Protein-Protein Interactions Mediated by Helical Tertiary Structure Motifs

Andrew M. Watkins, Michael G. Wu, and Paramjit S. Arora*

Department of Chemistry, New York University, New York, New York 10003, United States

SUPPORTING INFORMATION

Table of Contents

Figure S1: Distribution of Amino Acids in Helical Dimers S2
Figure S2: Distribution of Hot Spot Residues in Helical Dimers S3
Figure S3: Average ΔΔG of Hot Spot Residues in Helical Dimers S4
Figure S4: SOCKET Analysis Applied to Helical Dimers S5
Figure S5: Non-Canonical Packing at Dimer Interfaces S6
Figure S6: Frequency of and SASA Burial by Hydrophilic Inner Groove Residues S7
Figure S7: ΔΔG of Inter-Strand Contacts Versus Hole Composition S8
Table S1: Helical Dimer Interfaces Implicated in Transcription or Transcription Regulation S9-11
Common Motifs Linking Parallel Coiled Coils S11
Methods S12-S14
Figure S5: Flow Chart Describing Dataset Curation S12
HippDB Dataset Curation S12
Derivation of a Helical Dimer Dataset S12-13
Interpreting the metric of "percent complex ΔΔG" for inhibitor design S13
SOCKET Analysis S13
Figure S1. The distribution of amino acids within Case 1 (a), Case 2 (b), and Case 3 (c), which comprise 262, 261, and 490 helices respectively (left; bars colored by amino acid type) is broadly similar to the distribution across the entire database of 37,055 high affinity helices (right; gray bars).
Figure S2. The distribution of hot spot residues in Case 1 (a), Case 2 (b), and Case 3 (c) helical dimers (bars colored by residue type) versus high affinity interface helices (gray bars) normalized to the overall frequency of the residue in those classes.
Figure S3. Average ΔΔG for hot spots in high affinity helix bundles (left; colored bars) compared to all helices with at least 6.0 R.E.U. total ΔΔG (right; grey bars). (Only one cysteine hot spot is present in the helix bundles.) We employed this restricted set for our alpha helical baseline because on average higher affinity helices will have higher ΔΔG per hot spot residue.
Figure S4. Plot describing how well each Case of helical dimers (Case 1, light gray; Case 2, gray; Case 3, black) compares to coiled coil packing, as found by SOCKET analysis. “Coiled Coil” refers to the SOCKET coiled coil assignment; “One Interaction” to the identification of a single complementary knob-in-hole packing interaction; “Non-CC” to SOCKET failing to identify any canonical packing at all.
Figure S5. Non-canonical packing at helix dimer interfaces. Key residues to helical dimer interaction affinity are depicted in stick. (a) The helices of the Case 1 helical dimer from 1,2-hydroquinol dehydrogenase have low affinity for each other and exhibit no energetically important aliphatic packing. (b) The Case 3 dimer governing allantoinase homodimerization is high-affinity but obtains that affinity predominantly from a hydrogen bonding network. PDB Codes: 1TMX, 3HM7. The Case 1 dimer (Figure S5a) contains a well-oriented Arg-Glu salt bridge, an Arg-Gln hydrogen bond, and a Gln-carbonyl hydrogen bond, as well as two interactions where aliphatic residues pack against nonpolar atoms of charged side chains (Leu against the beta carbon of an Arg and Val against the beta carbon of a His). The Case 3 dimer (Figure S5b) features an entirely charged and polar inner groove. Pairs of powerful Arg-Glu salt bridges secure both sides of the helix-helix interface, and secondary hydrogen bonding interactions from flanking Gln residues stabilize the Arg side chain conformations.
Figure S6. a) The distribution of hydrophilic contacts in the inner groove of the 262 Case 1 helical dimers. b) The average % SASA buried by each hydrophilic inner groove residue type.
Figure S7. Average ΔΔG for each Case 3 contact if it packs into a primarily complementary (left, dark grey bars) versus a primarily noncomplementary hole (right, light grey bars). For each amino acid, we required two of three hole residues to be of a complementary residue type (e.g. for a cation, either an aromatic or anionic residue type; for a polar, another polar; for an anion, a cation). [Histidine is the only residue type for which the ΔΔG for complementary and noncomplementary holes are not much different. This unusual case turns entirely on a set of high affinity XYH holes, where X and Y were usually alanine and glycine. Omitting these drops histidine’s average in noncomplementary holes to 0.84 R.E.U.]
| PDB Code | Dimer Case | Title                                                                                                                                 |
|----------|------------|---------------------------------------------------------------------------------------------------------------------------------------|
| 1WWM     | 1          | Crystal structure of conserved hypothetical protein tt2028 from an extremely thermophilic bacterium thermus thermophilus hb8              |
