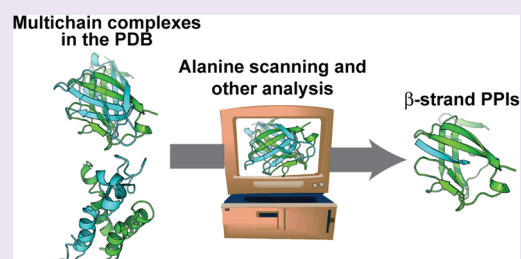


# Anatomy of $\beta$ -Strands at Protein–Protein Interfaces

Andrew M. Watkins and Paramjit S. Arora\*

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

**ABSTRACT:** The development of inhibitors for protein–protein interactions frequently involves the mimicry of secondary structure motifs. While helical protein–protein interactions have been heavily targeted, a similar level of success for the inhibition of  $\beta$ -strand and  $\beta$ -sheet rich interfaces has been elusive. We describe an assessment of the full range of  $\beta$ -strand interfaces whose high-resolution structures are available in the Protein Data Bank. This analysis identifies complexes where a  $\beta$ -strand or  $\beta$ -sheet contributes significantly to binding. The results highlight the molecular recognition complexity in strand-mediated interactions relative to helical interfaces and offer guidelines for the construction of  $\beta$ -strand and  $\beta$ -sheet mimics as ligands for protein receptors. The online data set will potentially serve as an entry-point to new classes of protein–protein interaction inhibitors.



Protein–protein interactions (PPIs) regulate a wide array of cellular processes and are attractive targets for drug design.<sup>1–3</sup> Successful approaches to the design of PPI inhibitors include high throughput screening of compound libraries, fragment-based screening, and the mimicry of interfacial protein segments that promote complex formation.<sup>4–6</sup> Rational design of protein domain mimetics requires access to high-resolution structures of protein complexes and understanding of the features of protein interfaces.<sup>5,7–9</sup> Mimicry of helical domains has successfully yielded inhibitors of intractable protein–protein interactions.<sup>4,10–13</sup> This success has required knowledge of the helical regions key to the interactions as well as the design of synthetic scaffolds that can mimic the attributes of protein helices. To guide the design of helix mimics, we previously described analysis of protein complexes whose high-resolution structures have been deposited in the Protein Data Bank and cataloged interfacial helices that feature clusters of residues that contribute significantly to binding.<sup>14–17</sup> The hot spot residues were characterized using computational alanine scanning mutagenesis analysis.<sup>18–20</sup> Here, we describe a similar effort to decode protein interfaces that contain  $\beta$ -strand and  $\beta$ -sheet segments. To our knowledge, one similar effort to identify  $\beta$ -sheet interfaces has been reported.<sup>21</sup> In this survey, Nowick, Baldi, and co-workers created a database of  $\beta$ -strands that form sheets with  $\beta$ -strands from another protein chain.

We identified roughly 15 000  $\beta$ -strand motifs that make valuable contributions to the overall stability of the complex; mimicry of these strands will potentially lead to potent inhibitors. Our analysis shows that, as with  $\alpha$ -helical interfaces, aromatic and hydrophobic hot spots are critical for strand-mediated protein–protein interactions.<sup>16</sup> Backbone hydrogen bonding, which is utilized by strands but not  $\alpha$ -helices for recognition, introduces unsystematic diversity to  $\beta$ -strand binding interactions.<sup>22</sup> Single  $\beta$ -strands can be localized in nonenzymatic protein pockets much as is seen in canonical enzyme–substrate complexes.<sup>23</sup> Strands participating in PPIs

can also exist as part of a  $\beta$ -sheet where the partner protein recognizes side chains on one or multiple strands.  $\beta$ -strands may utilize side chain functionality on one face or both faces of a single strand for interaction with the partner protein. Lastly,  $\beta$ -strands may form hydrogen-bonding interactions with the partner or they may interact exclusively with side chain functionality. Our results offer impetus for the construction of synthetic strand scaffolds but suggest that multiple design strategies will be required to match the diversity present in  $\beta$ -strand-mediated protein–protein interactions.<sup>24</sup>

## RESULTS AND DISCUSSION

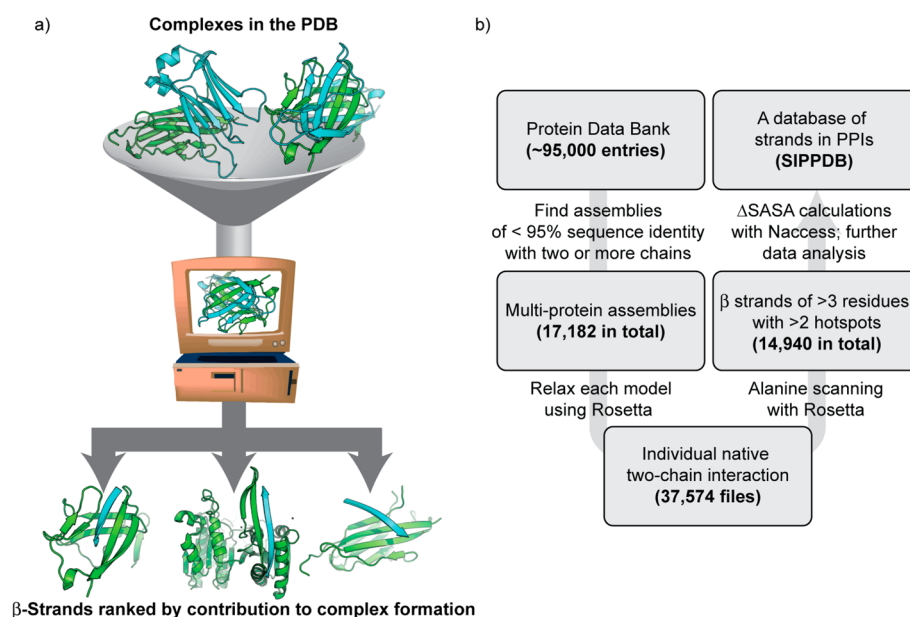
Peptidomimetics have a rich history as  $\beta$ -strand mimetics to inhibit enzyme activity.<sup>24,25</sup> For example, the HIV-1 protease has been successfully targeted with a diverse set of clinically useful strand mimics.<sup>24,26,27</sup> Similarly, synthetic strand and sheet mimics have been described as modulators of protein aggregation.<sup>28</sup> Cell surface  $\beta$ -strand and  $\beta$ -sheet proteins have been targeted by macrocyclic peptides and synthetic antibodies. AP33 is a neutralizing antibody that binds a  $\beta$ -hairpin peptide epitope on the E2 envelope glycoprotein of hepatitis C virus, pertuzumab binds the receptor tyrosine kinase ErbB2, and cetuximab binds to EGFR's extracellular domain.<sup>29–31</sup> Grb2's SH2 domain and the E3 ubiquitin ligase E6AP have both been the subjects of peptide or peptidomimetic macrocycle development.<sup>32–34</sup> Many successful efforts to develop  $\beta$ -strand,  $\beta$ -hairpin, and  $\beta$ -sheet mimetic scaffolds have been undertaken,<sup>22,24,35–45</sup> although the applications of these scaffolds to the disruption of PPIs remains limited.<sup>28,46–49</sup>

