JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

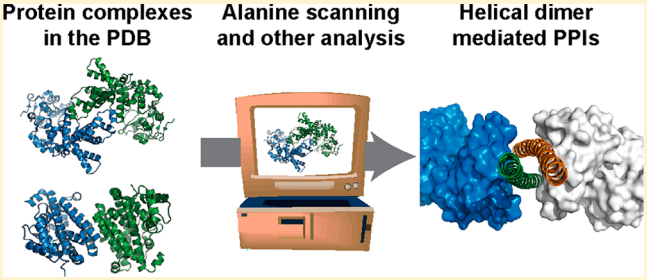
Protein–Protein Interactions Mediated by Helical Tertiary Structure Motifs

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

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

Supporting Information

ABSTRACT: The modulation of protein–protein interactions (PPIs) by means of creating or stabilizing secondary structure conformations is a rapidly growing area of research. Recent success in the inhibition of difficult PPIs by secondary structure mimetics also points to potential limitations, because often, specific cases require tertiary structure mimetics. To streamline protein structure-based inhibitor design, we have previously described the examination of protein complexes in the Protein Data Bank where α-helices or β-strands form critical contacts. Here, we examined coiled coils and helix bundles that mediate complex formation to create a platform for the discovery of potential tertiary structure mimetics. Though there has been extensive analysis of coiled coil motifs, the interactions between pre-formed coiled coils and globular proteins have not been systematically analyzed. This article identifies critical features of these helical interfaces with respect to coiled coil and other helical PPIs. We expect the analysis to prove useful for the rational design of modulators of this fundamental class of protein assemblies.



INTRODUCTION

Mimicry of interfacial protein segments has led to new classes of rationally designed inhibitors of protein–protein interactions (PPIs).[1−8] The identification and analysis of protein complexes mediated by protein secondary structures provide a platform for these explorations.[3,4,9] We have recently examined the full set of protein complexes in the Protein Data Bank mediated by α-helices[10−13] and β-strands.[14] Our work, along with efforts by Kritzer et al.[15] to define loop motifs at protein interfaces, aims both to *describe* the interactions present in the Protein Data Bank and to *prescribe* effective starting points for the design of PPI inhibitors.[4,9]

Individual secondary structures are critical elements of protein interfaces; however, many PPIs feature more complex modes of binding, suggesting a potential role for synthetic tertiary structure mimetics[16,17] or miniproteins[18,19] as attractive candidates for the design of new classes of PPI inhibitors. Miniproteins consisting of helical bundles, β-sheet barrels, and loops, along with synthetic antibodies,[20−22] are now routinely used to enrich ligands for protein targets, especially for extracellular receptors. In an effort to expand our atomic analysis of protein structural data beyond interactions that can be mediated by a single secondary structure element alone, we have developed new methodology to create a database of helical **d**imers at **p**rotein–**p**rotein interfaces (DippDB).

We chose to begin our survey of protein tertiary interactions by focusing on helix dimers because the dimer is the simplest all-helical tertiary structure stoichiometry. Coiled coils and helical bundles are well understood and have been extensively studied in diverse biochemical and biophysical contexts.[23−27] Dimeric coiled coils or similarly structured motifs such as bundles play essential roles in mediating biological processes, iconically driving the multimerization and stabilization of proteins involved in transcription factor complexes and vesicular trafficking, among other critical functions.[25,28] Several computational approaches have been implemented to predict coiled coil-mediated interactions by their pairwise and multimeric residue correlations.[29−32] Seminal studies have produced a comprehensive dataset of the coiled coil interactome.[33−35] However, computational and experimental methods for the analysis of coiled coils described thus far are largely devoted to characterization of forces that *lead to* coiled coil formation. To complement these studies, we sought to analyze interactions of helical dimers with globular proteins as a step toward the rational design of coiled coil mimetics as PPI inhibitors.[16,17] Though canonical coiled coils possess supercoiling and particular packing properties, we did not impose these requirements, stipulating only that the helices be proximal and well-oriented. Since our motivation for developing this dataset is to identify interactions that may not be inhibited by secondary structure mimics, we also required that critical binding residues be located on *both* helices. These criteria retain structures of high structural similarity to a coiled coil but eliminate canonical all-alpha tertiary structure motifs like the helix-loop-helix and helix-turn-helix DNA binding domains, whose interhelical angles are far from parallel or antiparallel.

Received: May 28, 2015
Published: August 24, 2015

DOI: 10.1021/jacs.5b05527
J. Am. Chem. Soc. 2015, 137, 11622−11630

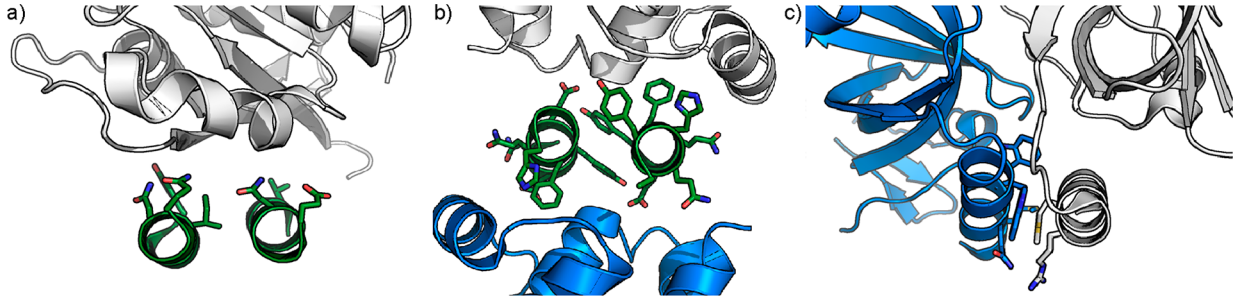


Figure 1. Models depicting three families of coiled coil-like structures at protein−protein interfaces. (a) In the first family (Case 1), a coiled coil entirely from chain A forms an interaction with protein B. (b) In the second family (Case 2), a coiled coil, which may come from one or two proteins, interacts with two different proteins partners. (c) In the third family (Case 3), a coiled coil forms across a protein interface.

Examination of the helix dimer dataset suggests that coiled coil interfaces can be divided into three broad categories (Figure 1) according to their interaction stoichiometry. Case 1 features a helical dimer from one protein interacting with a single partner protein. In Case 2, a helical dimer from one protein interacts with two different protein partners. In Case 3, a single helical dimer motif *forms* at the interface between partner proteins. We anticipate that helical dimers in Case 3 would favor different interacting residues from examples in Cases 1 and 2, because in Case 3 high-affinity interactions must form between two individual helices rather than a helix dimer and a globular protein. This taxonomy reflects the different properties demanded of potential designed inhibitors: Case 1 features interactions on predominantly one dimer face; Case 2 generally interacts with two faces; Case 3 dimer interfaces may be interrupted by a single helix.

We examined the biophysical properties of each class in relation to each other, to typical interface helices in general, and to *canonical* coiled coils. Coiled coils are defined as two or more α-helices that wind around each other to form supercoils.[23] Classical coiled coils are characterized by a heptad repeat, $(abcdefg)_n$, where buried *a* and *d* positions form the interface between partner helices. We sought to identify all helix dimers that are in contact with a globular protein irrespective of whether such helices would meet the strict definition. This study has revealed the existence of a set of biologically relevant complexes as potential targets for inhibitor design. We also analyzed the biophysical properties of dimer interfaces, such as the composition of hot spot residues and the degree to which helical dimers differ from coiled coils and protein helices in general. This analysis shows that hot spot residues are concentrated over compact areas, potentially allowing the mimicry of these dimers by small or medium-sized molecules.

