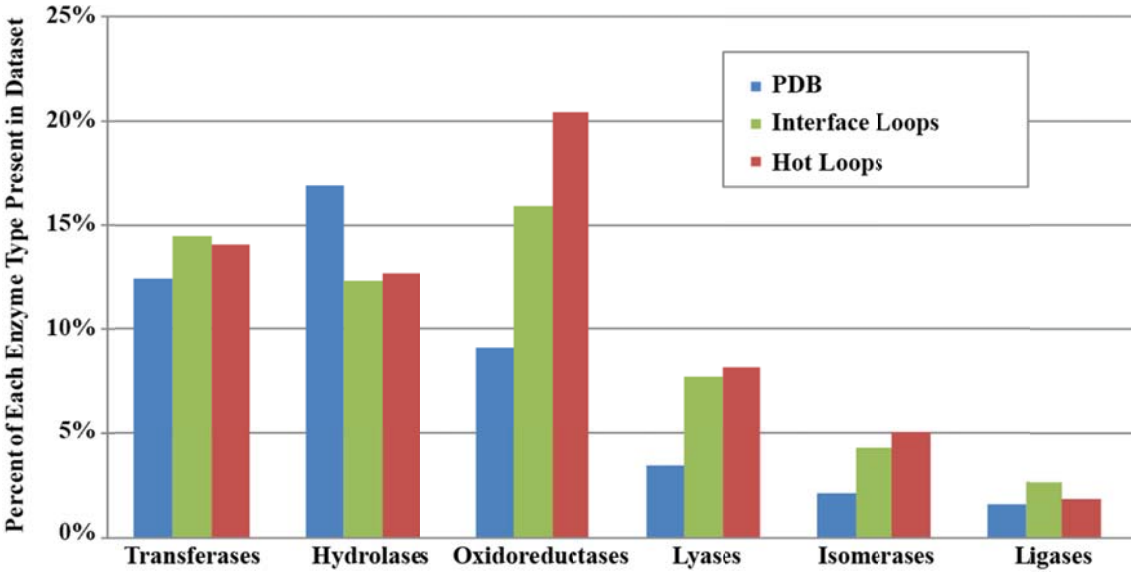
Supplementary Figure 3. Selected hot loops containing unusual turn motifs. a) An interface loop from *Borrelia burgdorferi* BbCRASP-1 with three non-consecutive hydrophobic hot spots constrained by an (*i*,*i*+1) salt bridge and a central threonine with an (*i*,*i*-3;*i*,*i*-6) ST-staple (1W33). b) An interface loop from the *Mycobacterium tuberculosis* toxin-antitoxin complex RelBE2 with a bidentate (*i*,*i*+4;*i*,*i*+5) main-chain hydrogen bond at the C-terminal end of an alpha helix (3G5O). c) An interface loop that contains a β-turn with non-standard backbone torsional angles and two (*i*,*i*+1) side-chain-to-backbone hydrogen bonds (3OA8); d) An interface loop from *Plasmodium falciparium* thioredoxin reductase that contains a β-turn with non-standard backbone torsional angles at the C-terminal end of an α-helix (4B1B). e) An interface loop from potassium ion channel Trek2 with a side-chain-to backbone hydrogen bond from lysine at the N-terminal end of an α-helix (4BW5). f) An interface loop from R-spondin-1 with a β-hairpin-like structure containing an (*i*,*i*+4) hydrogen bond and side-chain-to-backbone hydrogen bonds among Arg (*i*,*i*+3;*i*,*i*+4), Asn (*i*,*i*+1), and Asp (*i*,*i*+1), with both charged residues also being hot spots (4KNG). All structures, rendered in Pymol,[46] show the hot loop in green, and hot spots in orange ($\Delta\Delta G_{residue} \geq 1$ kcal/mol) or yellow ($\Delta\Delta G_{residue} \geq 2$ kcal/mol).



Supplementary Figure 4. Amino acid abundance relative to surface propensity for the hot loop set. The percent abundance of each amino acid in the hot loop data (blue), hot spot residues (red) and non-hot spot residues (green) was compared directly to the natural surface percent abundance of each amino acid as identified previously by Janin *et al.* in order to analyze which amino acids are over-represented at surface hot loops that take part in PPIs.[4] This figure is similar to Figure 3 in the main text, but uses only hot loop data instead of the full interface loop set.



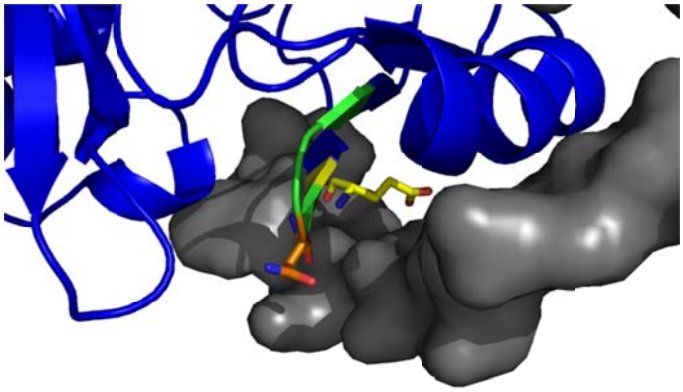
Supplementary Figure 5. Comparison of amino acid abundance between hot loops and all interface loops. A comparison of the amino acid percent abundance, normalized to surface propensity as measured by Janin *et al.*, for total interface loops (blue) and hot loops (red). These data are re-plotted here in one plot to show similarities in amino acid distribution at hot loops compared to the total interface loop set.[4]