Here, we present a comprehensive analysis of  $\beta$ -strands in PPIs to construct a list of suitable targets. Our studies are centered on the identification of clusters of interfacial residues

Received: April 1, 2014

Accepted: May 28, 2014

Published: May 28, 2014



**Figure 1.** (a) Schematic for identification of protein interfaces in the Protein Data Bank (PDB) where a  $\beta$ -strand contributes significantly to complex formation. (b) Protocol for extracting  $\beta$ -strand complexes from multichain entries in the PDB. Biological assemblies from the PDB were analyzed using computational alanine scanning mutagenesis to create the data set hosted at [www.nyu.edu/projects/arora/sippdb/](http://www.nyu.edu/projects/arora/sippdb/). The data set allows division of  $\beta$ -strand interfaces into high-affinity and low-affinity interactions.

that contribute to  $\Delta G_{\text{binding}}$  with the hypothesis that mimicry of the disposition and orientation of these hot spot residues will lead to successful inhibitors.<sup>15,16</sup> Several methods to evaluate the relative importance of different residues to a binding interaction have been described, with alanine scanning mutagenesis<sup>20</sup> and  $\Delta$ SASA<sup>50–52</sup> calculations the most commonly employed. Alanine scanning consists of the serial generation (either computational or experimental) of alanine point mutants; point mutants leading to a large change in binding energy upon mutation to alanine ( $\Delta\Delta G$ ) are likely to be important wild-type residues.<sup>18,19,53</sup> Commonly, residues whose mutation to alanine results in a decrease in binding energy of  $\Delta\Delta G \geq 1$  kcal/mol are designated hot spot residues. In contrast,  $\Delta$ SASA describes the amount of solvent-accessible surface area that is buried by the residue in question upon binding; the more surface area buried, the greater the entropy decrease upon binding due to desolvation. For this study, much as for our recent work on helical interfaces,<sup>17</sup> both  $\Delta\Delta G$  and  $\Delta$ SASA are evaluated, though  $\Delta\Delta G$  alone was used to determine if a strand interacts strongly with a binding partner.<sup>54</sup>

**Procedure for Identifying  $\beta$ -Strand Interfaces in Protein–Protein Interactions.** An overview of the approach is depicted in Figure 1. Crystal and NMR structures of protein complexes were obtained from the Protein Data Bank.<sup>55</sup> For crystal structures, the multimodel biological assembly files were acquired. Each individual model was refined using the Rosetta “relax” protocol, which involves iterations of all-atom minimization with restraints followed by side-chain repacking. Following this procedure, the best-scoring model was retained for further analysis. Each pair of protein chains was extracted and separately processed by computational alanine scanning using RosettaScripts<sup>56</sup> and by the  $\Delta$ SASA calculation utility NACCESS.<sup>57</sup> The secondary structure content of the chains in question was evaluated using the Dictionary of Secondary Structure Prediction (DSSP).<sup>58</sup> Three or more consecutive residues with a hydrogen bonding pattern characteristic of  $\beta$ -strands, assigned “E” by DSSP, were considered to be  $\beta$ -strands

for the purpose of this study. The online data set contains a listing of all high affinity  $\beta$ -strand complexes. For each high affinity strand entry, we recorded the total, per-residue average, and percent contribution of both  $\Delta\Delta G$  and  $\Delta$ SASA. We also recorded parameters such as length, distance between terminal hot spots, and the relative hot spot positions. The “Mimetic-Score” represents the sum of the three highest- $\Delta\Delta G$  hot spot residues, and may be used to compare entries in the entire data set.

**Classification of Interfacial  $\beta$ -Strands.** We classified interfacial strands as *strongly interacting* or *high affinity  $\beta$ -strands* if they contained two or more hot spot residues. Hot spot residues are defined as residues whose mutation to alanine results in a concomitant reduction in the binding affinity of the complex by  $\Delta\Delta G \geq 1$  Rosetta Energy Unit (or REU). Apparent “hot spots” of glycine and proline were omitted from consideration, since such point mutations likely lead to nonlocal effects that are poorly modeled by a fixed-backbone protocol; in particular, glycine to alanine mutations can result in clashes only present in the bound complex that could be relieved by backbone perturbations.  $\Delta\Delta G$  values exceeding 8.0 REU were reduced to 8.0. We obtained an initial data set of 17,182 multiprotein entries, clustered to ensure <95% sequence identity, from the PDB. These entries resulted in 37 574 pairs of two-chain complexes; using these criteria, 14 940 high affinity  $\beta$ -strands were found. Roughly 38 000 interface  $\beta$ -strands feature zero or one hot spot residues and are classified as *weakly interacting  $\beta$ -strands*.

The average length of  $\beta$ -strands in protein complexes that are part of the PDB is five residues, and each strand averages at least one hot spot residue (Table 1). Comparison of the strongly and weakly interacting  $\beta$ -strands suggests that high affinity  $\beta$ -strands are critical to the overall stability of the complex;  $\beta$ -strands in these interfaces contribute 22% of the overall binding interactions, which is significant given the low number of interacting residues. The strongly interacting strands feature 2.5 hot spot residues with a high average per residue