RESULTS AND DISCUSSION

We started from the previously developed HippDB dataset.[10−12] HippDB was developed to comprise all interface helices of four or more residues that contain two or more hot spot residues, where hot spot residues are defined via computational alanine scanning, performed with Rosetta, as those that result in a loss of binding energy ($\Delta\Delta G$) of at least 1.0 Rosetta energy unit (REU, which scales approximately as 1 kcal/mol).[36,37]

In this work, we expanded the HippDB dataset to include over 37 000 high-affinity interfaces and imposed stringent geometric and energetic criteria to obtain a set of over 1000 high-affinity helical dimers. We made modifications to our prior methodology to tailor it to the coiled coil motif (Figure 2). In

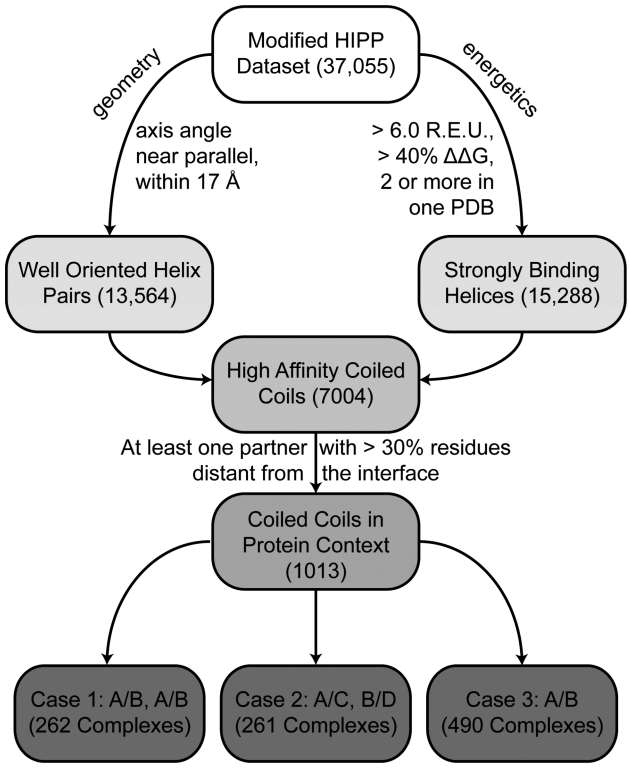


Figure 2. Schematic for identification of protein interfaces in the Protein Data Bank (PDB) where a helix dimer contributes significantly to complex formation. Interfacial helices from the previously described HippDB dataset were culled to produce a set of structures for detailed analysis via stringent distance, orientation, and energetic criteria. On the basis of our evaluations, we classified the interactions among three classes (Cases 1−3) of helical tertiary structure-mediated PPIs. The energetic contribution of each interface helix dimer and individual residues was approximated using Rosetta. The complete dataset is hosted at www.nyu.edu/projects/arora/dippdb/cc.php.

HippDB, we were interested in identifying minimal inhibitory motifs to aid the design of synthetic inhibitors. This goal requires identification of helical segments of a protein present at an interface without the sequences not in contact. In the context of helical dimers, the challenge is to capture the defining geometric parameters as accurately as possible so any energetically irrelevant residues far from the interface can be discarded separately. We altered our analysis such that any continuous stretch of helical residues counts as a helical element as long as part of the helix is present at the interface. This modified method identifies as a single element of interest

DOI: 10.1021/jacs.5b05527
J. Am. Chem. Soc. 2015, 137, 11622−11630

helix dimers in which one helix makes contact with a partner protein at multiple separate points along its length. Our earlier work would have split this helix into multiple helical elements.

We imposed both geometric and energetic criteria to obtain a dataset of interfaces where coiled coil-like structures play an important role. We stipulated that each helix must contribute at least 6.0 Rosetta energy units (REU) of $\Delta\Delta G$ in its interaction with its partner, that the angle between their helical axes must be within 30° of parallel or antiparallel, and that the two helices are within 17 Å of each other. These conditions ensured coiled coil-like geometry[24] and a substantial energetic contribution, equivalent to at least three strong hot spot residues, from each helix. The 6 REU threshold ensures that the selected coiled coil interfaces contain a sufficient number of hot spot residues to merit their mimicry by a potential synthetic inhibitor. Lower energetic thresholds yield additional compelling complexes, but far too many to investigate individually. We also imposed conditions on the percentage of the interaction's overall $\Delta\Delta G$ contributed by each helix. When both helices come from the same protein chain, we required each helix to contribute at least 20% of the chain's $\Delta\Delta G$. This requirement excluded any helical dimers that did not make a substantial total contribution to the binding interface, and ensured that the complexes in the database will truly require a dimer and are not amenable to disruption by mimic of a single helix.

Finally, we aimed to ensure that the dataset was not dominated by pairs of chains from multimeric helix bundles, as our interest lies in complexes where at least one partner is a globular protein. We observed that in structures of conventional helix bundles, almost every residue of each chain is present at the protein−protein interface. To distinguish dimer−protein interfaces from helix bundles, we required that at least one partner in every complex must have at minimum 30% of its residues distant from the interface.

The complete dataset is hosted at www.nyu.edu/projects/arora/ppidb/dippdb/cc.php. Dimers may be queried using PDB identification, total $\Delta\Delta G$ and ΔSASA, interhelical distances, and angles (Table 1).

Hot Spot Residues across Helix Dimer Interfaces Are Concentrated over Relatively Compact Areas. Archetypical coiled coil motifs consist of multiple heptad repeats. A cursory analysis would suggest that the hot spot residues may be distributed evenly over many heptads, requiring the design of large molecules or biologics for inhibition, as has been true for the design of coiled coil assembly inhibitors.[38] We were surprised to find that a plurality of complexes in DippDB possessed hot spot residues over a relatively small region of the interface (Figure 3). The average length of Case 1, 2, and 3 helical dimers is 19, 27, and 22 residues, respectively, while their average hot spot spans are 13, 17, and 16.5 residues. The critical residues are limited to a single heptad in 15% of Case 1 dimers and to two heptads in two-thirds of examples. The trend of compact hot segments[3] in helical dimers is observed in each Case with two-thirds of critical contact residues averaging fewer than three heptads. Dimers that span three heptads or less (roughly 30% of the dataset) average a hot spot per 4.7 residues. This signature is strongly suggestive of an interface amenable to inhibition by designed peptidomimetics. However, as peptidic coiled coils of these length scales are not generally stable,[39] we have undertaken an experimental effort to develop cross-linked helix dimers (CHDs) as minimal inhibitors of coiled coils and other helical PPIs.[17]

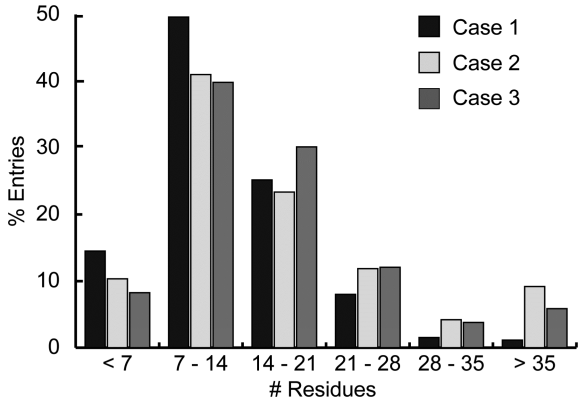


Figure 3. Histogram of the distance between the first and final hot spot residue in Case 1 (black), Case 2 (light gray), and Case 3 (dark gray) helical dimers.

