Structure
Article

The Structural Basis of Peptide-Protein Binding Strategies

Nir London,[1] Dana Movshovitz-Attias,[2] and Ora Schueler-Furman[1,*]
[1]Department of Microbiology and Molecular Genetics, Institute for Medical Research IMRIC, Faculty of Medicine, The Hebrew University, POB 12272, Jerusalem 91120, Israel
[2]School of Computer Science and Engineering, The Hebrew University, Jerusalem 91904, Israel
*Correspondence: oraf@ekmd.huji.ac.il
DOI 10.1016/j.str.2009.11.012

SUMMARY

Peptide-protein interactions are very prevalent, mediating key processes such as signal transduction and protein trafficking. How can peptides overcome the entropic cost involved in switching from an unstructured, flexible peptide to a rigid, well-defined bound structure? A structure-based analysis of peptide-protein interactions unravels that most peptides do not induce conformational changes on their partner upon binding, thus minimizing the entropic cost of binding. Furthermore, peptides display interfaces that are better packed than protein-protein interfaces and contain significantly more hydrogen bonds, mainly those involving the peptide backbone. Additionally, "hot spot" residues contribute most of the binding energy. Finally, peptides tend to bind in the largest pockets available on the protein surface. Our study is based on peptiDB, a new and comprehensive data set of 103 high-resolution peptide-protein complex structures. In addition to improved understanding of peptide-protein interactions, our findings have direct implications for the structural modeling, design, and manipulation of these interactions.

INTRODUCTION

Protein-protein interactions play an important role in the living cell. These interactions are versatile and come in various flavors. Although many interactions between proteins involve the classical, well-characterized binding between two globular domains, an increasing number of interactions have recently been reported to involve peptide-protein interactions, where short linear peptides bind to globular protein receptors (Pawson and Nash, 2003). These linear peptides might origin from a loop within a structured domain, or from a disordered region in protein termini or between defined domains. Peptide-protein interactions have been acknowledged as important mediators of protein-protein interactions, predominantly in signaling and regulatory networks (Pawson and Nash, 2003), and a number of examples for such interactions have been studied (Neduva and Russell, 2006; Petsalaki and Russell, 2008). Peptide-protein interactions are also attractive drug targets both for small molecules and for designed inhibitory peptides (Hayouka et al., 2007; Parthasarathi et al., 2008; Zhao and Chmielewski, 2005). For these reasons, it is very compelling to investigate basic principles that govern these interactions.

What are the strategies that peptides use for binding? How do they compensate for the configurational entropy lost upon binding? What stabilizes these interactions? What is the recognition process for these binding events? To address these questions, there is a need for a comprehensive structural data set of peptide-protein interactions. At the sequence level, an increasing amount of sequences of peptide-protein interaction has been deposited in databases such as DOMINO (Ceol et al., 2007). In addition, short, linear motifs have been shown to define certain peptide-protein interactions (e.g., ELM [Puntervoll et al., 2003]). These motifs were indeed used to extract a collection of peptide-protein complex structures (Parthasarathi et al., 2008; Stein and Aloy, 2008); other peptide-protein complexes data sets have been collected as well (Petsalaki et al., 2009). However, a broad structural characterization of protein-peptide complexes is still lacking. Pioneering structural studies of peptide-protein interactions are more than a decade old and thus based on very small data sets (e.g., Stanfield and Wilson, 1995; Zvelebil and Thornton, 1993). Thus, while several analyses have characterized protein-protein interfaces (e.g., Bahadur et al., 2004; Jones et al., 2000; Lo Conte et al., 1999; Nooren and Thornton, 2003; Reynolds et al., 2009; Rodier et al., 2005; Xu et al., 1997), and recent studies have also investigated the binding characteristics of intrinsically unstructured proteins (IUPs) to proteins (Fong et al., 2009; Meszaros et al., 2007; Mohan et al., 2006; Vacic et al., 2007), no thorough examination of unique characteristics of peptide-protein complexes is yet available.