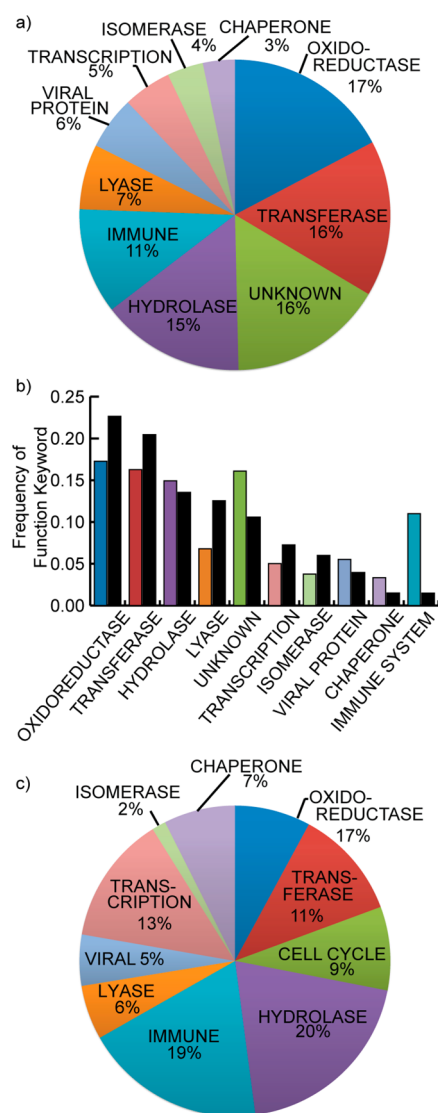
**Table 1. Comparison of Statistics for Strongly and Weakly Interacting Interfacial  $\beta$ -Strands**

	strongly interacting $\beta$ -strands (14 940)	weakly interacting $\beta$ -strands (38 027)	all interfacial $\beta$ -strands in PPIs (52 967)
interface strand length, residues	5.9 $\pm$ 2.3	4.4 $\pm$ 1.6	4.8 $\pm$ 2.0
no. of hot spot residues	2.5 $\pm$ 0.87	0.53 $\pm$ 0.49	1.1 $\pm$ 1.1
per-residue $\Delta\Delta G$ , REU	1.2 $\pm$ 0.65	0.47 $\pm$ 0.40	0.67 $\pm$ 0.57
per-residue $\Delta SASA$ , $\text{\AA}^2$	-41 $\pm$ 20	-22 $\pm$ 16	-27 $\pm$ 19
hot spot residue $\Delta\Delta G$ , REU	2.4 $\pm$ 1.6	2.0 $\pm$ 0.96	2.1 $\pm$ 1.1
% contribution to $\Delta\Delta G_{\text{complex}}$	22 $\pm$ 15	9.7 $\pm$ 12	13 $\pm$ 14
% contribution to $\Delta SASA_{\text{complex}}$	17 $\pm$ 13	9.4 $\pm$ 11	11 $\pm$ 12
distance between first and last hot spot residue, $\text{\AA}$	3.9 $\pm$ 1.8	n/a	n/a

$\Delta\Delta G$  of 2.4 REU, implying that mimicry of these strands will lead to potent inhibitors of protein–protein interactions. Because backbone hydrogen bonds are a critical component of  $\beta$ -strand complexes,<sup>59</sup> the alanine scanning  $\Delta\Delta G$  values likely underrate the overall interaction energy but remain a good measure for assessing the *sequence-specific* contribution of each residue. Indeed, seminal work on  $\beta$ -sheet interactions, in which both strands are hydrogen bonded, has included *in vitro* alanine scanning and has found substantial sequence-specific effects.<sup>60,61</sup> We have compiled the results in an online database hosted at [www.nyu.edu/projects/arora/sippdb/](http://www.nyu.edu/projects/arora/sippdb/). The database provides descriptions of each high affinity strand, including number of hot spot residues and per residue  $\Delta\Delta G$ , with links to the PDB entry for each strand.

**Functional Distribution of  $\beta$ -Strands in All Multiprotein Complexes.**  $\beta$ -Strand interactions are often associated with enzymes because protease substrates adopt extended conformations.<sup>23</sup> In the construction of this database, we removed all enzyme/ $\beta$ -strand substrate interactions to have the results reflect participation of strands mediating protein–protein interactions.  $\beta$ -Strand interfaces in fact participate widely in protein–protein interactions that involve enzyme partners where the strand is *not* manipulated by the enzyme action. Only such  $\beta$ -strand interactions are included in our data set.  $\beta$ -strand interfaces are also involved in a diverse set of other biological processes, including transcriptional regulation and protein folding (Figure 2a). Comparison of  $\beta$ -strand and helix mediated interactions reveals that enzymes such as oxidoreductases, transferases, hydrolases, and lyases amount to 55% of  $\beta$ -strand interactions and 67% of  $\alpha$ -helix mediated interactions, suggesting that protein–protein interactions represented in the PDB broadly involve at least one enzyme partner. This finding indicates that  $\beta$ -strand and helix mimetics may both serve as inhibitors of enzymatic function not by direct targeting of the catalytic pocket but by interrupting signal transduction pathways mediated by certain enzymes.

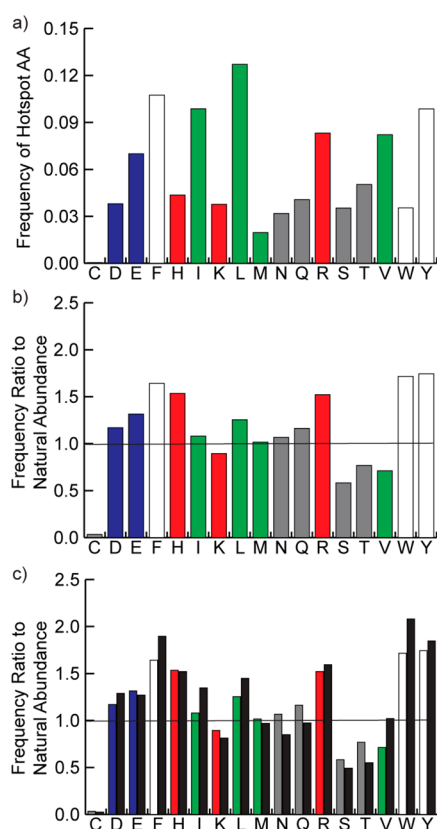
We compared the functional distribution of  $\beta$ -strands in all multiprotein complexes in the PDB to a subset consisting of only heterodimeric  $\beta$ -strand interfaces to determine if different classes of strand-mediated complexes (i.e., homodimers and heterodimers) diverge in function. We considered a set of 2330



**Figure 2.** (a) Analysis of high affinity  $\beta$ -strand interfaces shows that these complexes participate in a wide-range of biological functions, but protein–protein interactions where one partner is an enzyme are the dominant category. (b) Comparison of the functions associated with the helix and strand data sets (helix in black; strand in colors) shows that the distribution of functions is similar across these two classes of secondary structure-mediated PPIs, with strands dominating interactions related to the immune system. (c) The functional distribution of strand-mediated nonimmunoglobulin heterodimers shows that PPIs with enzymatic function play a prominent role in different subsets of protein complexes, but other categories such as transcription are better represented in heterodimers without the influence of immunoglobulin dimers.

heterodimers and further eliminated entries whose protein names contained “heavy chain” and “light chain” to diminish the influence of immunoglobulin interactions, in which  $\beta$ -sheets are dominant secondary structures. The pruned data set contains 1221 complexes. Comparisons of functions associated with all biological assemblies (Figure 2a) and heterodimeric complexes only (Figure 2c) indicates that enzymatic complexes dominate  $\beta$ -strand PPIs across different types of complexes, but categories such as transcriptional complexes are better represented in the pruned set of heterodimers.