Table 1. Selected Fields Recorded in DippDB[a]

field name	description
Title	title of the original PDB entry
Keywords	keywords included in the original PDB structure file, which often reflect function, localization, or other biology of interest
%ΔΔG[b]	percent of the total interaction $\Delta\Delta G$ contributed by one helix, as estimated by Rosetta
ΔSASA[b]	change in solvent-accessible surface area identified by NACCESS on one helix
Hot spot residues[b]	list of the hot spot residues (single letter code, residue number, and $\Delta\Delta G$) on the helix
Inter-helical angle	angle between the helix axes, approximated by the N- to C-terminal CA−CA vectors
%Residues at interface[b]	percentage of all the residues on the chain of one dimer helix that are found at the protein−protein interface

[a]Additional fields are available. Each field may be searched and sorted, and data may be downloaded in CSV, PDF, or XLS format. Complexes of weaker affinity ($\Delta\Delta G$ cutoff of 4 REU) are included for comparison in the website dataset. [b]Data provided for both participating helices.

Amino Acid Composition of Helix Dimers Differs from Coiled Coils. Further examination of DippDB interfaces suggests that helical dimers at protein−protein interfaces have comparable amino acid composition to other high-affinity interface helices[10,13] but significantly differ from canonical coiled coils. In general, classical coiled coil motifs possess a distribution of amino acids similar to the α-helix but with considerable additional bias toward aliphatic residues, owing to their obligate interior packing interactions.[23,26,30] We wished to understand whether helical dimers more closely reflected the amino acid distribution on high-affinity helices or classical coiled coils.

In examining the distribution of amino acids in these structures, we found it largely consistent with the distribution on helices in general (Supporting Information, Figure S1).[10,13] Though there is considerable selective pressure for coiled coil motifs to possess high proportions of aliphatic amino acids, especially leucine and isoleucine, for optimal knobs-into-holes packing arrangements, these residue types are not overwhelmingly enriched in the α-helices that are part of DippDB.

We also examined the distribution of hot spot residues on each family of helix dimers. A summary of these data is shown in Figure 4, and details for all three cases by residue types are included in the Supporting Information, Figure S2. Aliphatic and charged residues are moderately enriched as hot spot

DOI: 10.1021/jacs.5b05527
J. Am. Chem. Soc. 2015, 137, 11622−11630

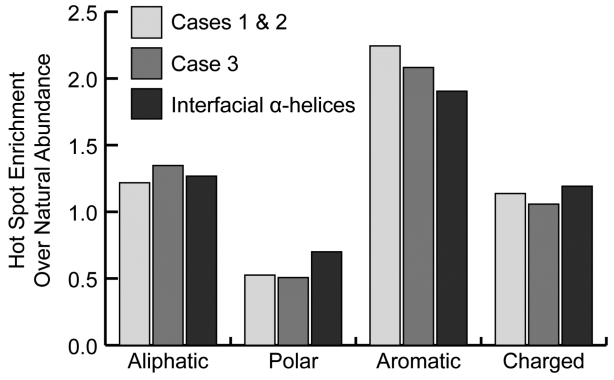


Figure 4. Frequency of hot spot residues, normalized to natural abundance, in helix dimers and interfacial single helices, respectively. The plot shows distribution of aliphatic residues (Leu, Ile, Val), polar residues (Gln, Asn, Ser), aromatic residues (Phe, Trp, Tyr), and charged residues (Arg, Asp, Glu, Lys) in the three contexts.

residues, though generally consistent with the helical baseline. In contrast, polar residues were greatly depleted, and more so than the general helical case; conversely, aromatic residues were more enriched in helical dimers.

Separate from our analysis of amino acid composition in general, we examined the inter-helix contacts made by Case 1/2 and Case 3 helical dimers and conventional coiled coils (Figure 5a). Helix dimers feature a larger proportion of interstrand contacts between polar residues than typical coiled coils. In the classical coiled coil motif, core polar mutations may be tolerated via changes in stoichiometry, local distortions in geometry, or an increased inner void volume.[30] The presence of polar contacts at the interior of helical dimers emphasizes the importance of their specific interactions. The leucine zipper coiled coil (Figure 5b) contains four paired aliphatic knob-into-hole packing interactions, which dominate the interaction. In the particular Case 1 dimer example depicted in Figure 5c, aliphatic packing interactions are limited and energetically insignificant; one helix only possesses one "aliphatic" residue (a threonine) facing its partner over five entire turns; in contrast, the nonpolar residues from the other helix pack into non-canonical holes. The Case 3 dimer example (Figure 5d) possesses energetically important nonpolar residues, but they are not organized into the classic interior groove of a coiled coil, and the phenylalanines are more than twice as energetically important as the leucines by $\Delta\Delta G$. Visual inspection of these examples and others inspired us to quantify the degree to which the helical dimer forms non-canonical packing interactions.

Quantifying the Non-canonicity of the Helical Dimer Motif. We found the examples of non-canonical inner grooves compelling and performed a more comprehensive analysis to determine how non-canonical these motifs are as compared to the classical coiled coil. We employed Woolfson's SOCKET analysis to explore the dataset (Supporting Information, Figure S4).[30] The SOCKET algorithm identifies knobs-into-holes packing of coiled coils to distinguish them from helix dimers. Of the 262 Case 1 dimers, only 24 dimers were identified as being coiled coils by SOCKET (9.2%), of which 22 were antiparallel and 2 were parallel. An additional 17 dimers contained only *one* complementary knob-into-hole packing interaction. In contrast, canonical coiled coils might have four such interactions per heptad (two per partner). Moreover, more than half of those 24 coiled coils identified were fewer than three heptads in length. This analysis implies that while