Supplementary Figure 6. Loop-mediated PPIs by functional class. Enzymes make up the largest class of proteins with loop-mediated PPIs. Percentages of each enzyme type are shown for the total PDB (blue), the 25,005 interface loops identified by LoopFinder (red), and the 1,407 hot loops (green). The 25,005 interface loops identified by LoopFinder covers 499 different functional classes of proteins from all kingdoms of life, as well as viral proteins. The 1,407 hot loops, by contrast, cover only 132 functions of proteins (Supplementary Data Set 1). The functional classes with the highest number of hot loops are the common classes of enzymes: hydrolases, isomerases, lyases, oxidoreductases, transferases, and ligases. This is not surprising given their prominence in the PDB as a whole. When representation of each enzyme type within the interface loop set and the hot loop set are compared to representation in the input PDB set (as shown here), it is clear that oxidoreductases are over-represented in the interface loop set, and are even more prominent in the hot loop set. Lyases and isomerases show a similar but less prominent trend, and hydrolases show the opposite trend. We conclude that oxidoreductases form loop-mediated PPIs more often than other types of enzymes. Further analysis will illuminate the reason for this over-representation.

```
HUMAN        551   TSFFPEPDDVESLMITPFLPVVAFGR-PLPKLTPQNFELPWLDERSR---
MOUSE        554   TSFFPEPDDVESLLITPFLPVVAFGR-PLPKLAPQNFELPWLDERSR---
CHIMP        289   TSFFPEPDDVESLMITPFLPVVAFGR-PLPKLTPQNFELPWLDERSR---
ZEBRAFISH    427   LSFHAEPEDVEYIMVTPFLPVVAFGS-PLPNLKQQDFDLPWLDERSR---
SALMON       473   SSFYPDTEDVESIIITPFLPVVAFGR-PLPKLTQQNFDLPWQ--RSR---
TETRAODON    364   SSFYPDTEDVETIVITPFLPVVAFGR-PLPKLSQENFELPWLDDRSR---
DROSOPHILA   943   STFYPLPEDIEAIQFVNEVTVQAFGENVVNMEARDDFGVPWVDAIEAPTS
```

Supplementary Figure 7. An alignment of MSL1 across species shows that the region identified as a hot loop is highly conserved across species. In the green box are the loop members that do not contribute significantly to binding. In the yellow box are the hot spot residues identified to contribute significantly to the binding energy.



Supplementary Figure 8. A second hot loop in the MOF complex. The binding interaction between MOF and MSL1 was also identified by LoopFinder (4DNC).[8] The epitope of MOF isolated by LoopFinder ranges from His183 to Glu188 (HIGNYE) with hot spots identified in yellow (Glu188, $\Delta\Delta G_{residue}$ = 3.23 kcal/mol) and orange (Asn186, $\Delta\Delta G_{residue}$ = 1.54kcal/mol). This crystal structure was made using only MSL1, a single component of the MSL complex. It is hypothesized that this same interaction would be present in the full complex structure and allow for the design of an inhibitor aimed at disrupting MOFs ability to bind its activating complex. All structures, rendered in Pymol,[9] show the hot loop in green, and hot spots in orange ($\Delta\Delta G_{residue} \geq 1$ kcal/mol) or yellow ($\Delta\Delta G_{residue} \geq 2$ kcal/mol).

Supplementary Table 3. Comparison to experimental alanine scanning for hGH-hGHbp complexes. Alanine scanning mutagenesis has been extensively done to study the binding of hGH to its hGHbp partner.[5–7] A comparison of mutagenesis results from experimental data is made to the computational data used to identify hot loop for both hGH and hGHbp in two crystal structures isolated in the LoopFinder process. Though the values from the computational alanine scan are not exactly the same as experimental values, the magnitudes are similar. Experimentally, these loops were identified as highly important the for the hGH:hGHbp interaction, confirming our approach for identifying hot loops.

hGH Loop	Experimental $\Delta\Delta G_{residue}$ (kcal/mol)	Computational		hGHbp Loop	Experimental $\Delta\Delta G_{residue}$ (kcal/mol)	Computational	
		1HWG	1AXI			1HWG	1AXI
P61	1.2	0	0	I165	2.0	0.0	4.5
S62	0.1	-0.15	0	Q166	0	0.0	0.1
N63	0.3	4.5	0.4	K167	0	0.3	0.3
R64	1.6	1.9	1.8	G168	n/a	0.0	0.0
E65	-0.5	0.22	0	W169	4.5	-1.1	2.4
E66	n/a	0	0	M170	n/a	0.0	0.0

n/a = no experimental data available

Supplementary Table 4. Checking reproducibility using the online server Robetta. 19 hot loops generated by LoopFinder, each with 3 non-consecutive hotspots, were compared with hot spots calculated independently using the Robetta computational alanine scan server (http://robetta.bakerlab.org/). 57 hot spots were identified using our score function and 54 were identified by Robetta. In total, 44 hot spots (77%) were identified by both methods. Of the hot loops identified, only one (2GE7:A58-63) would be excluded from the hot loop set using this alternate set of ΔΔG values. Underlined residues were identified by both methods; bold by LoopFinder only, italicized by Robetta only.