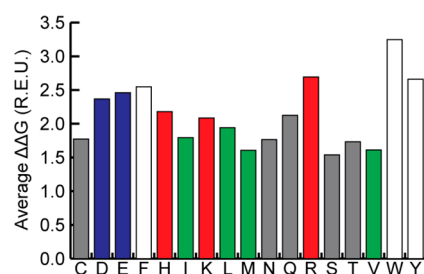
**Contribution of Individual Residues to  $\beta$ -Strand Interactions.** We analyzed the high affinity strand interfaces to assess the relative contribution of individual residues to protein complex formation (Figure 3). Hydrophobic and



**Figure 3.** (a) Distribution of hot spot residues in high affinity  $\beta$ -strands. (b) The data from part a normalized to the natural abundance in strands of each amino acid. (c) The normalized distribution of strand hot spots compared to the normalized distribution of helical hot spots (black bars).

aromatic residues dominate the binding energy landscape, accounting for roughly 40% of the hot spot residues (Figure 3a). When normalized for natural abundance,<sup>62</sup> we find that nonpolar aromatic residues, histidine, and arginine are overrepresented as hot spots at strand interfaces in comparison to polar residues (Figure 3b). These results correlate with the types of amino acids appearing as hot spot residues in protein interfaces,<sup>53,63,64</sup> although our data set is considerably larger than those previously examined. The normalized distribution for  $\beta$ -strand hot spots can be compared to those of  $\alpha$ -helices (Figure 3c). Interestingly, the energetic contribution of individual residues to helical or strand interfaces is roughly the same for individual amino acids. The comparison suggests that complexation of strand and helical interfaces share similar side chain recognition principles.

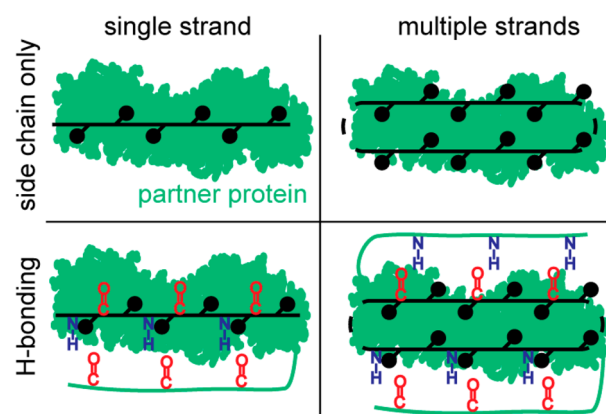
We analyzed the relative contribution of each hot spot amino acid to  $\beta$ -strand binding interactions (Figure 4). The relative binding contribution ( $\Delta\Delta G$ ) of different hot spot amino acids correlates well with the normalized occurrence shown in Figure 3b but not with the raw data in Figure 3a. While leucine is the most prevalent hot spot residue in  $\beta$ -strands, its contribution to binding is less significant when it does appear. In contrast, the aromatic residues and arginine are both more prevalent as hot



**Figure 4.** Average  $\Delta\Delta G$  of different hot spot amino acids found in strongly interacting  $\beta$ -strands. Aromatic residues and arginine make for the strongest interactions, while small polar and aliphatic residues are the weakest.

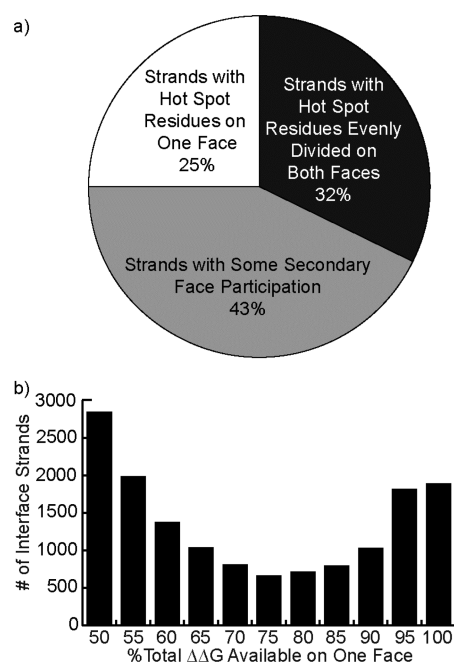
spots compared to their natural abundance, and they serve as strong hot spot residues when they do appear.

**Geometrical Diversity in  $\beta$ -Strand Interfaces.**  $\beta$ -Strand-mediated interfaces are geometrically diverse,<sup>22</sup> especially in comparison to the order presented by helical PPIs, which primarily differ in the number of helical faces involved in binding interactions.<sup>16</sup> The contribution of backbone hydrogen bonding to strand interactions, but not helical motifs, is an important factor that enhances diversity of the recognition motifs available to strands (Figure 5).



**Figure 5.** Schematic representing simplified  $\beta$ -strand interactions with a complementary protein receptor shown as green surface.  $\beta$ -Strands utilize a diverse range of binding strategies; for example,  $\beta$ -strands may interact alone or as part of a  $\beta$ -sheet of two or more strands, and  $\beta$ -strands may form hydrogen bonding interactions with the partner or they may interact exclusively with the side chain functionality. Additionally, a  $\beta$ -strand may interact with the side chains of only one face, or it may employ both faces for interaction with the partner protein.

To categorize this diversity, we began by examining the binding energetics of  $\beta$ -strands on different faces to determine if hot spots are largely featured on one face, as with  $\alpha$ -helices. The data reveals that high affinity  $\beta$ -strands can be divided into three categories (Figure 6): those with two faces contributing equally to the interaction, those with one face that contributes almost all of the  $\Delta\Delta G$  (90% or more), and those where both faces contribute unequally but significantly to the interaction. The results indicate that only a quarter of the  $\beta$ -strands employ only one face for interactions. Although hot spot residues are typically unequally divided on the two strand faces, the analysis suggests that synthetic scaffolds would typically need to mimic both faces to be broadly useful as inhibitors of PPIs.