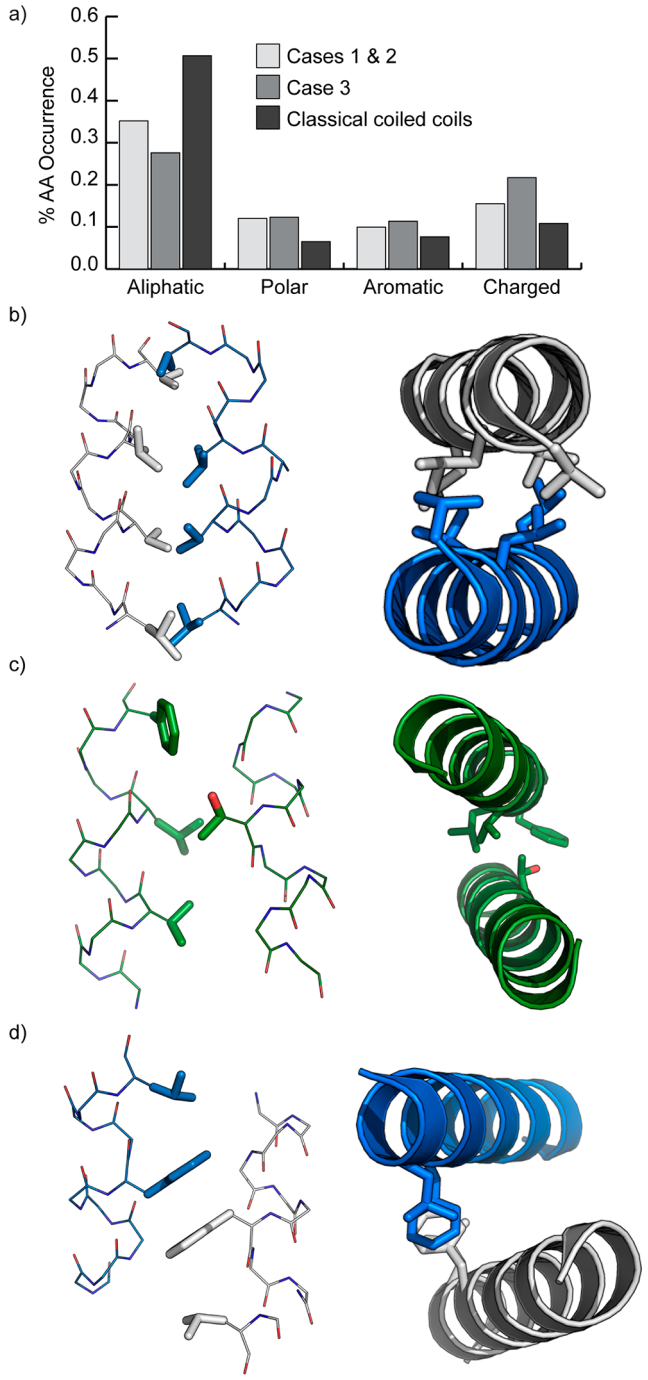


Figure 5. (a) Amino acid composition of inter-helical contact residues in helix dimers as opposed to true coiled coils. (b) The packing of a classic leucine zipper, GCN4, features an aliphatic *a*/*d* groove with each residue packing into a complementary hole. (c) This Case 1 helical dimer from 1,2-hydroquinol dehydrogenase homodimer illustrates highly non-canonical packing interactions. (d) The orphan nuclear receptor Nur77 contains an aliphatic core but lacks any heptad repeat structure and knob/hole packing orientation. PDB codes: 1LLM, 1TMX, and 3V3E.

helical dimers may occasionally exhibit packing characteristic of coiled coil motifs, they are too short to be stable on their own.[39] We also analyzed Case 2 and Case 3 dimers with SOCKET. 27/261 Case 2 dimers contained significant coiled coil structure (10.3%) and 21 contained one complementary interaction. Of Case 3 dimers, 133/919 were identified as coiled

DOI: 10.1021/jacs.5b05527
J. Am. Chem. Soc. 2015, 137, 11622−11630

coils by SOCKET (14.4%) and an additional 50 contained a single knob-in-hole interaction.

To follow up on the disparities in inner groove composition demonstrated in Figure 5, we specifically studied the frequency of hydrophilic inner grooves. We found hydrophilic inner groove residues quite common: each case averaged at least three, and 40% of complexes overall had at least four such residues. We profile two complexes whose interfacial $\Delta\Delta G$ results almost exclusively (>90%) from hydrophilic contacts in the Supporting Information, Figure S5. Such features are uncommon in canonical coiled coils. Polar residues may be found at the interior of "inside-out" coiled coil motifs found in membrane proteins,[40,41] though there is debate regarding the extent of polarity on the inside.[42−44] Because the vast majority of the proteins in DippDB are cytosolic, including these unusual examples, we hypothesized that a similar environment may make this inside-out geometry possible: the helical dimer is surrounded by hydrophobic residues presented by the protein in which it is found and by its binding partner, serving an analogous role to membrane lipids. The networks of buried hydrogen bonds that may form at the interior of a coiled coil are well studied, although the partially polar interiors observed here present an extreme case.[45]

We obtained the set of hydrophilic contact residues in Case 1 helical dimers. A total of 79.4% of hydrophilic contact residues were flanked by at least one nonpolar residue. Buried polar residues averaged under 30% relative SASA (Supporting Information, Figure S6). Typical SASA burial for these residue types is markedly lower. This degree of burial is highly destabilizing if hydrogen bonds are left buried but unsatisfied as a result, but it concomitantly increases the value of satisfied hydrogen bonds, contributing to the strength of these interfaces.[46,47] Overall, though it is possible to find recognizably canonical packing at the interior of helical dimers, the majority do not succumb to the same generalizations as the coiled coil motif.

Helical Dimer Affinity Depends on Complementary Packing Interactions. As discussed above, the knobs-into-holes packing of coiled coils distinguishes them from helical dimers.[30,48] Because the two helices of a Case 3 helical dimers come from different chains, we were able to compare knob−hole packing interactions to the $\Delta\Delta G$ of the knob residue by identifying the nearest three-residue hole on the partner helix (i, i+1, i+4 or i, i+3, i+4) to each knob. We restricted this analysis to the inner groove of each helix and furthermore recomputed the knob $\Delta\Delta G$ values on complexes only including the dimer helices so as to omit any interactions with other components of the protein. Though the packing is not conventional enough to identify the dimers as coiled coils by SOCKET, we anticipated that we would still be able to identify some trends. In contrast to the classical knobs-into-holes aliphatic model, residues of all types could form inner-groove contacts of considerable $\Delta\Delta G$. Instead, we found the key feature, tightly correlated to $\Delta\Delta G$, was that knob residues made contact with chemically complementary holes. Although aliphatic contact residues of low to moderate $\Delta\Delta G$ do frequently pack into polar holes and vice versa, every aliphatic and aromatic residue type had higher average $\Delta\Delta G$ when packing in a mostly aliphatic or aromatic hole than in a mostly polar one (Supporting Information, Figure S7).

Helix-Turn-Helix Motifs Commonly Secure the Orientation of Antiparallel Helical Dimers. In the creation of this dataset, we made no stipulation about the relative position of these helix bundles in the protein(s) in question. We conjectured that helix-turn-helix motifs, or helical hairpins, which are common sources of ideal antiparallel coiled coils,[49,50] might be particularly prominent interface elements. Of the 262 Case 1 helical dimers, 81 are separated by two to eight nonhelical residues. This substantial proportion of helix-turn-helix motifs is encouraging, as it suggests that the tertiary structure present at the interface is largely governed by local forces that may be mimicked by a designed inhibitor. Overall, there are 133 parallel and 129 antiparallel dimers; thus, development of scaffolds appropriate for *both* motifs is critical. In contrast, of the 115 Case 2 helical dimers in which both dimer helices come from the same chain, only 25 exhibit helix-turn-helix motifs. The possible interplay between interface tertiary structure geometry and the stoichiometry of formed complexes merits further study, as it suggests a difference in folding cooperativity.[51]

Parallel Helical Dimers Are Connected by Diverse Motifs. Even though the N and C termini of parallel dimers on the same chain are distant, the majority of parallel dimers are connected by a small number of motifs. Of the 133 parallel Case 1 dimers, 55 are connected by a two to four residue loop, a strand whose length varies with that of the dimer, a single residue turn, a short 3_{10} or α-helix, followed by another short loop (Figure 7). The second most prevalent motif includes 16