In order to investigate the structural basis of peptide-protein interactions, we have created a non-redundant database of high-resolution structures of peptide-protein complexes (termed *peptiDB)*. Analysis of this database shows that peptides bind to proteins in a fashion that minimizes the conformational changes of the protein partner, while maximizing the enthalpy gained by hydrogen bonds and packing. This might allow the peptide to overcome its own configurational entropy loss upon binding. Similar to protein interactions, binding is mediated by "hot spot" residues (as suggested by Stein and Aloy, 2008). Finally, we show that peptides usually bind within the largest pocket on the protein surface, a finding that can complement recent attempts to computationally predict the binding sites of peptides (e.g., Petsalaki et al., 2009). Insights from this study

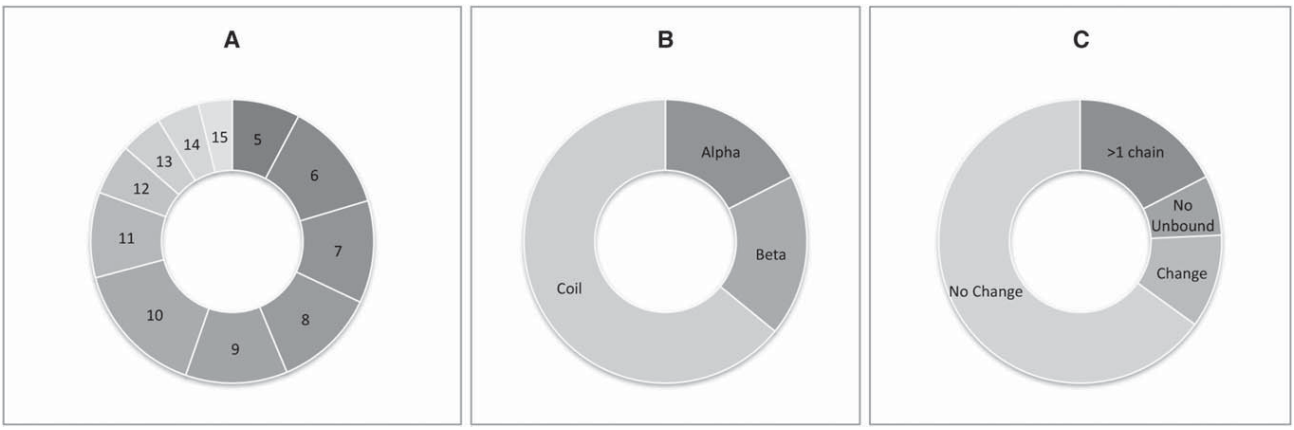

Figure 1. The peptiDB Data Set of Peptide-Protein Interactions
The distribution of different features is shown.
(A) Peptide sizes in the data set.
(B) Peptide secondary structure.
(C) Conformational changes of the binding protein.
" > 1 chain" indicates cases where the binding protein is composed of two chains; in these cases no free conformations were evaluated. Most peptides are between 6 and 11 residues long, adapt a coiled conformation, and induce no significant conformational changes in the protein partner upon binding.

can be incorporated to improve both the structural modeling of peptide-protein interactions, as well as their structure-based manipulation.

RESULTS

PeptiDB: A Database of High-Resolution Protein-Peptide Complexes

We created a database of 103 high-resolution peptide-protein complexes, named *peptiDB* (see Table S1 available online). This set is the basis for our analysis to understand the binding strategies of short peptides (5 to 15 residues long; Figure 1A) to proteins. The peptide-protein interactions included are involved in various cellular activities, such as signal transduction, protein trafficking and transport, antigen binding, enzyme substrates/inhibition, and others, and it appears that many of the peptides are in fact linear recognition motifs (LMs) that are derived from a larger protein (see Results below, and Fuxreiter et al., 2007).

The data set of 103 peptide-protein complexes contains no two protein monomers that share more than 70% sequence identity. To ensure that no bias due to fold overrepresentation is introduced, we also created a smaller set of 61 protein-peptide complexes, in which no two protein monomers share the same fold (according to CATH [Orengo et al., 1997]; see Experimental Procedures and Table S1). Analysis of these two data sets yielded overall similar results. We therefore report here our results from analysis of the set of 103 complexes (and detail about the small set if results differ significantly).

Most of the peptides bind in either extended or coiled conformation; merely 18 of the peptides were found to bind as helices, while 19 bind as a β strand adjacent to or within a β sheet in the binding partner (Figure 1B). Examination of Ramachandran plots for the remaining 66 coiled peptides that do not adopt a regular secondary structure shows a ϕ/ψ distribution that is similar to that of β strand peptides and includes only few outliers (see Figure S1). Certain binding features can differ between peptides that adopt different secondary structures, and will be described in detail. Despite variability in structure and function, our analysis of the peptide binding features below suggests that, in general, similar binding strategies are shared by most peptides.