**Figure 6.** Hot spot residues are unevenly distributed among the two faces of a  $\beta$ -strand, with the higher affinity face on average controlling 70% of the total  $\Delta\Delta G$ . (a) A pie chart describing three approximate classifications of  $\beta$ -strands. (b) A histogram of the percentage of the total  $\Delta\Delta G$  associated with the stronger-interacting face, providing a higher-resolution account of the data summarized by the pie chart.

We evaluated the different types of interactions  $\beta$ -strands participate in with binding partners in the context of transcriptional complexes. Manually filtering for unique strands with three or more hot spot residues resulted in a set of 172 entries describing 133 distinct protein–protein interactions. Two-thirds of these entries described a  $\beta$ -sheet-formation interaction, 95% of which were antiparallel in orientation. Across the PDB, antiparallel  $\beta$ -sheet orientations are more prevalent, on the order of 3:1, as compared to parallel sheets.<sup>65</sup> Antiparallel arrangements are thought to possess better hydrogen bonding geometry<sup>66</sup> and overall energetics.<sup>67</sup> The greater prevalence of antiparallel strand orientations in protein–protein interactions may reflect the more stringent energetic criteria for quaternary versus tertiary structure association.

In these  $\beta$ -sheet interactions, strands made an average of ten hydrogen bond acceptors and donors available to the partner protein, with two-thirds of the carbonyls and amide N–H's participating in hydrogen bonding. Hydrogen bonding potential between partner carbonyl C=O and amide N–H groups was judged using the standard distance and angle criteria: a hydrogen–oxygen distance less than 2.7 Å (i.e., their van der Waals radii are in contact) and a N–H–O angle greater than 110°. In contrast, hydrogen bonding potential was not fulfilled in interface strands that do not form sheets, with only 19% of the exposed hydrogen bond donors and acceptors forming hydrogen bonds. Multivalent interactions are also common in this set, with roughly two-thirds of the interfaces using more than one high affinity strand for binding interactions. This result suggests that hairpin and  $\beta$ -sheet mimetics will have a substantial role as protein–protein interaction inhibitors.<sup>22</sup>

**Survey of the Data Set.** Table 2 depicts five representative examples present in the database that display the diversity of  $\beta$ -strand interfaces. For each entry, a cartoon representation of the complex with the strand(s) highlighted in magenta is provided alongside a representation of the strands in sticks and the partner protein in cartoon with pertinent side-chains displayed in sticks. Entry 1 depicts a high affinity  $\beta$ -strand found in the ternary complex between IKB $\beta$  and NFkappaB p65 homodimer, at the p65 homodimer interface. Except for a carbonyl-histidine hydrogen bond, the entire interaction is mediated by side-chain packing interactions and a glutamate-arginine charged interaction. Entry 2 depicts the bacillus transcriptional regulator C-125, as an example of a  $\beta$ -strand that forms a sheet at the interface. Entry 3 depicts part of Bcl6 corepressor making a unique interaction with RING finger protein 1. Exactly one face of hydrogen bonds (alternating amides and carbonyls) is occupied by a strand in the partner protein. Meanwhile, one face of the strand (an F, F, and L) pack in a hydrophobic groove formed by a helix of RING, while the other face (M, E, S) interact with one face of the aforementioned partner strand. Bcl6 contains an additional high affinity strand, antiparallel to the one depicted, which does not form hydrogen bonds with the partner; the two may be mimicked with a  $\beta$ -hairpin. Entry 4 shows a  $\beta$ -hairpin found in the Rap30/Rap74 complex of human transcription initiation factor IIF. Entry 5 shows a pair of nonhairpin strands from the HTH-type transcriptional regulator LRPC, as an example of a complex in which one chain possesses multiple high affinity strands without that do not participate in a hairpin relationship.

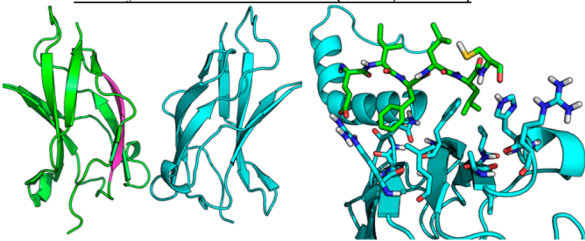
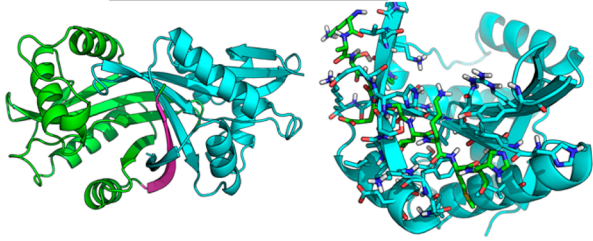
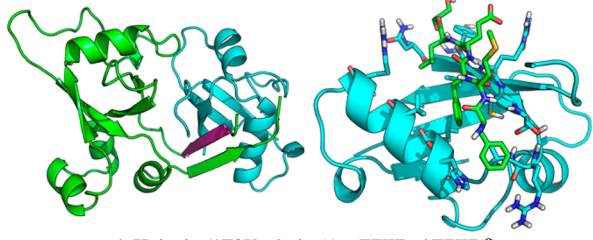
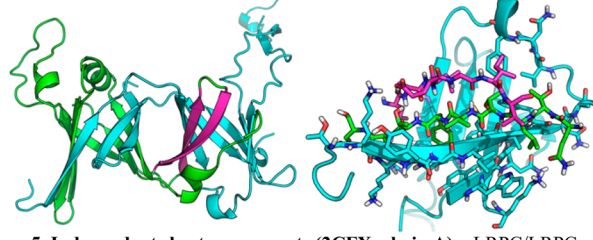
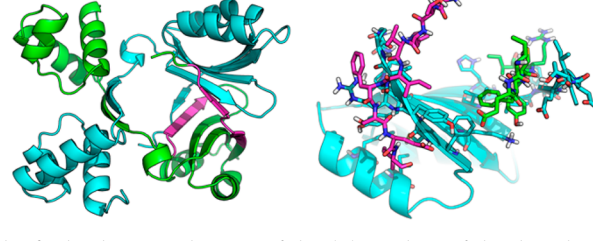
In summary, we have cataloged  $\beta$ -strand interfaces whose structures are available in the Protein Data Bank according to the relative contribution of individual strands to complex formation. Our evaluations suggest that there are roughly 15 000 high affinity  $\beta$ -strands mediated protein–protein interactions. Mimics of these strands would lead to potent inhibitors as has been illustrated for helical PPIs; however, the variety and complexity of  $\beta$ -strand-based interface structures is staggering and suggests that a range of synthetic approaches will be required for inhibition. A large proportion of  $\beta$ -strands participate in sheet interactions, which points to an important role for  $\beta$ -hairpin and  $\beta$ -sheet mimics.