PDB	CHAIN	FIRST	LAST	AA1	AA2	AA3	AA4	AA5	AA6	AA7	AA8
1IAI	L	HIS91	PHE96	HIS	TYR	SER	**THR**	PRO	PHE		
1J34	A	GLU82	SER87	GLU	TRP	SER	ASP	GLY	**SER**		
1MTP	B	ARG347	LEU353	ARG	*ARG*	ARG	GLY	ALA	*ILE*	LEU	
1XIM	D	SER106	ARG112	SER	ASN	ASP	ARG	SER	VAL	ARG	
2DVT	A	LEU182	TRP187	LEU	LEU	GLY	**PRO**	*THR*	TRP		
2GE7	A	PRO58	CYS63	**PRO**	SER	SER	HIS	ALA	**CYS**		
2IUF	E	ALA33	ALA38	**ALA**	GLY	*GLN*	ARG	GLY	**ALA**		
2OIZ	D	ARG147	LEU154	ARG	PRO	GLY	TYR	GLU	PHE	PHE	LEU
2QLZ	A	LEU84	PHE89	LEU	*THR*	**PRO**	GLU	N/A	PHE		
2QNR	B	LEU257	VAL262	LEU	TYR	PRO	TRP	GLY	**VAL**		
3B7E	A	ALA138	HIS144	ALA	LEU	LEU	ASN	**ASP**	*LYS*	HIS	
3FSL	A	ARG292	SER296	ARG	ARG	ASN	TYR	SER			
3GWA	A	ASP92	SER98	ASP	TYR	*VAL*	LEU	PRO	THR	SER	
3HMU	A	THR90	THR95	THR	PHE	*PHE*	LYS	THR	THR		
3OCD	A	ARG238	ALA244	ARG	GLN	MET	ARG	MET	**PRO**	ALA	
3STH	B	MET305	LYS312	MET	LEU	ASN	ASP	THR	PHE	VAL	LYS
3VTO	C	HIS108	ARG113	HIS	HIS	GLU	GLY	HIS	ARG		
4EHI	A	HIS361	PHE366	HIS	*ILE*	ASP	GLY	GLY	PHE		
4HST	B	LEU37	ASP43	LEU	ALA	**CYS**	ASP	**ARG**	*PHE*	ASP	
Loopfinder Alanine Scan											
PDB	**CHAIN**	**FIRST**	**LAST**	**ΔΔG1**	**ΔΔG2**	**ΔΔG3**	**ΔΔG4**	**ΔΔG5**	**ΔΔG6**	**ΔΔG7**	**ΔΔG8**
1IAI	L	HIS91	PHE96	1.88	0.00	0.00	4.50	0.04	1.46		
1J34	A	GLU82	SER87	0.00	3.24	0.96	1.80	0.00	1.23		
1MTP	B	ARG347	LEU353	2.67	-0.51	2.88	0.00	0.01	0.45	1.54	
1XIM	D	SER106	ARG112	0.00	2.52	0.32	2.02	0.54	0.13	1.90	
2DVT	A	LEU182	TRP187	1.18	0.00	0.00	2.81	0.90	2.93		
2GE7	A	PRO58	CYS63	4.15	0.27	-0.07	1.77	0.00	1.48		
2IUF	E	ALA33	ALA38	1.16	0.00	0.50	1.77	0.00	4.50		
2OIZ	D	ARG147	LEU154	3.12	-0.05	0.00	1.61	-0.87	4.50	0.00	0.00
2QLZ	A	LEU84	PHE89	1.13	0.95	4.50	0.81	0.00	1.10		
2QNR	B	LEU257	VAL262	-0.01	1.30	0.54	2.19	0.00	4.50		
3B7E	A	ALA138	HIS144	0.00	1.77	0.00	0.00	1.58	0.33	3.75	
3FSL	A	ARG292	SER296	4.50	0.59	1.59	0.30	1.16			
3GWA	A	ASP92	SER98	0.72	1.18	-0.90	1.61	0.89	4.50	0.00	
3HMU	A	THR90	THR95	-0.15	3.11	0.00	1.67	0.38	1.28		
3OCD	A	ARG238	ALA244	1.64	0.00	0.00	1.93	0.43	3.83	0.00	
3STH	B	MET305	LYS312	0.00	1.56	0.00	0.00	0.00	2.13	0.00	4.50
3VTO	C	HIS108	ARG113	3.59	0.40	1.51	0.00	2.35	0.00		
4EHI	A	HIS361	PHE366	1.44	0.79	1.44	0.00	0.00	2.45		
4HST	B	LEU37	ASP43	1.33	0.00	4.50	0.27	1.04	0.00	0.00	

Robetta Alanine Scan

PDB	CHAIN	FIRST	LAST	ΔΔG1	ΔΔG2	ΔΔG3	ΔΔG4	ΔΔG5	ΔΔG6	ΔΔG7	ΔΔG8
1IAI	L	HIS91	PHE96	2.32	0.00	0.00	0.29	0.00	1.51		
1J34	A	GLU82	SER87	0.00	5.39	0.53	1.43	0.00	0.50		
1MTP	B	ARG347	LEU353	4.50	4.50	4.02	0.00	0.00	1.75	2.09	
1XIM	D	SER106	ARG112	-0.02	2.61	0.29	4.18	0.55	0.34	4.50	
2DVT	A	LEU182	TRP187	2.07	0.00	0.00	0.00	1.56	4.50		
2GE7	A	PRO58	CYS63	0.00	0.88	0.12	2.47	0.00	-0.05		
2IUF	E	ALA33	ALA38	0.00	0.00	2.21	3.43	0.00	0.00		
2OIZ	D	ARG147	LEU154	2.62	0.00	0.00	4.50	-0.26	2.46	0.00	0.00
2QLZ	A	LEU84	PHE89	2.43	1.37	0.00	0.56	0.00	2.39		
2QNR	B	LEU257	VAL262	0.00	4.37	0.00	4.07	0.00	0.00		
3B7E	A	ALA138	HIS144	0.00	1.60	0.00	0.00	0.75	1.77	3.19	
3FSL	A	ARG292	SER296	1.91	0.40	3.78	0.50	1.28			
3GWA	A	ASP92	SER98	0.90	2.64	1.27	3.31	0.00	3.49	0.00	
3HMU	A	THR90	THR95	-0.23	2.23	2.76	1.16	0.48	3.01		
3OCD	A	ARG238	ALA244	1.34	0.08	0.00	3.60	0.74	0.00	0.00	
3STH	B	MET305	LYS312	0.00	2.66	0.58	0.00	0.00	3.35	0.00	4.50
3VTO	C	HIS108	ARG113	1.79	0.19	2.04	0.00	1.89	0.00		
4EHI	A	HIS361	PHE366	2.22	1.73	3.73	0.00	0.00	3.23		
4HST	B	LEU37	ASP43	2.63	0.00	-0.24	0.16	0.21	1.02	0.00	

Supplementary Table 5. Comparison of LoopFinder results to HippDB results. These show limited overlap between the two databases. Rows highlighted in light gray contain overlapping sequences, with boxed sequences highlighting specific overlapping sequences.

A. PDB IDs for all of the protein complexes identified to have hot loops by LoopFinder and also have helical interface regions as identified by HippDB

B. The interface chains involved in the PPI as identified by LoopFinder, the first letter listed is the chain that contains the interface loop

C. Sequence of hot loop as identified by LoopFinder

D. The interface chains involved in the PPI as identified by HippDB, the first letter listed is the chain that contains the interface helix (reference 10)