Characterization of Peptide-Protein Binding versus Protein-Protein Binding

Various studies have tried to characterize features of homodimeric, heterodimeric, and transient protein-protein interactions (Bahadur et al., 2004; Jones et al., 2000; Lo Conte et al., 1999; Nooren and Thornton, 2003; Reynolds et al., 2009; Rodier et al., 2005; Xu et al., 1997), as well as interactions between IUPs and their partners (Meszaros et al., 2007; Mohan et al., 2006; Vacic et al., 2007). A noticeable variance exists between the reported results, due to different data sets and different implementations of the measurements. In this study we evaluated a set of features on the data sets of Mintseris and Weng (2003) (PPI, 205 complexes), Meszaros et al. (2007) (IUPPI, 35 complexes), and the present peptiDB as representatives of protein-protein, protein-IUP, and peptide-protein interactions, respectively (see Table 1 for a summary of values, standard deviations, and statistical significance of observed differences; and Experimental Procedures for a description of evaluated features, as well as statistical tests). In addition, we compare our calculated values to those reported in the literature (see Table S2, and Bahadur et al., 2004; Jones et al., 2000; Lo Conte et al., 1999; Nooren and Thornton, 2003; Reynolds et al., 2009; Rodier et al., 2005; Xu et al., 1997).

Although both peptides and IUPs lack a defined structure when unbound, and gain structure upon binding, and although there are cases in which the peptide is but a subset of a larger IUP (or protein), peptides and IUPs define two different classes of interaction and indeed their binding features are significantly

Table 1. Interface Characteristics of Peptide-Protein Complexes, Compared with Protein-Protein Interactions, and Interactions between Intrinsically Unstructured Proteins and Proteins

Complex Type	Peptide-Protein (n = 103)[a]		PPI (n = 205)			IUPPI (n = 35)		
Size	Mean	±SD	Mean	±SD	Significance[b]	Mean	±SD	Significance
Peptide length	9.2	±2.7				40.3	±26.5	
ASA (Å^2)	512	±177	1151	±605	< < e-100	1361	±707	< < e-100
Interface residues	20.8	±8.6	33.2	±18.5		27.3	±15.8	
Interface atoms	56.1	±20.5	100.1	±47.7		104.7	±57.7	
% Atoms polar	34.2	±9.7	32.9	±9.5	NS[c]	28.7	±8.5	0.03
% Atoms nonpolar	44.7	±9.9	45.2	±13.5	NS	55.9	±13.4	5.1e-6
% Atoms neutral	19.8	±8.3	20.3	±8.9	NS	15.2	±8.0	0.049
Shape								
Planarity	2.6	±0.7	3.1	±1.1	5.0e-4	3.1	±1.2	0.016
Eccentricity	0.7	±0.1	0.8	±0.2	NS	0.8	±0.2	1.8e-3
RosettaHoles score	3.0	±0.3	4.1	±1.7	1.2e-8	4.1	±1.2	3.5e-12
Polar interactions								
Hydrogen bonds (HB)	8.1	±5.0	9.7	±4.6		9.3	±7.7	
HB per 100 Å^2	1.6	±0.7	1.0	±0.6	7.5e-12	0.7	±0.4	9.6e-10
Peptide side-chain HB	3	±2.6	6.0	±3.7		4.7	±4.8	
Peptide main-chain HB	5.1	±4.0	3.7	±2.8		4.6	±4.4	
Salt bridges (SB)	1.2	±1.0	2.6	±2.7		2.9	±3.1	
SB per 100 Å^2	0.2	±0.2	0.2	±0.3	NS	0.3	±0.25	NS
Bridging H_20	4	±2.8	6.6	±13				
Bridging H_20 per 100 Å^2	0.8	±0.5	0.7	±1.5	NS			

The different measures are described in Experimental Procedures.
[a] The size of each data set is indicated in parentheses.
[b] Statistical significance between the peptiDB set and other sets was assessed by a two-sample *t*-test. We report p-values after Bonferroni correction for multiple testing (e.g., the p-value obtained from the statistical test is multiplied by the number of independent evaluations).
[c] NS indicates no significant difference between the two distributions.