## METHODS

Given a PDB file with <95% sequence homology to those already in the database, the following general procedure was employed. First, the PDB was “cleaned,” removing HETATM entries and any crystal water. Then, the PDB was separated into its distinct MODEL entries. One to four models are common for biological assemblies derived from X-ray structures, while 20 are typical for NMR structures; viral capsid structures can have 60. Each model was subjected to refinement using Rosetta’s relax protocol, with atoms restrained to their initial coordinates and with extra side chain rotamers generated for the first and second  $\chi$ -dihedral angles. Five refined structures were generated for each model, and the lowest-energy model (using Rosetta’s talaris2013 scoring function) was selected for further processing. This relaxation procedure identifies the model from a multimodel PDB file that is best compatible with the Rosetta scoring function. Moreover, it helps to correct structural errors like steric clashes, misassigned side chain amides in glutamine and asparagine, and problematic rotamers. Correcting these errors in the starting structure prior to alanine scanning avoids an asymmetrical comparison between a poor input structure and a repaired alanine point mutant.

We applied the DSSP program to generate secondary structure assignments for that best model and then split it into individual files for each chain and for each pair of chains. We performed alanine

Table 2. Selection of the Diverse  $\beta$ -Strands Catalogued by SIPPDB<sup>a</sup>

Complex	Hot spots	% Contribution (MimeticScore)
<b>1. Only side-chain interactions (1OY3, chain B) – P65/P65</b>		
	<i>i, i+2, i+4</i> E211, 1.6881 F213, 3.1826 L215, 2.1768	39.5% $\Delta\Delta G$ (6.0 REU)
<b>2. Sheet-forming (1Z4E, chain A) – C125/C125</b>		
	<i>i, i+1, i+2, i+4, i+5, i+6, i+7, i+8</i> S145, 1.0588 H146, 2.1864 E147, 1.0397 L149, 1.6620 K150, 1.6949 M151, 1.2744 H152, 1.9825	17.5% $\Delta\Delta G$ (5.3 REU)
<b>3. Both side-chain and sheet-forming (4HPM, chain A) – PCGF/BCORL1</b>		
	<i>i, i+2, i+4</i> F1596, 3.3213 F1598, 4.3172 L1600, 1.4547	33.2% $\Delta\Delta G$ (9.1 REU)
<b>4. Hairpin (1F3U, chain A) – TFIIIF <math>\alpha</math>/TFIIIF <math>\beta</math></b>		
	<u>green:</u> <i>i, i+2, i+3, i+5</i> L93, 1.2049 V95, 1.0093 F96, 3.0797 E98, 2.3400 <u>magenta:</u> <i>i, i+2, i+3</i> L104, 1.4687 L106, 1.4915 E107, 2.1201	25.1% $\Delta\Delta G$ (5.7 REU) 5.0% $\Delta\Delta G$ (5.1 REU)
<b>5. Independent sheet components (2CFX, chain A) – LRPC/LRPC</b>		
	<u>green:</u> <i>i, i+2, i+4</i> F93, 1.4132 Y95, 4.2787 I97, 1.0193 <u>magenta:</u> <i>i, i+2, i+3, i+6</i> H131, 1.0704 H133, 1.9872 F134, 1.2298 I137, 2.8886	17.4% $\Delta\Delta G$ (6.7 REU) 16.2% $\Delta\Delta G$ (5.3 REU)

<sup>a</sup>MimeticScore, in the final column, is the sum of the  $\Delta\Delta G$  values of the three highest  $\Delta\Delta G$  hot spot residues.

scanning using the RosettaScripts AlaScan filter, averaging the result of 20 applications for better convergence and again using talaris2013 for scoring.

The subsequent data were processed with Perl scripts to identify strands (i.e., sequences of three or more residues long) with two or more hot spot residues ( $\Delta\Delta G > 1.0$  REU). Those interfaces with such qualifying strands were processed using the SASA calculation program NACCESS (the program was run on the files for each individual chain and on the complex file) and further analyzed to produce the data set available online at <http://www.nyu.edu/projects/arora/sippdb/>.

## AUTHOR INFORMATION

### Corresponding Author

\*Email: arora@nyu.edu.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was financially supported by the National Institutes of Health (R01GM073943) and the National Science Foundation (CHE-1151554). A.M.W. thanks New York University for the Kramer Fellowship.

## REFERENCES

(1) London, N., Raveh, B., and Schueler-Furman, O. (2013) Druggable protein–protein interactions—From hot spots to hot segments. *Curr. Opin. Chem. Biol.* 17, 952–959.