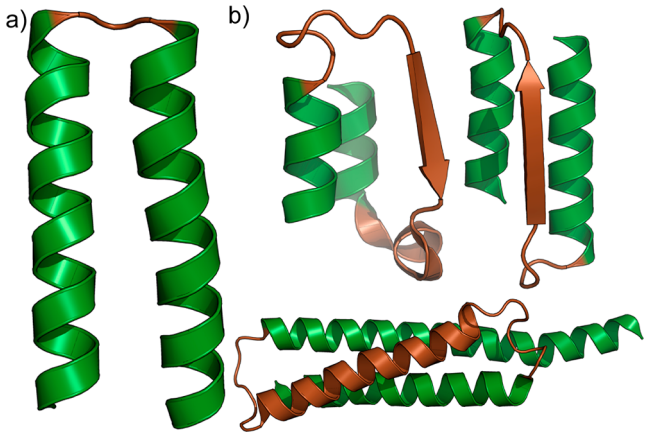


Figure 7. Common antiparallel (a) and parallel (b) helical dimer motifs feature linkers of a simple turn, a loop-strand-helix-loop, a loop-strand-loop, and a loop-helix-loop. PDB codes: 1I4Y, 2UYG, 2CBY, and 1OAH.

examples of a strand bracketed by two 4−5-residue loops. The two helices are linked by a loop-helix-loop in only four instances. The 55 parallel Case 2 dimers are more heterogeneous. Five possess the loop-strand-turn-helix-loop linker, five are connected by a loop-helix-loop, eight have a loop-strand-loop, but 11 contain a loop-helix-loop-helix-loop-helix-loop. (Specific PDBs for each motif are listed in the Supporting Information.) To our knowledge, this is the first effort to establish common protein folds connecting helices of defined orientation outside of canonical coiled coils.[52] Redesigned and optimized derivatives of these motifs—particularly the loop-strand-turn-helix-loop motif, as it is by far the most common and the most structurally interesting—may serve as miniprotein scaffolds.

Helical Dimers Typically Have Positive Net Charge. Given the importance of salt bridge interactions holding coiled coils together, we investigated the distribution of charge across

DOI: 10.1021/jacs.5b05527
J. Am. Chem. Soc. 2015, 137, 11622−11630

helix monomers in each Case. In Case 1, the average helix dimer has helices with net charges of 1.19 and −1.05, and 61% are net neutral or positive in total. In contrast, though Case 2 features dimers with net charges per helix of 0.89 and −0.33, and thus a higher average charge, 66% are net neutral or positive. Case 3 dimers are 70% net neutral or positive and have net charges for each helix of 0.88 and −0.08. These percentages suggest that the surface bound by such helical dimers may also be frequently negative or neutral. We calculated the surface charge on proteins bound by Case 1 dimers to be 70% net neutral or negative, with an average net charge of −0.75. Case 2 dimers bind surfaces that are 80% net neutral or negative with an average net charge of −0.68. Case 3 dimers bind surfaces that are 68% neutral or negative with an average charge of −0.61. The net positive charge in dimers is consistent with the higher number of positively charged protein helices in general.[55]

Helix Dimer Interfaces Mediate Fundamental Biological Processes. The interfaces collated in DippDB have the traits of prime targets for drug design. We categorized the functions of PPIs as defined in the PDB (Figure 8a) and observed that they are implicated in biological processes from enzymatic function to transcription to the immune response. We identified a total of 523 interactions (Cases 1 and 2) that would require a helix dimer mimetic or miniprotein for inhibition. Case 3 dimers, in contrast, feature a pair of single helices interacting with each other across an interface; thus, mimics of a single helix can disrupt these interactions. Intracellular PPIs dominate the dataset; thus, inhibition of these complexes will require development of synthetic analogues that can permeate the cell membrane. Our study reveals new classes of previously unidentified targets for helix dimer mimetics. Some of these newly identified targets will potentially aid efforts in drug discovery. In particular, it is interesting to note that the largest category, various enzymes, accounts for 63% of DIPP interactions. This category contains many hydrolases, oxidoreductases, and transferases, among other enzymes. Although enzyme function has typically been controlled using substrate or transition state analogues, helix dimers offer a potentially attractive alternative scaffold. Figure 8b,c highlights two examples of interactions involving UDP-galactose 4′-epimerase[53] and PCSK9[54] where helical dimers at dimeric interfaces in enzyme biological assemblies of considerable medical relevance bind near enzyme active sites.

Several targets in DippDB are critically relevant to cancer phenotypes.[56−58] In particular, we explored classic cases leading to "Hallmarks of Cancer" and discovered a set of possible targets mediated by helical dimers.[59,60] For example, three complexes—MHF histone tetramer/FANCM helicase (PDB code 4E45),[61] Mre11 nuclease/Rad50 ABC ATPase (PDB code 3QF7),[62] and the N- and C-terminal domains of the Mms21 subunit of Smc5 (PDB code 3HTK)[63]—have a role in DNA repair. The interaction between EPO and its receptor (PDB code 1EER) is implicated in the hypoxic response;[64] the complex between murine Ifnar1 and interferon-beta (PDB code 3WCY) is implicated in the regulation of apoptosis,[65,66] and the inhibition of the catalytic subunit p110β by the SH2 domain of the regulatory subunit p85β of phosphoinositide 3-kinase (PDB code 2Y3A) is implicated in angiogenesis[67] and invasive cell growth.[68] In each case, mimicry of *both* helices is predicted to be essential to optimize binding affinity. Beyond these targets, helical dimer interfaces include bacterial transcription and metabolism, quorum sensing, cell signaling, and more. A list of targets implicated in transcription is included in the Supporting Information, Table S1. We examined the 354 PDB structures that contain at least one Case 1 or Case 2 helical dimer and explored the Gene Ontology (GO) terms annotating each complex.[69] For example, two structures (PDB codes 2P5T and 1GVN) were toxins annotated with "cell killing", four possess "nucleic acid binding transcription factor activity", and seven