| 1YY7     | 1          | Crystal structure of stringent starvation protein a (sspa), an rna polymerase-associated transcription factor                          |
| 2IW5     | 1          | Structural basis for corest-dependent demethylation of nucleosomes by the human lsd1 histone demethylase                              |
| 2Q0O     | 1          | Crystal structure of an anti-activation complex in bacterial quorum sensing                                                        |
| 2XA6     | 1          | Structural basis for homodimerization of the src-associated during mitosis, 68 kd protein (sam68) qua1 domain                      |
| 3BNI     | 1          | Crystal structure of tetr-family transcriptional regulator from streptomyces coelicolor                                          |
| 4ADZ     | 1          | Crystal structure of the apo form of a copper-sensitive operon regulator (csor) protein from streptomyces lividans                 |
| 2FMM     | 2          | Crystal structure of emsy-hp1 complex                                                                                             |
| 2H8N     | 2          | Structure of a glutamine-rich domain from histone deacetylase 4                                                                     |
| 3BRV     | 2          | Nemo/ikkb association domain structure                                                                                             |
| 4FN6     | 2          | Structural characterization of thiaminase type ii tena from staphylococcus aureus                                                    |
| 1B9M     | 3          | Regulator from escherichia coli                                                                                                     |
| 1EYV     | 3          | The crystal structure of nusb from mycobacterium tuberculosis                                                                      |
| 1FT9     | 3          | Structure of the reduced (feii) co-sensing protein from r. Rubrum                                                                  |
| 1LJ9     | 3          | The crystal structure of the transcriptional regulator slya                                                                       |
| 1OJL     | 3          | Crystal structure of a sigma54-activator suggests the mechanism for the conformational switch necessary for sigma54 binding    |
| 1PQ9     | 3          | Human lxr beta hormone receptor complexed with t0901317 complex                                                                     |
| 1U9H     | 3          | Heterocyclic peptide backbone modification in gcn4-pli based coiled coils: replacement of e(22)(l(23)                                    |
| 1Y7Q     | 3          | Mammalian scan domain dimer is a domain-swapped homologue of the hiv capsid c-terminal domain                                        |
| 1YY7     | 3          | Crystal structure of stringent starvation protein a (sspa), an rna polymerase-associated transcription factor                         |
| 2AZE     | 3          | Structure of the rb c-terminal domain bound to an e2f1-dp1 heterodimer                                                            |
| 2BG6     | 3          | Prfa-g145s, a constitutive active mutant of the transcriptional regulator in L.monocytogenes                                       |
| 2FSW     | 3          | Crystal structure of the putative transcriptional regualator, marr family from porphyromonas gingivalis w83                         |
| 2IW5     | 3          | Structural basis for corest-dependent demethylation of nucleosomes by the human lsd1 histone demethylase                            |
| 2LD7     | 3          | Solution structure of the msin3a pah3-sap30 sid complex                                                                                 |
| 2OQG     | 3          | Arsr-like transcriptional regulator from rhodococcus sp. Rha1                                                                        |
| 2P4W     | 3          | Crystal structure of heat shock regulator from pyrococcus furiosus                                                                    |
| 2PQQ     | 3          | Structural genomics, the crystal structure of the n-terminal domain of a transcriptional regulator from streptomyces coelicolor a3(2) |
| 2Q0O     | 3          | Crystal structure of an anti-activation complex in bacterial quorum sensing                                                       |
| 2QWW     | 3          | Crystal structure of multiple antibiotic-resistance repressor (marr) (yp_013417.1) from listeria monocytogenes 4b f2365 at 2.07 a resolution |
| 2WG5     | 3          | Proteasome-activating nucleotidase (pan) n-domain (57-134) from                                                                        |
archaeoglobus fulgidus fused to gcn4