- (2) Ryan, D. P., and Matthews, J. M. (2005) Protein–protein interactions in human disease. *Curr. Opin. Struct. Biol.* 15, 441–446.
- (3) Zinzalla, G., and Thurston, D. E. (2009) Targeting protein–protein interactions for therapeutic intervention: A challenge for the future. *Future Med. Chem.* 1, 65–93.
- (4) Azzarito, V., Long, K., Murphy, N. S., and Wilson, A. J. (2013) Inhibition of  $\alpha$ -helix-mediated protein–protein interactions using designed molecules. *Nat. Chem.* 5, 161–173.
- (5) Arkin, M. R., and Wells, J. A. (2004) Small-molecule inhibitors of protein–protein interactions: Progressing towards the dream. *Nat. Rev. Drug. Discovery* 3, 301–317.
- (6) Cesa, L. C., Patury, S., Komiyama, T., Ahmad, A., Zuiderweg, E. R. P., and Gestwicki, J. E. (2013) Inhibitors of difficult protein–protein interactions identified by high-throughput screening of multiprotein complexes. *ACS Chem. Biol.* 8, 1988–1997.
- (7) Thompson, A. D., Dugan, A., Gestwicki, J. E., and Mapp, A. K. (2012) Fine-tuning multiprotein complexes using small molecules. *ACS Chem. Biol.* 7, 1311–1320.
- (8) Raj, M., Bullock, B. N., and Arora, P. S. (2013) Plucking the high hanging fruit: A systematic approach for targeting protein–protein interactions. *Bioorg. Med. Chem.* 21, 4051–4057.
- (9) Wells, J. A., and McClendon, C. L. (2007) Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* 450, 1001–1009.
- (10) Verdine, G. L., and Hilinski, G. J. (2012) Stapled peptides for intracellular drug targets. *Methods Enzymol.* 503, 3–33.
- (11) 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., and Fairlie, D. P. (2010) Downsizing human, bacterial, and viral proteins to short water-stable  $\alpha$ -helices that maintain biological potency. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11686–11691.
- (12) Kushal, S., Lao, B. B., Henchey, L. K., Dubey, R., Mesallati, H., Traaseth, N. J., Olenyuk, B. Z., and Arora, P. S. (2013) Protein domain mimetics as *in vivo* modulators of hypoxia-inducible factor signaling. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15602–15607.
- (13) 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., and Gellman, S. H. (2012) Evaluation of diverse  $\alpha/\beta$ -backbone patterns for functional  $\alpha$ -helix mimicry: Analogues of the Bim BH3 domain. *J. Am. Chem. Soc.* 134, 315–323.
- (14) Jochim, A. L., and Arora, P. S. (2009) Assessment of helical interfaces in protein–protein interactions. *Mol. Biosyst.* 5, 924–926.
- (15) Jochim, A. L., and Arora, P. S. (2010) Systematic analysis of helical protein interfaces reveals targets for synthetic inhibitors. *ACS Chem. Biol.* 5, 919–923.
- (16) Bullock, B. N., Jochim, A. L., and Arora, P. S. (2011) Assessing helical protein interfaces for inhibitor design. *J. Am. Chem. Soc.* 133, 14220–14223.
- (17) Bergey, C. M., Watkins, A. M., and Arora, P. S. (2013) HippDB: A database of readily targeted helical protein–protein interactions. *Bioinformatics* 29, 2806–2807.
- (18) Clackson, T., and Wells, J. A. (1995) A hot-spot of binding-energy in a hormone–receptor interface. *Science* 267, 383–386.
- (19) Cunningham, B. C., and Wells, J. A. (1989) High-resolution epitope mapping of hGH–receptor interactions by alanine-scanning mutagenesis. *Science* 244, 1081–1085.
- (20) Kortemme, T., Kim, D. E., and Baker, D. (2004) Computational alanine scanning of protein–protein interfaces. *Sci. STKE* 2004, pl2.
- (21) Dou, Y., Baisnee, P. F., Pollastri, G., Pecout, Y., Nowick, J., and Baldi, P. (2004) ICBS: A database of interactions between protein chains mediated by  $\beta$ -sheet formation. *Bioinformatics* 20, 2767–2777.
- (22) Cheng, P.-N., Pham, J. D., and Nowick, J. S. (2013) The supramolecular chemistry of  $\beta$ -sheets. *J. Am. Chem. Soc.* 135, 5477–5492.
- (23) Tyndall, J. D., Nall, T., and Fairlie, D. P. (2005) Proteases universally recognize  $\beta$ -strands in their active sites. *Chem. Rev.* 105, 973–999.
- (24) Loughlin, W. A., Tyndall, J. D. A., Glenn, M. P., and Fairlie, D. P. (2004)  $\beta$ -Strand mimetics. *Chem. Rev.* 104, 6085–6117.
- (25) Ripka, A. S., and Rich, D. H. (1998) Peptidomimetic design. *Curr. Opin. Chem. Biol.* 2, 441–452.
- (26) Ghosh, A. K., Chapsal, B. D., Weber, I. T., and Mitsuya, H. (2007) Design of HIV protease inhibitors targeting protein backbone: An effective strategy for combating drug resistance. *Acc. Chem. Res.* 41, 78–86.
- (27) Smith, A. B., Hirschmann, R., Pasternak, A., Guzman, M. C., Yokoyama, A., Sprengeler, P. A., Darke, P. L., Emini, E. A., and Schleif, W. A. (1995) Pyrrolinone-Based HIV Protease Inhibitors. Design, Synthesis, and Antiviral Activity: Evidence for Improved Transport. *J. Am. Chem. Soc.* 117, 11113–11123.
- (28) Cheng, P.-N., Liu, C., Zhao, M., Eisenberg, D., and Nowick, J. S. (2012) Amyloid  $\beta$ -sheet mimics that antagonize protein aggregation and reduce amyloid toxicity. *Nat. Chem.* 4, 927–933.
- (29) Potter, J. A., Owsianka, A. M., Jeffery, N., Matthews, D. J., Keck, Z.-Y., Lau, P., Fong, S. K. H., Taylor, G. L., and Patel, A. H. (2012) Toward a Hepatitis C Virus Vaccine: the Structural Basis of Hepatitis C Virus Neutralization by AP33, a Broadly Neutralizing Antibody. *J. Virol.* 86, 12923–12932.
- (30) Franklin, M. C., Carey, K. D., Vajdos, F. F., Leahy, D. J., de Vos, A. M., and Sliwkowski, M. X. (2004) Insights into ErbB signaling from the structure of the ErbB2–pertuzumab complex. *Cancer Cell* 5, 317–328.
- (31) Li, S., Schmitz, K. R., Jeffrey, P. D., Wiltzius, J. J. W., Kussie, P., and Ferguson, K. M. (2005) Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* 7, 301–311.
- (32) Gao, Y., Voigt, J., Wu, J. X., Yang, D. J., and Burke, T. R. (2001) Macrocyclization in the design of a conformationally constrained Grb2 SH2 domain inhibitor. *Bioorg. Med. Chem. Lett.* 11, 1889–1892.
- (33) Phan, J., Shi, Z. D., Burke, T. R., Jr., and Waugh, D. S. (2005) Crystal structures of a high-affinity macrocyclic peptide mimetic in complex with the Grb2 SH2 domain. *J. Mol. Biol.* 353, 104–115.
- (34) Yamagishi, Y., Shoji, I., Miyagawa, S., Kawakami, T., Katoh, T., Goto, Y., and Suga, H. (2011) Natural product-like macrocyclic N-methyl-peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-expressed *de novo* library. *Chem. Biol.* 18, 1562–1570.
- (35) Rose, G. D., Gierasch, L. M., and Smith, J. A. (1985) Turns in peptides and proteins, in *Advances in Protein Chemistry* (Anfinsen, C.B., Edsall, J. T., and Frederic, M. R., Eds.), pp 1–109, Academic Press, New York.
- (36) Freire, F., and Gellman, S. H. (2009) Macrocyclic design strategies for small, stable, parallel  $\beta$ -sheet scaffolds. *J. Am. Chem. Soc.* 131, 7970–7972.
- (37) Stanger, H. E., and Gellman, S. H. (1998) Rules for antiparallel  $\beta$ -sheet design: D-Pro-Gly is superior to L-Asn-Gly for  $\beta$ -hairpin nucleation. *J. Am. Chem. Soc.* 120, 4236–4237.
- (38) Robinson, J. A. (2008)  $\beta$ -Hairpin peptidomimetics: Design, structures, and biological activities. *Acc. Chem. Res.* 41, 1278–1288.
- (39) Smith, A. B., Guzman, M. C., Sprengeler, P. A., Keenan, T. P., Holcomb, R. C., Wood, J. L., Carroll, P. J., and Hirschmann, R. (1994) *De-novo* design, synthesis, and X-ray crystal-structures of pyrrolinone-based  $\beta$ -strand peptidomimetics. *J. Am. Chem. Soc.* 116, 9947–9962.
- (40) Phillips, S. T., Rezac, M., Abel, U., Kossenjans, M., and Bartlett, P. A. (2002) “@-tides”: The 1,2-dihydro-3(6H)-pyridinone unit as a  $\beta$ -strand mimic. *J. Am. Chem. Soc.* 124, 58–66.
- (41) Wyrembak, P. N., and Hamilton, A. D. (2009) Alkyne-linked 2,2-disubstituted-indolin-3-one oligomers as extended  $\beta$ -strand mimetics. *J. Am. Chem. Soc.* 131, 4566–4567.
- (42) Angelo, N. G., and Arora, P. S. (2005) Nonpeptidic foldamers from amino acids: Synthesis and characterization of 1,3-substituted triazole oligomers. *J. Am. Chem. Soc.* 127, 17134–17135.
- (43) Lingard, H., Han, J. T., Thompson, A. L., Leung, I. K. H., Scott, R. T. W., Thompson, S., and Hamilton, A. D. (2014) Diphenylacetylene-linked peptide strands induce bidirectional  $\beta$ -sheet formation. *Angew. Chem., Int. Ed.* 53, 3650–3653.
- (44) Fuller, A. A., Du, D., Liu, F., Davoren, J. E., Bhabha, G., Kroon, G., Case, D. A., Dyson, H. J., Powers, E. T., Wipf, P., Grubele, M., and