E. Sequence of interface helix as identified by HippDB (reference 10)

F. Identical to D, provided for interfaces with multiple interface helices (reference 10)

G. Identical to E, provided for interfaces with multiple interface helices (reference 10)

A. PDB ID	B. Interface Chains	C. LoopFinder Hot Loops	D. Interface Chains	E. HippDB Interface Helices	F. Interface Chains	G. HippDB Interface Helices
1BOU	DC	DEGWG	DA	DLAWHIAQSLIL	DC	IQYLRE
1CPC	AB	ADSQGRFL	BA	RMAACLRDMEIILRY VTYAIFA	LK	RMAACLRDMEIILRYVTYAI FA
1DD4	BA	LTVSEL	CA	TIDEIIEAI	DB	TIDEIIEAIE
			DB	ELAELVKKLEDK		
1DOA	AB	THHCP	BA	SLRKYKEALL		
1XDT	RT	HGERC	BA	LGRLLVV	D C	LGRLLVV
1EEX	ML	DYPLANK	BA	ARPKYQAKSAILHIKET		
1EEX	AL	QRDLKV	BA	ARPKYQAKSAILHIKET		
1EEX	LA	AHGSKD	BA	ARPKYQAKSAILHIKET		
1EFR	AB	HLGESTV	GE	LTLTFNRTRQAVITKELIEIISGA		
1XEY	BA	TIGSSEA	BA	IILRYISYALLA	LK	EIILRYISYALLA
1XFS	AB	LVKNYRP	BA	LGRLLVV	BA	FKLLGNVLVVVLARNF
			DC	GRLLVV		
1GVN	BA	YADDP	DC	NRLNDNLEE	DC	RYETMYAD
1H0H	D	LEAEP	BA	GCQVACKQWH		
1H2K	SA	QGEEL	SA	EELLRAL		

1IWP	AG	FTDGDDT	BA	RPKFMAKAALFHIKE TK	ML	LEKVL
			ML	ERILAIYNAL	ML	FVRESAEVYQQ
1IWP	AL	SHSDIRR	BA	RPKFMAKAALFHIKE TK	ML	LEKVL
			ML	ERILAIYNAL	ML	FVRESAEVYQQ
1IWP	AL	QRDLMV	BA	RPKFMAKAALFHIKE TK	ML	LEKVL
			ML	ERILAIYNAL	ML	FVRESAEVYQQ
1JNR	BA	GYVDYS	CA	VRLQKIMDEY		
1KF6	PO	SAIIA	DC	APVMILLVG		
1KF6	DC	ILLVG	DC	APVMILLVG		
1LIA	LK	LDAFSR	LK	IILRYVSYALLA		
1LQB	CB	YTLKER	CB	LKERCLQVVRSL		
1LTS	ED	MAGKRE	CA	EETQNLSTIYLREYQS KVKRQI	ED	KDTLRITYLT
			FE	KDTLRITYLT	GF	KDTLRITYLT
			HG	KDTLRITYLT		
1M34	JI	RDGFE	DA	KERGRLVDMMTDSHTWL		
1M34	FE	VVCGGF	DA	KERGRLVDMMTDSHTWL		
1MXH	CA	VPLGQ	FB	LRKQR		
1MHY	BD	RWHHPY	GD	GLRKER		
1MHY	DG	WLIEP	GD	GLRKER		
1MTY	BD	KFHGGRPS	HE	LRKQR		
1NVM	AC	VDRETL	CA	YTLMDAADD	GE	YTLMDAADD
1P84	CD	FVFYS	BA	YTKL	DC	RKRLGLKTVIILSSLYLLSIW V...
			GC	ARAYRIIRAHQTELT	HC	VLIPAGIYWYWWKNGNEY NEFL
			FD	EEFFHLQHYLDTAT	GD	AYRIIRAHQTELT
			ID	DTAITSWYENH	ID	WKDVK
1POI	BA	PRSVGD	DC	LRFM		
1PYA	EF	ETKNAYI	CD	DVLDGIVSYDRAET	ED	DVLDGIVSYDRAET
1Q7L	BA	LLHDHDE	D C	DNRYIRA	DC	EAVFLRGVDIYTRLLPALA
1QH8	BA	NRHFKE	DA	ERGRLVDMMLDSHTWL		
1QHH	AB	AKNELL	BA	EKVVSDVYQEYQQRL L	BA	DLIMTTIQLFDR
			CB	AGALAAFRSQLEQWT QL	CB	AQSRLENLDEFLSVTKH
			CB	LIAFLT		
1QHH	BC	YDRKEI	BA	EKVVSDVYQEYQQRL L	BA	DLIMTTIQLFDR
			CB	AGALAAFRSQLEQWT QL	CB	AQSRLENLDEFLSVTKH

			CB	LIAFLT		
1QHH	BC	VIANP	BA	EKVVSDVYQEYQQRLL	BA	DLIMTTIQLFDR
			CB	AGALAAFRSQLEQWTQL	CB	AQSRLENLDEFLSVTKH
			CB	LIAFLT		
1RM6	CA	TQCGFCT	BA	WWRSG		
1RP3	AB	QLIFY	DC	TLSKIAQELS	DC	DEKVVKGLIEFF
			FE	DKVTLSKIAQEL	FE	DEKVVKGLIEFF
			HG	LSKIAQEL	HG	EKKVKELKEKIE
			HG	DEKVVKGLIEFF		
1S5D	ED	LAGKRE	ED	KDTLRIAYLT	HD	IERMKDTLRIAYLT
			FE	DTLRIAYL		
1SDK	BA	AHHFGKEF	D C	LGRLLVV		
1TQY	BA	LWSEG	DC	FTHREFRKLWS	FE	FTHREFRKLWS
1TWF	AH	LTLRDT	KC	HTLGNLIRAE	KC	ALKNACNSIINKLGALKTNFETE
1TZY	FE	KQVHP	BA	IYVYKVLKQV	BA	KAMGIMNSFVNDIFERIAGEAS...
			BA	ELAKHAVSEGTKAVTKYTS	DB	VTYTEH
			DC	PAIRRLARR	DC	EETRGVLKVFLENVIRDAVTYTE
			GC	DIQLARRIR	HF	VTYTEH
			HG	PAIRRLARR	HG	ETRGVLKVFLENVIRDAVTYTE
1ULI	CE	QCRHRG	DC	QMMRGRLRKI	EC	DGENWVEIQQV
			FD	QHEIEQFYYWEAKLLN		
1ULI	CE	EEQAF	DC	QMMRGRLRKI	EC	DGENWVEIQQV
			FD	QHEIEQFYYWEAKLLN		
1UMD	BC	GGHHH	CA	GDWYAGINFAAV	CB	LRQEALL
			DC	LRYR		
1UMD	AC	AHAFGI	CA	GDWYAGINFAAV	CB	LRQEALL
			DC	LRYR		
1YE9	AE	YTEEGI	DA	LREKITHFD	HA	PLLQGRLFSYTD
			FC	PLLQGRLFSYTDTQISR	ED	PLLQGRLFSYTD
			GE	FFAE	HF	FFAE
			LI	REKITHFD	KJ	REKITHFD
			OJ	PLLQGRLFSYTD	NK	ELWEAIE
1YFN	EA	QKMPFW	BA	RPYLLRAFYEWL	DC	PYLLRAFYEWLLD
1YWH	FE	YLWSS	DC	YLWS	HG	YLWS
			NM	YLWS	PO	YLWS
1Z3E	BA	LKRAGI	BA	VRSYNCLKR		
2A6H	ED	VDSKYRL	BA	TLGNPLRRILLS	LK	TLGNPLRRILLS