different, as described below. This difference might be explained by findings in a recent study, which reported that although LMs tend to locate to unstructured regions, their amino acid composition differs substantially from the composition observed both in their surrounding flanking regions, as well as in IUPs (Fuxreiter et al., 2007). Analysis of the amino acid propensities in our data set indeed showed the highest degree of similarity to LMs, in particular to the subset of specificity determining residues (RSs in the study of Fuxreiter et al.), and the lowest correlation to the distribution in IUPs or flanking regions around LMs, indicating that the predominant part of our data set consists of LMs (see Figure S2).

The average solvent-accessible surface area that is buried upon peptide binding to a protein (ASA) is around 500 Å^2, compared with an area twice as large in protein-protein complexes, and almost three times as large in IUP-protein interactions (Table 1). However, within this small buried area, peptides optimize different binding features, and in particular hydrogen bonds. Binding shape analysis shows that peptides tend to bind in a more planar fashion than proteins and IUPs. They also display better packing at the interface, with an average RosettaHoles score (Sheffler and Baker, 2009) of 3.0 compared with 4.1 and 4.0 in IUPPI and PPI, respectively. We should note that the RosettaHoles score is sensitive to the structure's resolution. Therefore, structures in the IUPPI set receive on average poor scores (because the resolution of the structures in this set is lower). A subset of the PPI set with resolution < 2 Å (64/205) obtained an average score of 3.3, which is still significantly higher than the peptide-protein set (p_{val} = 5.8e-14).

Peptides Use More Hydrogen Bonds Than Proteins in Binding to Their Protein Partner, and Many of the Hydrogen Bonds Involve the Peptide Backbone

An average of 8.1 hydrogen bonds are formed in peptide-protein interfaces, compared with 9.7 hydrogen bonds in protein-protein interfaces (11.0 in a high-resolution subset of the PPI set, see Experimental Procedures), and 9.3 in IUP-mediated interactions. Considering the smaller interface, this clearly indicates that peptides form more hydrogen bonds per interface area (about 50% more than in protein-protein interactions, and more than double of IUP-protein interactions per 100 Å^2 ASA; see Table 1). This increase in hydrogen bond density, however, cannot be attributed to different amino acid propensities, because peptides display a similar distribution of polar/nonpolar atoms at the interface as protein-protein interfaces do, whereas interestingly, IUP-mediated interactions show significantly more hydrophobic regions in the interface, as already noted by Meszaros et al. (2007) (see Table 1).

The number of hydrogen bonds formed upon peptide binding, as well as the distribution of different types of atoms at the

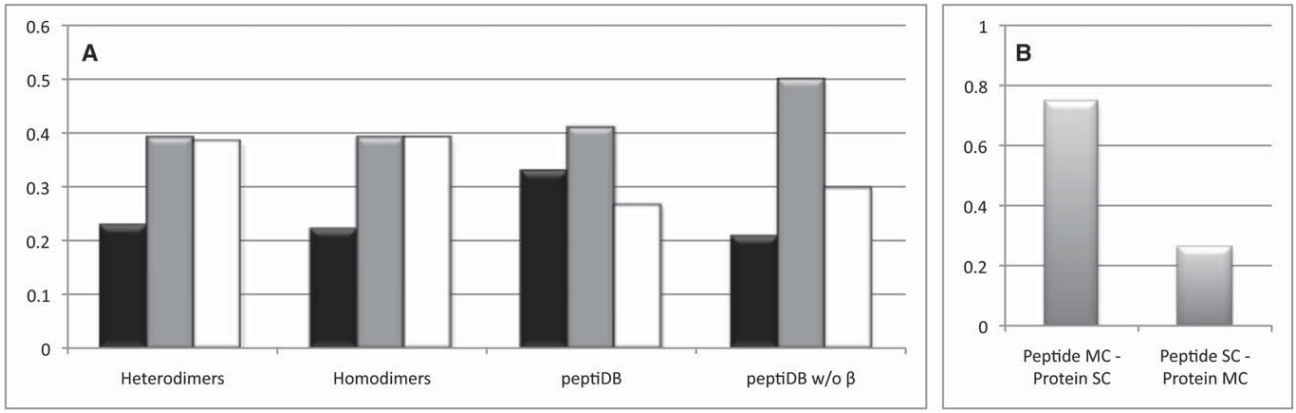

Figure 2. Substantial Involvement of the Peptide Main Chain in Peptide-Protein Hydrogen Bonds