2WPZ 3  Gcn4 leucine zipper mutant with two vxxxxxx motifs coordinating chloride

2WUQ 3  Crystal structure of blab protein from streptomyces cacaoi

2XHK 3  Crystal structure of transcription factor ntca from synechococcus elongatus bound to 2-oxoglutarate

2YY0 3  Crystal structure of ms0802, c-myc-1 binding protein domain from homo sapiens

3A7M 3  Structure of flit, the flagellar type iii chaperone for flid

3BRV 3  Nemo/ikkb association domain structure

3BY6 3  Crystal structure of a transcriptional regulator from oenococcus oeni

3CUO 3  Crystal structure of the predicted dna-binding transcriptional regulator from e. Coli

3D0S 3  Human p53 core domain with hot spot mutation r249s (ii)

3ESU 3  Ocpa complexed crpk (c200s)

3ECO 3  Crystal structure of mepr, a transcription regulator of the staphylococcus aureus multidrug efflux pump mepa

3ER6 3  Crystal structure of a putative transcriptional regulator protein from vibrio parahaemolyticus

3FXQ 3  Crystal structure of the lysr-type transcriptional regulator tsar

3ISP 3  Crystal structure of argp from mycobacterium tuberculosis

3KCC 3  Crystal structure of d138l mutant of catabolite gene activator protein

3L0L 3  Crystal structure of orphan nuclear receptor rorgamma in complex with natural ligand

3LA7 3  Crystal structure of ntca in apo-form

3LHR 3  Crystal structure of the scan domain from human znf24

3M8J 3  Crystal structure of e.coli focb at 1.4 a resolution

3NS0 3  Human early b-cell factor 3 (ebf3) ipt/tig and hlhlh domains

3PQK 3  Crystal structure of the transcriptional repressor bigr from xylella fastidiosa

3QP8 3  Crystal structure of cvir (chromobacterium violaceum 12472) ligand-binding domain bound to c10-hsl

3R6S 3  Crystal structure of glxr transcription factor from corynebacterium glutamicum with camp

3R84 3  Structure of the mediator head subcomplex med11/22

3TGN 3  Crystal structure of the zinc-dependent marr family transcriptional regulator adcr in the zn(ii)-bound state

3V3E 3  Crystal structure of the human nur77 ligand-binding domain

3VP9 3  Crystal structure of the n-terminal domain of the yeast general corepressor tup1p mutant

4ADZ 3  Crystal structure of the apo form of a copper-sensitive operon regulator (csor) protein from streptomyces lividans

4AIH 3  Crystal structure of rova from yersinia in its free form

4BG7 3  Bacteriophage t5 homolog of the eukaryotic transcription coactivator pc4 implicated in recombination-dependent dna replication

4DEJ 3  Crystal structure of glutathione transferase-like protein il0419 (target efi-501089) from idiomarina loihiensis l2tr

4FBI 3  Crystal structure of an r46a mutant of the restriction-modification controller protein c.esp1396i (trigonal form)

4HBL 3  Crystal structure of abfr of staphylococcus epidermidis

4HH2 3  Structure of ppsr without the hth motif from rb. Sphaeroiides
Common motifs linking parallel Case 1/2 coiled coils

Loop/turn, strand, loop/turn, helix, loop/turn:
Case 1: 1C41, 1EK6, 1GQO (5 examples), 1GTZ, 1GY8, 1IY8, 1KZ1, 1MXH, 1UQR (4 examples), 1XHL, 2BD0, 2CFC (2 examples), 2F59, 2JAQ, 2OKF, 2UYG (4 examples), 2VP4 (2 examples), 3A28 (2 examples), 3AWD, 3AY7, 3GDF, 3GUY (2 examples), 3KIP (3 examples), 3N8K, 3O9Z, 3SSS, 3UCE, 3TZQ (2 examples), 4GKB, 4IBO (2 examples), 4IIN, 4TU (2 examples), 4J07
Case 2: 3GRK, 4EIT (2 examples), 4J07 (2 examples)

Strand bracketed by two loops/turns:
Case 1: 1DI0 (2 examples), 1V4V, 1XG4, 2CBY (4 examples), 2C92, 3BOF, 3IBG, 3NNN, 3NTL, 3NYW, 4BNW, 4JCQ
Case 2: 2NWQ, 2OBX, 2C92 (2 examples), 3B8F, 3GEM, 3R7F, 3UFX

Loop-helix-loop:
Case 1: 1NXH (2 examples), 1OAH, 1QDB
Case 2: 1GVN, 1J1J, 2M6I, 2P5T, 3ZC1

Loop-helix-loop-helix-loop-helix-loop:
Case 1: no examples
Case 2: 1J1G, 109R, 1TJ0, 1WZ8, 2CHP, 2D5K, 2IY4, 2VXX, 2VZB, 3AK8 (2 examples)
Figure S8. The process (reproduced from Figure 2) by which the multichain protein structures in the PDB are analyzed to produce the present dataset.