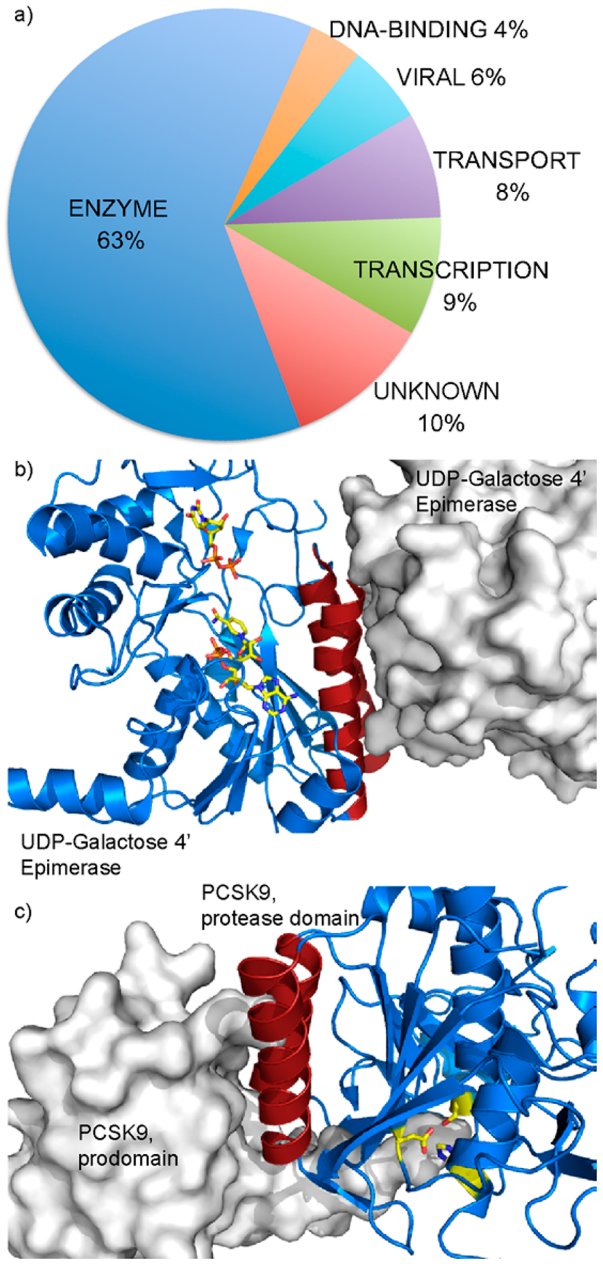


Figure 8. (a) Interfaces mediated by Case 1 and Case 2 helical dimers possess diverse functions and are dominated by enzyme complexes. (b) *T. brucei*'s UDP-galactose 4′-epimerase features a mutation in the active site relative to the human enzyme, potentially permitting specific targeting. Galactose metabolism is essential to the parasite's ability to cause African sleeping sickness, a neglected tropical disease.[53] The enzyme is active in dimeric and tetrameric forms; the target monomer is shown as gray surface. (c) PCSK9 propeptide (surface) binds with high affinity to PCSK9 (cartoon) and inhibits its proteolytic activity. This enzyme is linked to atherosclerosis and cardiovascular disease.[54] In both figures, the nearby active site from the dimer chain is highlighted in yellow sticks. PDB codes: 1GY8 and 2QTW.

DOI: 10.1021/jacs.5b05527
J. Am. Chem. Soc. 2015, 137, 11622−11630

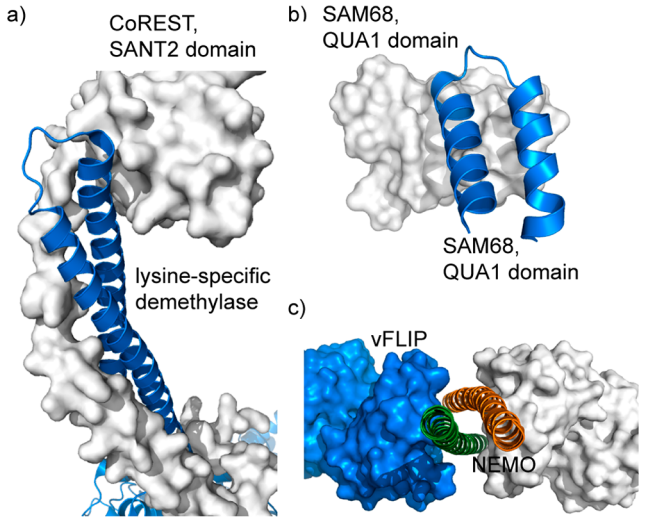


Figure 9. Examples of protein−protein interactions mediated by helix dimers. (a) LSD1-mediated nucleosome demethylation requires complexation with CoREST; existing inhibitors of LSD1 bind its substrate pocket, and predicted binding sites have thus far omitted the helical dimer interface.[73,74] (b) Sam68 RNA-binding protein is implicated in pro-oncogenic activity and modulates alternative splicing of CD44 and Bcl-xL. (c) The Kaposi's sarcoma protein vFLIP forms an A_2B_2 heterotetramer with NF-kappa-B essential modulator (NEMO) coiled coil; thus far, inhibition of this complex's function has only been achieved through geldanamycin inhibition of Hsp90.[75] PDB codes: 2IW5, 2XA6, and 3CL3.

are implicated in organismal development; two are involved in immune responses.

We anticipate that these and more targets will become tractable for modulation by designed tertiary structure mimetics; in Figure 9, we depict several interactions of particular pharmacological interest whose tertiary structure binding sites have not been drugged. These complexes, or the pathways they modulate, are known to be of therapeutic interest.[70−72] We illustrate them here to highlight the importance of helix dimer domains in coaxing the formation of these complexes and the potential of dimer mimics as inhibitors.

CONCLUSION

The aim of any systematic study of protein structures is both to uncover general principles governing protein geometry and to develop new insight into how to practically modulate protein function. Topologically defined segments often mediate protein−protein interactions,[11−15] and mimicry of these regions has emerged as a successful strategy for inhibitor design. Several examples of PPI inhibitors derived from mimicry of interfacial α-helical and β-strand domains have been described.[7,76−90] Emerging examples of tertiary mimetics as PPI inhibitors illustrate the broad potential of moving beyond secondary structure mimicry.[16,91] We were motivated to analyze protein complexes featuring helical dimers to create a list of potential targets where mimics of a single helix may not be sufficient.

Modification of the protocol used to develop a database of interface helices (HippDB)[10] provided a set of protein−protein interactions where the critical binding residues reside on two helices oriented in parallel or antiparallel configurations. The dataset includes some helical dimers that would be classified as true coiled coils because of their heptad repeats and supercoiling as well as helical dimers that find other means of making contacts with each other. We find that helical dimer complexes exhibit uncommon structural features. Some helical dimers violate the typical expectations for coiled coil interiors and are held together largely by salt bridges and hydrogen bonds, while others violate expectations for typical protein interfaces, which contain mostly large, aliphatic hot spot residues.

The length of dimers in contact with the protein partners spans one to three heptads, suggesting that medium-sized molecules or miniproteins will be able to disrupt these complexes. The online database provides a list of all entries in the dataset along with their PDB identifiers and the energetic contributions of the hot spot residues. We anticipate that this analysis will enable discovery of new classes of protein−protein interaction inhibitors as potential therapeutics.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b05527.

Protocol for the creation of DippDB, figures with detailed data, and tables of PDB complexes with particular features (PDF)

AUTHOR INFORMATION

Corresponding Author

*arora@nyu.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the National Institutes of Health (R01GM073943) for financial support. A.M.W. is supported by a Dean's Dissertation Fellowship. M.G.W. is grateful for the Sokol Predoctoral Fellowship from the NYU Chemistry Department.

REFERENCES

(1) Arkin, M. R.; Tang, Y.; Wells, J. A. *Chem. Biol.* **2014**, *21*, 1102.
(2) Wells, J. A.; McClendon, C. L. *Nature* **2007**, *450*, 1001.
(3) London, N.; Raveh, B.; Schueler-Furman, O. *Curr. Opin. Chem. Biol.* **2013**, *17*, 952.
(4) Raj, M.; Bullock, B. N.; Arora, P. S. *Bioorg. Med. Chem.* **2013**, *21*, 4051.
(5) Milroy, L.-G.; Grossmann, T. N.; Hennig, S.; Brunsveld, L.; Ottmann, C. *Chem. Rev.* **2014**, *114*, 4695.
(6) Ko, E.; Liu, J.; Burgess, K. *Chem. Soc. Rev.* **2011**, *40*, 4411.
(7) Azzarito, V.; Long, K.; Murphy, N. S.; Wilson, A. J. *Nat. Chem.* **2013**, *5*, 161.
(8) Chang, Y. S.; Graves, B.; Guerlavais, V.; Tovar, C.; Packman, K.; To, K. H.; Olson, K. A.; Kesavan, K.; Gangurde, P.; Mukherjee, A.; Baker, T.; Darlak, K.; Elkin, C.; Filipovic, Z.; Qureshi, F. Z.; Cai, H.; Berry, P.; Feyfant, E.; Shi, X. E.; Horstick, J.; Annis, D. A.; Manning, A. M.; Fotouhi, N.; Nash, H.; Vassilev, L. T.; Sawyer, T. K. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E3445.
(9) Watkins, A. M.; Arora, P. S. *Eur. J. Med. Chem.* **2015**, *94*, 480.
(10) Bergey, C. M.; Watkins, A. M.; Arora, P. S. *Bioinformatics* **2013**, *29*, 2806.
(11) Jochim, A. L.; Arora, P. S. *Mol. BioSyst.* **2009**, *5*, 924.
(12) Jochim, A. L.; Arora, P. S. *ACS Chem. Biol.* **2010**, *5*, 919.
(13) Bullock, B. N.; Jochim, A. L.; Arora, P. S. *J. Am. Chem. Soc.* **2011**, *133*, 14220.
(14) Watkins, A. M.; Arora, P. S. *ACS Chem. Biol.* **2014**, *9*, 1747.