2AFH	AF	DIVFGG	DA	KERGRLVDMMTDSHTWL		
2AFH	CD	FQKMGI	DA	KERGRLVDMMTDSHTWL		
2B1X	AE	YISEDQ	BA	DSMEMRVLRL	FB	WLYMEAELLD
2B1X	AE	ACRHRG	BA	DSMEMRVLRL	FB	WLYMEAELLD
2B7Y	CD	WDSETGN	D C	LKDV		
2BR2	EB	DYAKKADG	DC	RREIELSKVIREALE	LK	RREIELSKVIREALE
			PO	RREIELSKVIREALE	WX	RREIELSKVIREALE
2BW3	BA	VVRDCR	BA	DCKKEAIEKCAQWVVRD		
2BWE	CD	LRRSGG	BA	EHQLRQLND	CB	EHQLRQLND
			DC	EHQLRQLND	ED	EHQLRQLND
			FE	EHQLRQLN	HG	EHQLRQLN
			IH	EHQLRQLND	KJ	EHQLRQLND
			ML	EHQLRQLND	ON	EHQLRQLN
			QP	EHQLRQLND	RQ	EHQLRQLND
2FM8	AC	DLFALPS	BA	YEILMTI	CB	PALIKQASLDALF
2G38	AB	AADLVS	BA	AAARAWRSLDVEMT AVQRSFNRTL	DC	AAARAWRSLDVEMTAVQR SFNRTL
2GL9	CD	STWNFG	DB	HLVVELMN		
2H5K	BA	RDGAGKY	BA	NELVDY		
2HZS	AB	VFNGSSTG	IB	VDDVLKFTFT	JD	VDDVLKFTFT
			KF	VDDVLKFTFT	LH	VDDVLKFTFT
2J3T	DC	ETDTFK	DC	LASMFHSLFAIGSQ		
2JDI	HG	ASPTQV	IH	YIRYSQICAKAVR		
2P5T	AB	LNPVED	CA	ERYSGYLDGIERMLEI SEKR	DC	HALARNLRSLT
			HD	SYLSTLIRYE		
2XPP	BA	FEIFG	BA	GDRDSLFFEIF		
2QRD	GA	DGETGS	DC	AFLT		
2RF4	AB	IHDAF	BA	LSSSISQLKRIQRDF	DC	LSSSISQLKRIQRDF
2V7Q	FC	HLGESTV	JD	FGKREQAEEERYFRA RAKEQLA...	IH	YIRYSQICAKAVR
2V7Q	AD	SLLLRR	JD	FGKREQAEEERYFRA RAKEQLA...	IH	YIRYSQICAKAVR
2VX8	AB	KSLLG	BA	KFLM		
2WG3	AC	ESRNHV	DB	LDDMEE		
2WNR	BE	FSVEER	FE	SVEISKITAEAL		
2YEV	AB	LSMTPLD	BA	RLEVVWTLIPLAIVFV LFGLTA...	CA	GAALVTLFFYLIL
			CA	DLRFVLFMLLLILLAA GTVALM...	ED	RLEVVWTLIPLAIVFVLFGL TA...
2YFI	EA	CRHRGMRI	FD	QNEIEQFYYREAQLLD	HG	ETMYGRIRKV
			LH	QNEIEQFYYREAQLLD		

2YIU	FA	CTHLG	BA	MDRKQVGFVSVIFLIV LAALLY...	DA	WLHRR
			DA	EVTWIV	ED	MDRKQVGFVSVIFLIVLAAL LY...
2Z5C	CA	NLYYD	CB	AVTHNLYY	FE	AVTHNLYY
2ZC3	AB	YDRNGV	CB	PGDLAEELRR	FE	PGDLAEELRR
3A1G	AB	EKFFP	DC	ERIKELRNL		
3AJV	CB	VDRTGL	CB	WAAAVEVIAG	DB	EIVRAGRL
3AYX	AB	KNPHP	DB	FDEAISE	DC	YMAKLAEQA
			IG	FDEAIS	JH	FDEAISE
			JI	VYMAKLAEQ		
3CF4	AG	AETWQEA	GA	KFYINQVLSAAKNF		
3CIP	AG	ASLSTF	GA	QDESGAAAIFTVQLDDY		
3CR3	CD	SHSPEIA	DC	EIASGLKKLIR		
3DD7	AB	ISRYG	BA	EFASLFDT	DC	AEFASLFDT
3DWL	DA	TDFDGVTF	FD	NGRARLVAETYLSC	KI	ILVRKFMQFL
			KI	IEFMEEVDAEISEMK	KI	NGRARLVAETYLS
3EXE	BA	YYMSGG	BA	TYYM	CA	GQIFEAYNMAAL
			DB	VGAEICARIME	DC	TYYM
			GE	GQIFEAYNMAALW	HF	VGAEICARIME
3EUH	AB	DYYIR	BA	PELVAWARK	BA	RLSFLLAVATLNG
			BA	LGIGITDYYI	BA	EGGDEFHWHRNVYAPLKY
3FXD	BA	YSEII	BA	NTDAVEVLTELNTKV ERA	DC	DNTDAVEVLTELNTKVERA A
3G5O	AB	RAEFGV	BA	LAAVVEFA	DA	WESLQETLYWL
			DC	RESIAEADADIAS	DC	EIRAEF
3H0L	BC	HEGDKT	CA	FQKQLSDILDF	FD	FQKQLSDILDF
			IG	FQKQLSDILDF	LJ	FQKQLSDILDF
			OM	FQKQLSDILDF	RP	FQKQLSDILDF
			US	FQKQLSDILDF	V	FQKQLSDILDF
3HVQ	AC	RIYGFY	CA	ASAEYELE		
3IAM	EG	FREGRY	43	MEAVIYHFKH	DC	MEAVIYHFKH
3K6G	AD	FLKNSG	DA	TLKAAFKTLS	DA	AFAKLDQ
			EB	TLKAAFKTL	EB	AFAKLDQ
			FC	TLKAAFKTL	FC	FAKLDQ
3MM5	ED	QGWIHC	EA	WERFFE		
3MM5	VA	EPPRW	EA	WERFFE		
3O4X	EH	GDYFVF	DA	FFDLKGRLDDIRME	DA	EVFQIILNTV
			DA	EPHFLSILQHL	DA	ARPQYYKLIEECVSQIV
			HE	FWTK		