Kelly, J. W. (2009) Evaluating  $\beta$ -turn mimics as  $\beta$ -sheet folding nucleators. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11067–11072.

(45) Kang, C. W., Sun, Y., and Del Valle, J. R. (2012) Substituted imidazo[1,2-a]pyridines as  $\beta$ -strand peptidomimetics. *Org. Lett.* 14, 6162–6165.

(46) Hammond, M. C., Harris, B. Z., Lim, W. A., and Bartlett, P. A. (2006)  $\beta$ -Strand peptidomimetics as potent PDZ domain ligands. *Chem. Biol.* 13, 1247–1251.

(47) Fasan, R., Dias, R. L., Moehle, K., Zerbe, O., Vrijbloed, J. W., Obrecht, D., and Robinson, J. A. (2004) Using a  $\beta$ -hairpin to mimic an  $\alpha$ -helix: Cyclic peptidomimetic inhibitors of the p53-HDM2 protein–protein interaction. *Angew. Chem., Int. Ed. Engl.* 43, 2109–2112.

(48) Huang, Z., Zhang, M., Burton, S. D., Katsakhyan, L. N., and Ji, H. (2014) Targeting the Tcf4 G13ANDE17 binding site to selectively disrupt  $\beta$ -catenin/T-cell factor protein–protein interactions. *ACS Chem. Biol.* 9, 193–201.

(49) Zutshi, R., Franciskovich, J., Shultz, M., Schweitzer, B., Bishop, P., Wilson, M., and Chmielewski, J. (1997) Targeting the dimerization interface of HIV-1 protease: Inhibition with cross-linked interfacial peptides. *J. Am. Chem. Soc.* 119, 4841–4845.

(50) Lee, B., and Richards, F. M. (1971) Interpretation of protein structures—Estimation of static accessibility. *J. Mol. Biol.* 55, 379–400.

(51) Lo Conte, L., Chothia, C., and Janin, J. (1999) The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* 285, 2177–2198.

(52) Fischer, T. B., Holmes, J. B., Miller, I. R., Parsons, J. R., Tung, L., Hu, J. C., and Tsai, J. (2006) Assessing methods for identifying pairwise atomic contacts across binding interfaces. *J. Struct. Biol.* 153, 103–112.

(53) Bogan, A. A., and Thorn, K. S. (1998) Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* 280, 1–9.

(54) Koes, D. R., and Camacho, C. J. (2012) Small-molecule inhibitor starting points learned from protein–protein interaction inhibitor structure. *Bioinformatics* 28, 784–791.

(55) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.

(56) Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E. M., Khare, S. D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G., Meiler, J., and Baker, D. (2011) RosettaScripts: A scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* 6, e20161.

(57) Hubbard, S. J., and Thornton, J. M. (1993) NACCESS; Department of Biochemistry and Molecular Biology, University College, London.

(58) Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637.

(59) Pedersen, S. W., Pedersen, S. B., Anker, L., Hultqvist, G., Kristensen, A. S., Jemth, P., and Stromgaard, K. (2014) Probing backbone hydrogen bonding in PDZ/ligand interactions by protein amide-to-ester mutations. *Nat. Commun.* 5, 3215.

(60) Smith, C. K., and Regan, L. (1995) Guidelines for protein design—The Energetics of  $\beta$ -sheet side-chain interactions. *Science* 270, 980–982.

(61) Blasie, C. A., and Berg, J. M. (1997) Electrostatic interactions across a  $\beta$ -sheet. *Biochemistry* 36, 6218–6222.

(62) Nelson, D. L., Lehninger, A. L., and Cox, M. M. (2008) *Principles of Biochemistry*, 5th ed., W.H. Freeman, New York.

(63) Jones, S., and Thornton, J. M. (1995) Protein–protein interactions—A review of protein dimer structures. *Prog. Biophys. Mol. Biol.* 63, 31–65.

(64) Guharoy, M., and Chakrabarti, P. (2007) Secondary structure based analysis and classification of biological interfaces: Identification of binding motifs in protein–protein interactions. *Bioinformatics* 23, 1909–1918.

(65) Caudron, B., and Jestin, J. L. (2012) Sequence criteria for the anti-parallel character of protein  $\beta$ -strands. *J. Theor. Biol.* 315, 146–149.

(66) Zhao, Y. L., and Wu, Y. D. (2002) A theoretical study of  $\beta$ -sheet models: Is the formation of hydrogen-bond networks cooperative? *J. Am. Chem. Soc.* 124, 1570–1571.

(67) Wang, Z. X., Wu, C., Lei, H. X., and Duan, Y. (2007) Accurate *ab initio* study on the hydrogen-bond pairs in protein secondary structures. *J. Chem. Theory Comput.* 3, 1527–1537.