DOI: 10.1021/jacs.5b05527
J. Am. Chem. Soc. 2015, 137, 11622−11630

(15) Gavenonis, J.; Sheneman, B. A.; Siegert, T. R.; Eshelman, M. R.; Kritzer, J. A. *Nat. Chem. Biol.* **2014**, *10*, 716.
(16) Checco, J. W.; Kreitler, D. F.; Thomas, N. C.; Belair, D. G.; Rettko, N. J.; Murphy, W. L.; Forest, K. T.; Gellman, S. H. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 4552.
(17) Wuo, M. G.; Mahon, A. B.; Arora, P. S. *J. Am. Chem. Soc.* **2015**, DOI: 10.1021/jacs.5b05525.
(18) Kariolis, M. S.; Kapur, S.; Cochran, J. R. *Curr. Opin. Biotechnol.* **2013**, *24*, 1072.
(19) Hosse, R. J.; Rothe, A.; Power, B. E. *Protein Sci.* **2006**, *15*, 14.
(20) Adams, J. J.; Sidhu, S. S. *Curr. Opin. Struct. Biol.* **2014**, *24*, 1.
(21) Gilbreth, R. N.; Koide, S. *Curr. Opin. Struct. Biol.* **2012**, *22*, 413.
(22) Lewis, R. J.; Garcia, M. L. *Nat. Rev. Drug Discovery* **2003**, *2*, 790.
(23) Crick, F. H. C. *Acta Crystallogr.* **1953**, *6*, 689.
(24) Lupas, A. N.; Gruber, M. *Adv. Protein Chem.* **2005**, *70*, 37.
(25) Strauss, H. M.; Keller, S. *Handb. Exp. Pharmacol.* **2008**, 461.
(26) Woolfson, D. N. *Adv. Protein Chem.* **2005**, *70*, 79.
(27) Burkhard, P.; Stetefeld, J.; Strelkov, S. V. *Trends Cell Biol.* **2001**, *11*, 82.
(28) Rose, A.; Meier, I. *Cell. Mol. Life Sci.* **2004**, *61*, 1996.
(29) Fong, J. H.; Keating, A. E.; Singh, M. *Genome Biol.* **2004**, *5*, R11.
(30) Walshaw, J.; Woolfson, D. N. *J. Mol. Biol.* **2001**, *307*, 1427.
(31) Berger, B.; Wilson, D. B.; Wolf, E.; Tonchev, T.; Milla, M.; Kim, P. S. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 8259.
(32) Kim, P. S.; Berger, B.; Wolf, E. *Protein Sci.* **1997**, *6*, 1179.
(33) Bohmann, D.; Tjian, R. *Cell* **1989**, *59*, 709.
(34) Barbara, K.; Willis, K.; Haley, T.; Deminoff, S.; Santangelo, G. *Mol. Genet. Genomics* **2007**, *278*, 135.
(35) Newman, J. R. S.; Keating, A. E. *Science* **2003**, *300*, 2097.
(36) Kortemme, T.; Kim, D. E.; Baker, D. *Sci. Signaling* **2004**, *2004*, pl2.
(37) Kortemme, T.; Baker, D. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14116.
(38) Johnson, L. M.; Horne, W. S.; Gellman, S. H. *J. Am. Chem. Soc.* **2011**, *133*, 10038.
(39) Su, J. Y.; Hodges, R. S.; Kay, C. M. *Biochemistry* **1994**, *33*, 15501.
(40) Engelman, D. M.; Zaccai, G. *Proc. Natl. Acad. Sci. U. S. A.* **1980**, *77*, 5894.
(41) Rees, D.; DeAntonio, L.; Eisenberg, D. *Science* **1989**, *245*, 510.
(42) Stevens, T. J.; Arkin, I. T. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 463.
(43) Rees, D. C.; Eisenberg, D. *Proteins: Struct., Funct., Genet.* **2000**, *38*, 121.
(44) Stevens, T. J.; Arkin, I. T. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 135.
(45) Akey, D. L.; Malashkevich, V. N.; Kim, P. S. *Biochemistry* **2001**, *40*, 6352.
(46) Lins, L.; Thomas, A.; Brasseur, R. *Protein Sci.* **2003**, *12*, 1406.
(47) Loladze, V. V.; Ermolenko, D. N.; Makhatadze, G. I. *J. Mol. Biol.* **2002**, *320*, 343.
(48) Joo, H.; Chavan, A. G.; Phan, J.; Day, R.; Tsai, J. *J. Mol. Biol.* **2012**, *419*, 234.
(49) Kohn, W. D.; Mant, C. T.; Hodges, R. S. *J. Biol. Chem.* **1997**, *272*, 2583.
(50) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. *Annu. Rev. Biochem.* **1999**, *68*, 779.
(51) Giri Rao, V. V.; Gosavi, S. *PLoS Comput. Biol.* **2014**, *10*, e1003938.
(52) Koga, N.; Tatsumi-Koga, R.; Liu, G.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D. *Nature* **2012**, *491*, 222.
(53) Urbaniak, M. D.; Turnock, D. C.; Ferguson, M. A. *Eukaryotic Cell* **2006**, *5*, 1906.
(54) Fan, D.; Yancey, P. G.; Qiu, S.; Ding, L.; Weeber, E. J.; Linton, M. F.; Fazio, S. *Biochemistry* **2008**, *47*, 1631.
(55) Sharadadevi, A.; Sivakamasundari, C.; Nagaraj, R. *Proteins: Struct., Funct., Genet.* **2005**, *59*, 791.
(56) Kahl, P.; Gullotti, L.; Heukamp, L. C.; Wolf, S.; Friedrichs, N.; Vorreuther, R.; Solleder, G.; Bastian, P. J.; Ellinger, J.; Metzger, E.; Schule, R.; Buettner, R. *Cancer Res.* **2006**, *66*, 11341.
(57) Lim, S.; Janzer, A.; Becker, A.; Zimmer, A.; Schule, R.; Buettner, R.; Kirfel, J. *Carcinogenesis* **2010**, *31*, 512.
(58) Lukong, K. E.; Richard, S. *Future Oncol.* **2007**, *3*, 539.
(59) Hanahan, D.; Weinberg, R. A. *Cell* **2011**, *144*, 646.
(60) Hanahan, D.; Weinberg, R. A. *Cell* **2000**, *100*, 57.
(61) Fox, D., 3rd; Yan, Z.; Ling, C.; Zhao, Y.; Lee, D. Y.; Fukagawa, T.; Yang, W.; Wang, W. *Cell Res.* **2014**, *24*, 560.
(62) Lammens, K.; Bemeleit, D. J.; Mockel, C.; Clausing, E.; Schele, A.; Hartung, S.; Schiller, C. B.; Lucas, M.; Angermuller, C.; Soding, J.; Strasser, K.; Hopfner, K. P. *Cell* **2011**, *145*, 54.
(63) Duan, X.; Sarangi, P.; Liu, X.; Rangi, G. K.; Zhao, X.; Ye, H. *Mol. Cell* **2009**, *35*, 657.
(64) Scholz, H.; Schurek, H. J.; Eckardt, K. U.; Bauer, C. *Experientia* **1990**, *46*, 1197.
(65) Tanaka, N.; Sato, M.; Lamphier, M. S.; Nozawa, H.; Oda, E.; Noguchi, S.; Schreiber, R. D.; Tsujimoto, Y.; Taniguchi, T. *Genes Cells* **1998**, *3*, 29.
(66) de Weerd, N. A.; Vivian, J. P.; Nguyen, T. K.; Mangan, N. E.; Gould, J. A.; Braniff, S. J.; Zaker-Tabrizi, L.; Fung, K. Y.; Forster, S. C.; Beddoe, T.; Reid, H. H.; Rossjohn, J.; Hertzog, P. J. *Nat. Immunol.* **2013**, *14*, 901.
(67) Zhang, X.; Vadas, O.; Perisic, O.; Anderson, K. E.; Clark, J.; Hawkins, P. T.; Stephens, L. R.; Williams, R. L. *Mol. Cell* **2011**, *41*, 567.
(68) Czauderna, F.; Fechtner, M.; Aygun, H.; Arnold, W.; Klippel, A.; Giese, K.; Kaufmann, J. *Nucleic Acids Res.* **2003**, *31*, 670.
(69) Harris, M. A.; Clark, J.; Ireland, A.; Lomax, J.; Ashburner, M.; Foulger, R.; Eilbeck, K.; Lewis, S.; Marshall, B.; Mungall, C.; Richter, J.; Rubin, G. M.; Blake, J. A.; Bult, C.; Dolan, M.; Drabkin, H.; Eppig, J. T.; Hill, D. P.; Ni, L.; Ringwald, M.; Balakrishnan, R.; Cherry, J. M.; Christie, K. R.; Costanzo, M. C.; Dwight, S. S.; Engel, S.; Fisk, D. G.; Hirschman, J. E.; Hong, E. L.; Nash, R. S.; Sethuraman, A.; Theesfeld, C. L.; Botstein, D.; Dolinski, K.; Feierbach, B.; Berardini, T.; Mundodi, S.; Rhee, S. Y.; Apweiler, R.; Barrell, D.; Camon, E.; Dimmer, E.; Lee, V.; Chisholm, R.; Gaudet, P.; Kibbe, W.; Kishore, R.; Schwarz, E. M.; Sternberg, P.; Gwinn, M.; Hannick, L.; Wortman, J.; Berriman, M.; Wood, V.; de la Cruz, N.; Tonellato, P.; Jaiswal, P.; Seigfried, T.; White, R.; Consortium, G. O. *Nucleic Acids Res.* **2004**, *32*, D258.
(70) Mould, D. P.; McGonagle, A. E.; Wiseman, D. H.; Williams, E. L.; Jordan, A. M. *Med. Res. Rev.* **2015**, *35*, 586.
(71) Wang, Q.; Li, Y.; Zhou, J.; Liu, J.; Qin, J.; Xing, F.; Zhang, J.; Cheng, J. *Tumor Biol.* **2015**, *36*, 4509.
(72) Habineza Ndikuyeze, G.; Gaurnier-Hausser, A.; Patel, R.; Baldwin, A. S.; May, M. J.; Flood, P.; Krick, E.; Propert, K. J.; Mason, N. J. *PLoS One* **2014**, *9*, e95404.
(73) Robertson, J. C.; Hurley, N. C.; Tortorici, M.; Ciossani, G.; Borrello, M. T.; Vellore, N. A.; Ganesan, A.; Mattevi, A.; Baron, R. *PLoS Comput. Biol.* **2013**, *9*, e1003158.
(74) Wang, J.; Lu, F.; Ren, Q.; Sun, H.; Xu, Z.; Lan, R.; Liu, Y.; Ward, D.; Quan, J.; Ye, T.; Zhang, H. *Cancer Res.* **2011**, *71*, 7238.
(75) Field, N.; Low, W.; Daniels, M.; Howell, S.; Daviet, L.; Boshoff, C.; Collins, M. *J. Cell Sci.* **2003**, *116*, 3721.
(76) Johnson, L. M.; Barrick, S.; Hager, M. V.; McFedries, A.; Homan, E. A.; Rabaglia, M. E.; Keller, M. P.; Attie, A. D.; Saghatelian, A.; Bisello, A.; Gellman, S. H. *J. Am. Chem. Soc.* **2014**, *136*, 12848.
(77) Boersma, M. D.; Haase, H. S.; Peterson-Kaufman, K. J.; Lee, E. F.; Clarke, O. B.; Colman, P. M.; Smith, B. J.; Horne, W. S.; Fairlie, W. D.; Gellman, S. H. *J. Am. Chem. Soc.* **2012**, *134*, 315.
(78) Horne, W. S.; Johnson, L. M.; Ketas, T. J.; Klasse, P. J.; Lu, M.; Moore, J. P.; Gellman, S. H. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 14751.
(79) Harrison, R. S.; Shepherd, N. E.; Hoang, H. N.; Ruiz-Gomez, G.; Hill, T. A.; Driver, R. W.; Desai, V. S.; Young, P. R.; Abbenante, G.; Fairlie, D. P. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 11686.
(80) Sinclair, J. K. L.; Denton, E. V.; Schepartz, A. *J. Am. Chem. Soc.* **2014**, *136*, 11232.