3OQY	Bb	ERQHM	Aa	AAKERQH	bB	AAKERQH
3P8C	BA	LGPYG	EA	ITSSIKKIADFLNSFDM SCRSR...	ED	SSIKKIADFLNSFDMSCRSRL A...
3P8C	BA	SYHIP	EA	ITSSIKKIADFLNSFDM SCRSR...	ED	SSIKKIADFLNSFDMSCRSRL A...
3RRL	DC	INAGKET	D C	DLVHG		
3RRR	AB	FYQSTCS	BA	IVNKQSCSISNIETVIEF QQK	FB	LHLEGEVNKIKSALLSTNKA VV...
			FD	LHLEGEVNKIKSALLS TNKAVV...	FD	SQVNEKINQSLAFIRKSDEL
			HG	VNKQSCSISNIETVIEF QQK	LH	LEGEVNKIKSALLSTNKAVV SL...
			NH	VSKVLHLEGEV	NH	IKSALLSTNKAVVSLSNGVS VL...
3SDE	AB	CGDGAF	BA	TLAEIAKVE	BA	RWKALIEMEKQQQDQVDR NIKE...
3SQG	AD	GHYGREP	FE	SMDVTAQIHWKRSVGGF		
3U52	BD	YLTRD	DC	MQESAETSFGFCEKR		
3UQY	MT	LGIFR	SM	MSAIITYMVTF		
3UQY	SL	VQSWDDD A	SM	MSAIITYMVTF		
3UQY	LS	GGKNPHPN	SM	MSAIITYMVTF		
3ZWL	BE	HKIFEE	EV	RELLKQWTEYREKIG QEMEKSM	FD	RELLKQWTEYREKIGQEME KSM
4F4O	FD	SCRTA	BA	LGRLLVV	BA	RLLGNVIVVVLARRL
			ED	LGRLLVV	ED	LLGNVIVVVLARRL
			HG	LGRLLVV	HG	LLGNVIVVVLARRL
			KJ	LGRLLVV		
4F6R	BC	KGHYT	BA	LRKLAVNM	CB	AMLERLQEKDKHAEEVRK NK
4FIP	HA	LKKYF	DA	AIVDHL	ED	YFIRHSM
4GD3	SL	VQSWDDD A	SM	SAIITYMVTF		
4GD3	LS	GGKNPHPN	SM	SAIITYMVTF		
4GD3	TM	CIQSGH	SM	SAIITYMVTF		
4GDK	CB	DRLQR	CB	RWKRHISEQLRRRD	CB	FEEIILQYN
			FE	WKRHISEQLRRRDRL	FE	AFEEIILQYN
4GDL	CB	DRLQR	CB	WKRHISEQLRRRDRL	CB	QAFEEIILQYN

Supplementary Table 6. Analysis of all interface loops and the hot loop set with respect to protein function. Annotated function for each protein-protein interaction was identified for the total PDB input set of proteins, the total interface loop set of proteins, and the hot loop set of proteins Only oxidoreductase and lyase enzymes seemed to be more highly represented in the final hot loop data set of proteins compared to the input and total interface loop set. Only categories that had at least one protein of that function in the hot loop dataset are shown and tallied.

A. Functional categories contained in the dataset.

B. Number of each category of protein in the total PDB

C. Number of each category of protein in the total interface loop dataset

D. Number of each category of protein in the hot loop dataset

E. $\frac{\text{Number from B.}}{\text{96,692 (total number of proteins in the PDB)}} \times 100$

F. $\frac{\text{Number from C.}}{\text{25,005 (proteins in the total interface loop set)}} \times 100$

G. $\frac{\text{Number from D.}}{\text{1,407 (proteins in the hot loop set)}} \times 100$

A. CATEGORY	B. # in PDB	C. # in total loop set	D. # in hot loop set	E. % of PDB	F. % of total loop set	G. % of hot loop set
OXIDOREDUCTASE	8786	3980	287	9.09	15.92	20.40
LYASE	3324	1926	115	3.44	7.70	8.17
HYDROLASES	16320	3088	178	16.88	12.35	12.65
ISOMERASE	2026	1070	71	2.10	4.28	5.05
STRUCTURAL GENOMICS	2494	1140	71	2.58	4.56	5.05
TRANSFERASE	12032	3611	198	12.44	14.44	14.07
IMMUNE SYSTEM	2515	716	20	2.60	2.86	1.42
TRANSPORT PROTEIN	2040	311	10	2.11	1.24	0.71
TRANSFERASE/INHIBITOR	1403	83	4	1.45	0.33	0.28
HYDROLASE/HYDROLASE INHIBITOR	2233	344	15	2.31	1.38	1.07
LIGASE	1551	662	26	1.60	2.65	1.85
RNA	946	4	1	0.98	0.02	0.07
TRANSCRIPTION	2779	697	29	2.87	2.79	2.06