Methods

The original protocol for creating the HippDB dataset has been described: *Bioinformatics* 2013, 29, 2806-2807. The updated protocol that generated the expanded dataset in use here is described in *ACS Chem. Biol.* 2014, 9, 1747. The biological assemblies corresponding to NMR and X-ray crystal structures are acquired from the PDB subject to a 95% sequence redundancy filter. Each assembly is subjected to Rosetta’s all-atom refinement algorithm “relax” with restraints to the starting coordinates. Extra rotamers are generated for the first and second chi angles and the input side-chain is added to the rotamer set. Five models are generated with the talaris2013 scoring function and the best model by total score separated into two-chain files. Each two-chain complex is subjected to computational alanine scanning through RosettaScripts, where the ΔΔG of a point mutation is evaluated as

\[
\Delta \Delta G = (E_{\text{mut, complex}} - E_{\text{mut, separated}}) - (E_{\text{wt, complex}} - E_{\text{wt, separated}})
\]

and where the mutated and separated structures are repacked in an 8Å radius around the mutated residue. Twenty repetitions are averaged to obtain ideal convergence. Secondary structure is assigned for each complex using the dssp executable distributed by the Center for Molecular and Biomolecular Informatics. The results are parsed to identify helices four or more residues long containing two or more hot spot residues; in addition to generating SQL to populate a database table (accessible at http://www.nyu.edu/projects/arora/ppidb), it generates a tab-separated data table for further parsing.

Helical dimer data set curation

We wrote a Python script to parse the HippDB data set. This script first filters out any protein complexes containing only one helix. Subsequently, it identifies all the helices in a given protein complex with total ΔΔG ≥ 6.0 R.E.U. For each such pair of helices, it measures the distance and angle between the helical
axes. If the distance is less than $17\Delta$ and the angle is within $30^\circ$ of parallel or antiparallel, it computes the percentage of protein residues present at the protein-protein interface and the percentage of complex $\Delta\Delta G$ contributed by the helical interface in question.

A heuristic inter-helix distance is computed by finding the nearest residue on chain B to each residue on chain A by $Ca-Ca$ distance (and vice-versa). The largest of these “nearest residue distances” is then assigned as the inter-helix distance. Similarly, the angle between the helix axes is approximated as the angle between the first-to-last $Ca$-to-$Ca$ vectors for each helix.

Upon finding a compatible pair of helices, the script records essential data about the helices in question: the total $\Delta\Delta G$ and $\Delta SASA$, the start and end residue, the sequence, the distance and angle, the hot spot residues, and so forth.

**Interpreting the metric of “percent complex $\Delta\Delta G$” for inhibitor design**

Up to the approximation that we are relying on an approximately fixed backbone and that multiple alanine point mutants would combine linearly, the summed $\Delta\Delta G$ is the difference between the native interaction energy and the interaction energy for a complex whose interface is entirely mutated to alanine. The interaction energy for such an all-alanine interface is a useful baseline for both inhibitor design and structural characterization because it is featureless and not amenable to any form of sequence-specific mimicry.

Furthermore, the interaction energy for such an all-alanine interface is typically zero, and thus the summed $\Delta\Delta G$ approximately equals the native interaction energy, except in specific cases, such as where the interface makes multiple sequence-independent (e.g. backbone mediated) hydrogen bonds. This is obviously uncommon in helix-dominated interfaces.

We find this metric to be a concise summary statistic for constructing interaction inhibitors, as it may be interpreted as the $\Delta\Delta G$-weighted proportion of key residues for this interaction that are found on this helical dimer.

**SOCKET analysis**

We built the SOCKET program, using make, from source downloaded from the Woolfson group web site. We created PDB structures containing the residues from each helical dimer, as well as two residues flanking the N- and C-termini of each helix to eliminate any possible terminal effects in DSSP assignments. We ran SOCKET with command lines like

```
socket3.03/socket -f 4MP4_A.pdb -s 4MP4_A.dssp -o 4MP4_A.out
```

thus using the default packing radius of 7.0Å. Examining Case 1, we observed only 24 dimers classified as coiled coils—fewer than 10% of the category. (An additional 17 examples were not classified as coiled coils, but one mutual knob/hole interaction was identified). Increasing the packing radius to even 8.0Å only identified an additional 12 dimers as coiled coils.

In the original SOCKET paper, nine classical coiled-coil structures are contrasted to seven control structures. The former group all possess multiple complete layers of knob residues, identified at packing radii of 6.8Å or less, while the latter group possess only non-complementary interactions.