(81) Lao, B. B.; Grishagin, I.; Mesallati, H.; Brewer, T. F.; Olenyuk, B. Z.; Arora, P. S. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 7531.
(82) Lao, B. B.; Drew, K.; Guarracino, D. A.; Brewer, T. F.; Heindel, D. W.; Bonneau, R.; Arora, P. S. *J. Am. Chem. Soc.* **2014**, *136*, 7877.
(83) Kushal, S.; Lao, B. B.; Henchey, L. K.; Dubey, R.; Mesallati, H.; Traaseth, N. J.; Olenyuk, B. Z.; Arora, P. S. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 15602.
(84) Walensky, L. D.; Bird, G. H. *J. Med. Chem.* **2014**, *57*, 6275.
(85) Verdine, G. L.; Hilinski, G. J. *Methods Enzymol.* **2012**, *503*, 3.
(86) Burslem, G. M.; Kyle, H. F.; Breeze, A. L.; Edwards, T. A.; Nelson, A.; Warriner, S. L.; Wilson, A. J. *ChemBioChem* **2014**, *15*, 1083.
(87) Cheng, P.-N.; Liu, C.; Zhao, M.; Eisenberg, D.; Nowick, J. S. *Nat. Chem.* **2012**, *4*, 927.
(88) Hammond, M. C.; Harris, B. Z.; Lim, W. A.; Bartlett, P. A. *Chem. Biol.* **2006**, *13*, 1247.
(89) Zutshi, R.; Franciskovich, J.; Shultz, M.; Schweitzer, B.; Bishop, P.; Wilson, M.; Chmielewski, J. *J. Am. Chem. Soc.* **1997**, *119*, 4841.
(90) Huang, Z.; Zhang, M.; Burton, S. D.; Katsakhyan, L. N.; Ji, H. *ACS Chem. Biol.* **2014**, *9*, 193.
(91) Reinert, Z. E.; Lengyel, G. A.; Horne, W. S. *J. Am. Chem. Soc.* **2013**, *135*, 12528.

DOI: 10.1021/jacs.5b05527
J. Am. Chem. Soc. 2015, 137, 11622−11630