RIBOSOME	716	8	1	0.74	0.03	0.07
SIGNALING PROTEIN	2096	353	25	2.17	1.41	1.78
ELECTRON TRANSPORT	1144	144	7	1.18	0.58	0.50
MEMBRANE PROTEIN	1243	253	8	1.29	1.01	0.57
CHAPERONE	955	247	6	0.99	0.99	0.43
DNA BINDING PROTEIN	1261	340	11	1.30	1.36	0.78
TOXIN	890	184	18	0.92	0.74	1.28
FLAVOPROTEIN	89	25	8	0.09	0.10	0.57
METAL BINDING PROTEIN	1044	153	13	1.08	0.61	0.92
RNA BINDING PROTEIN	807	115	6	0.83	0.46	0.43
CELL ADHESION	915	189	7	0.95	0.76	0.50
CELL CYCLE	554	222	14	0.57	0.89	1.00
SUGAR BINDING PROTEIN	887	166	7	0.92	0.66	0.50
VIRAL PROTEIN	1984	421	24	2.05	1.68	1.71
OXIDOREDUCTASE/INHIBITOR	432	34	1	0.45	0.14	0.07
TRANSCRIPTION REGULATOR	405	165	4	0.42	0.66	0.28
PROTEIN BINDING	1249	263	20	1.29	1.05	1.42
APOPTOSIS	399	37	1	0.41	0.15	0.07
UNKNOWN FUNCTION	873	245	17	0.90	0.98	1.21
DE NOVO PROTEIN	342	23	1	0.35	0.09	0.07
ION TRANSPORT	16	11	4	0.02	0.04	0.28
BIOSYNTHETIC PROTEIN	339	136	9	0.35	0.54	0.64
ALLERGEN	112	29	5	0.12	0.12	0.36
HORMONE	371	31	3	0.38	0.12	0.21
OXIDOREDUCTASE/ELECTRON TRANSPORT	53	59	4	0.05	0.24	0.28
OXYGEN TRANSPORT	353	38	5	0.37	0.15	0.36
HORMONE/GROWTH FACTOR	270	23	1	0.28	0.09	0.07
SERINE PROTEASE	132	15	4	0.14	0.06	0.28
PROTEIN TRANSPORT	649	175	7	0.67	0.70	0.50
IMMUNOGLOBULIN	179	39	5	0.19	0.16	0.36
ANTIBIOTIC RESISTANCE	22	13	3	0.02	0.05	0.21
LIPID BINDING PROTEIN	356	45	3	0.37	0.18	0.21
STRUCTURAL PROTEIN	995	208	13	1.03	0.83	0.92
METAL TRANSPORT	378	62	3	0.39	0.25	0.21
LUMINESCENT PROTEIN	221	22	1	0.23	0.09	0.07
GENE REGULATION	410	67	4	0.42	0.27	0.28
TRANSPORT	69	15	3	0.07	0.06	0.21
VIRAL PROTEIN/DE NOVO PROTEIN	2	4	2	0.00	0.02	0.14
LIGHT HARVESTING PROTEIN	3	6	2	0.00	0.02	0.14
PHOTOSYNTHESIS	204	90	4	0.21	0.36	0.28
KETOLISOMERASE	3	21	2	0.00	0.08	0.14
DEHYDROGENASE	16	8	2	0.02	0.03	0.14

PENICILLIN BINDING PROTEIN	14	11	2	0.01	0.04	0.14
CALCIUM BINDING PROTEIN	164	11	1	0.17	0.04	0.07
HEME BINDING PROTEIN	30	19	2	0.03	0.08	0.14
PLANT PROTEIN	252	65	5	0.26	0.26	0.36
LIGASE/INHIBITOR	79	9	2	0.08	0.04	0.14
NEUROPEPTIDE	57	15	2	0.06	0.06	0.14
TRANSLATION	338	92	4	0.35	0.37	0.28
BIOTIN BINDING PROTEIN	144	19	1	0.15	0.08	0.07
HORMONE RECEPTOR	115	8	1	0.12	0.03	0.07
NUCLEAR PROTEIN	95	14	2	0.10	0.06	0.14
LECTIN	166	32	3	0.17	0.13	0.21
TRANSFERASE/HYDROLASE	16	25	1	0.02	0.10	0.07
REPLICATION	210	72	3	0.22	0.29	0.21
MONOOXYGENASE	6	21	1	0.01	0.08	0.07
OXIDOREDUCTASE/IMMUNE SYSTEM	8	21	1	0.01	0.08	0.07
BACTERIAL ANTIBIOTIC RESISTANCE	1	1	1	0.00	0.00	0.07
ADP-RIBOSYLATION	2	1	1	0.00	0.00	0.07
RECEPTOR	120	12	1	0.12	0.05	0.07
ENDOCYTOSIS	69	20	2	0.07	0.08	0.14
IMMUNE SYSTEM/ANTIMICROBIAL PROTEIN	1	2	1	0.00	0.01	0.07
MEMBRANE PROTEIN/CELL ADHESION	1	2	1	0.00	0.01	0.07
POSTSEGREGATIONAL KILLING SYSTEM	1	2	1	0.00	0.01	0.07
TRYPTOPHAN BIOSYNTHESIS	1	2	1	0.00	0.01	0.07
VIRAL PROTEIN/PROTEIN BINDING	5	1	1	0.01	0.00	0.07
NITRITE REDUCTASE	1	3	1	0.00	0.01	0.07
VIRAL PROTEIN/SIGNALING PROTEIN	4	2	1	0.00	0.01	0.07
PROTEIN TURNOVER	3	16	1	0.00	0.06	0.07
MEMBRANE PROTEIN/OXIDOREDUCTASE	2	3	1	0.00	0.01	0.07
PROTEIN BINDING/BLOOD CLOTTING	2	3	1	0.00	0.01	0.07
PROTEIN SYNTHESIS/TRANSFERASE	2	3	1	0.00	0.01	0.07
PROTEIN TRANSPORT/IMMUNE SYSTEM	2	3	1	0.00	0.01	0.07
IMMUNE SYSTEM/CYTOKINE	9	1	1	0.01	0.00	0.07
MEMBRANE TRANSPORT	3	3	1	0.00	0.01	0.07
MOLYBDENUM-IRON PROTEIN	1	4	1	0.00	0.02	0.07
STRUCTURAL PROTEIN/CHAPERONE	4	3	1	0.00	0.01	0.07
DOMAIN SWAPPING	2	4	1	0.00	0.02	0.07
RECOMBINATION	128	34	1	0.13	0.14	0.07
COMPLEMENT REGULATOR	5	3	1	0.01	0.01	0.07
IMMUNOLOGY	1	5	1	0.00	0.02	0.07
PROTEIN BIOSYNTHESIS	1	5	1	0.00	0.02	0.07
HALOPEROXIDASE	14	1	1	0.01	0.00	0.07
ADENOVIRUS	2	5	1	0.00	0.02	0.07

PORIN	10	2	1	0.01	0.01	0.07
IMMUNE SYSTEM/HORMONE RECEPTOR	1	6	1	0.00	0.02	0.07
OXYGENASE	5	4	1	0.01	0.02	0.07
RIBOSOME INHIBITOR	5	4	1	0.01	0.02	0.07
TRANSFERASE/RNA BINDING PROTEIN	5	4	1	0.01	0.02	0.07
OXYGEN TRANSPORT/TRANSPORT PROTEIN	1	7	1	0.00	0.03	0.07
CARBOXY-LYASE	2	6	1	0.00	0.02	0.07
TRANSCRIPTION REGULATION	124	16	1	0.13	0.06	0.07
ANTITOXIN	4	5	1	0.00	0.02	0.07
DNA RECOMBINATION	6	4	1	0.01	0.02	0.07
FORMYLTRANSFERASE	2	7	1	0.00	0.03	0.07
OXIDOREDUCTASE/OXIDOREDUCTASE	2	11	1	0.00	0.04	0.07
HYDROLASE/REPLICATION	3	7	1	0.00	0.03	0.07
LYASE/OXIDOREDUCTASE	3	11	1	0.00	0.04	0.07
TRANSFERASE/RECEPTOR	7	5	1	0.01	0.02	0.07
TRANSCRIPTION/HYDROLASE	4	10	1	0.00	0.04	0.07
HYDROLASE/TRANSFERASE	18	20	1	0.02	0.08	0.07
ACETYLCHOLINE-BINDING PROTEIN/AGONIST	17	3	1	0.02	0.01	0.07
TOXIN/ANTITOXIN	7	9	1	0.01	0.04	0.07
OXIDOREDUCTASE/PROTEIN BINDING	8	9	1	0.01	0.04	0.07
GLYCOGEN PHOSPHORYLASE	18	4	1	0.02	0.02	0.07
SIGNALING PROTEIN/PROTEIN BINDING	19	4	1	0.02	0.02	0.07
TRANSCRIPTION/ACTIVATOR	12	7	1	0.01	0.03	0.07
HYDROLASE/HYDROLASE REGULATOR	11	8	1	0.01	0.03	0.07
ENDOCYTOSIS/EXOCYTOSIS	101	12	1	0.10	0.05	0.07
DIOXYGENASE	18	13	1	0.02	0.05	0.07
CELL INVASION	79	30	1	0.08	0.12	0.07
COAGULATION	22	8	1	0.02	0.03	0.07
ANTITUMOR PROTEIN	47	5	1	0.05	0.02	0.07
TRANSCRIPTION REPRESSOR	25	8	1	0.03	0.03	0.07
KINASE	65	7	1	0.07	0.03	0.07
VIRAL PROTEIN/IMMUNE SYSTEM	102	29	1	0.11	0.12	0.07
CIRCADIAN CLOCK PROTEIN	38	20	1	0.04	0.08	0.07
GROWTH FACTOR	93	16	1	0.10	0.06	0.07
TRANSCRIPTION ACTIVATOR	49	11	1	0.05	0.04	0.07

Supplementary Information References:

(1) Kortemme, T.; Baker, D. *Proc. Natl. Acad. Sci.* **2002**, *99*, 14116–14121.
(2) Fersht, A. R.; Shi, J.-P.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G. *Nature* **1985**, *314*, 235–238.
(3) Golovin, A.; Henrick, K. *BMC Bioinformatics* **2008**, *9*, 312.
(4) Janin, J.; Miller, S.; Chothia, C. *J. Mol. Biol.* **1988**, *204*, 155–164.
(5) Cunningham, B.; Wells, J. *Science* **1989**, *244*, 1081–1085.
(6) Clackson, T.; Wells, J. *Science* **1995**, *267*, 383–386.
(7) Cunningham, B. C.; Wells, J. A. *J. Mol. Biol.* **1993**, *234*, 554–563.
(8) Huang, J.; Wan, B.; Wu, L.; Yang, Y.; Dou, Y.; Lei, M. *Cell Res* **2012**, *22*, 1078–1081.
(9) DeLano, W. L. *The PyMol Molecular Graphic System*; DeLano Scientific LLC: San Francisco, CA.
(10) Bergey, C. M.; Watkins, A. M.; Arora, P. S. *Bioinformatics* **2013,** *29*, 2806–2807.
(11) Kortemme, T.; Kim, D.; Baker, D. *Sci. STKE Signal Transduct. Knowl. Environ.* **2004**, *2004*, pl2.

Supplementary Data Set 1. The entire set of hot loops generated by LoopFinder. These are those loops that meet the requirements previously discussed and outlined in the methods description.

A. PDB ID numbers for each protein complex identified to have a hot loop at the interface
B. Structure title
C. Functional category of the protein that takes part in the interaction
D. The chain on which the hot loop is located
E. The partner that binds to the exposed hot loop
F. The length of the sequence identified as a hot loop, in number of amino acids
G. The number of the first residue in the hot loop
H. The number of the last residue in the hot loop
I. Linker length, the distance between the N and C terminus of the hot loop peptide, in Angstroms
J. The sequence of the hot loop
K. The calculated $\Delta\Delta G_{residue}$ as calculated using computation alanine methods developed by Kortemme et al. (reference 11)
L. The average energy of the loop, $\Delta\Delta G_{loop,avg.}$ is the average of all values from column K.
M. The sum of all $\Delta\Delta G_{residue}$ values from column K. (reference 11)
N. The hot loops percent contribution to the total interface energy
O. Comments: (*) denotes that the $\Delta\Delta G_{loop,sum}$ calculated for the hot loop is a negative value. Negative values denote the possibility that substitution of alanine may improve binding, as in the prevention of an optimal, buried hydrogen bond (as described in reference 1). (+) denotes that the calculated total interface energy is a negative value.
P. Loop types for as identified using PDBeMotif (reference 3)

Supplementary Data Set 2. The subset of 364 hot loops that do not contain two or more consecutive hot spots. These loops were the main focus for identifying biologically relevant protein complexes that may be best targeted using constrained macrocycles.

A. PDB ID numbers for each protein complex identified to have a hot loop at the interface
B. Structure title
C. Functional category of the protein that takes part in the interaction
D. The chain on which the hot loop is located
E. The partner that binds to the exposed hot loop
F. The length of the sequence identified as a hot loop, in number of amino acids
G. The number of the first residue in the hot loop
H. The number of the last residue in the hot loop
I. The sequence of the hot loop
J. The calculated $\Delta\Delta G_{loop}$ as calculated using computation alanine methods developed by Kortemme et al. (reference 11)
K. Number of hot spots located within the loop