(A) Comparison of the distribution of hydrogen bonds across interfaces of heterodimers, homodimers, and peptide-protein complexes. Black represents main-chain/main-chain hydrogen bonds; gray, side-chain/main-chain hydrogen bonds; white, side-chain/side-chain hydrogen bonds. Heterodimers and homodimers were taken from Cohen et al. (2008). PeptiDB, full data set; PeptiDB w/o β, set without β sheet forming peptides.

(B) Distribution of side-chain/main-chain hydrogen bonds in peptiDB: in most cases the main-chain atom involved in the hydrogen bond is contributed by the peptide.

interface (e.g., nonpolar, polar, and neutral atoms), differs substantially for both α-helical peptides and β strand peptides. α-Helical peptides form on average a significantly smaller number of hydrogen bonds with the protein (4.2), and contain many more nonpolar atoms at the interface (53%). Helical wheel analysis showed that most of these peptides form amphiphilic helices and bind with a hydrophobic surface, which explains the distinct properties of this class of peptides. β Strand peptides, on the other hand, form many more hydrogen bonds (12.5 on average), because they are involved in a β sheet that is stabilized by main chain hydrogen bonds (see below).

Hydrogen bonds across the peptide-protein interface involve substantially more main-chain atoms than in protein-protein interfaces (see Figure 2A): A comparison of the distribution of main-chain/main-chain, main-chain/side-chain and side-chain/side-chain hydrogen bonds across protein-protein interfaces (measured in a set of homodimers and a set of heterodimers described in Cohen et al., 2008) and peptide-protein interfaces revealed that the latter utilize significantly more main-chain/main-chain hydrogen bonds (32% of the hydrogen bonds at the interface), while keeping the same level of main-chain/side-chain hydrogen bonds (40%) (see Figure 2A; p_{val} = 6.2e-18, assessed by χ^2 test). This overrepresentation of main-chain/main-chain hydrogen bonds is mainly contributed by peptides that bind as part of a β sheet. When these are excluded from the data set, the distribution shifts toward a significant overrepresentation of main-chain/side-chain hydrogen bonds (p_{val} = 1.2e-7, Figure 2A). Most of these hydrogen bonds involve a peptide main-chain atom and a protein side-chain atom (Figure 2B). The few α-helical peptides contribute only a small number of hydrogen bonds, (see above; these mainly connect a main-chain and a side-chain atom), and therefore do not affect this distribution significantly.

Note that a similar analysis on *intra*chain hydrogen bonds over the proteins from these data sets showed identical distributions for proteins that bind peptides, homodimers, and heterodimers. Most hydrogen bonds were mediated by main-chain/main-chain interactions (66%), followed by 22% main-chain/side-chain and 12% side-chain/side-chain interactions. Thus, the observed differences are not due to differences of the proteins in the data sets used to evaluate protein-protein interactions and protein-peptide interactions.

The density of salt bridges, on the other hand, is very similar for the three different classes of interactions (0.2/100 Å^2 ASA), as well as the number of interface water-mediated hydrogen bonds.

Peptides Use Hot Spots to Bind to Proteins

It is well established that in protein-protein interfaces the critical contribution to binding energy is due to a small number of residues, which have been termed hot spot residues (Clackson and Wells, 1995; Dall'Acqua et al., 1996). We wanted to assess whether hot spots are also observed in peptide-protein interactions. For that purpose we used computational alanine scanning (see Experimental Procedures) to mutate in turn every peptide residue in each of the complexes to alanine. Residues that were predicted to harm binding by more than a given threshold (here ΔΔG > 1 kcal/mol, see Experimental Procedures) were defined as hot spots. Two possible strategies for peptide binding would be: (1) each amino acid in the peptide contributes a small amount to the binding energy (i.e., a uniform distribution of binding energies), or (2) a few hot spots are responsible for most of the binding energy.

As can be seen in Figure 3, the second option is observed in peptide-protein complexes. Few hot spots mediate the major part of the free energy of binding: more than 70% of the predicted decrease in binding free energy upon mutation to alanine is contributed by hot spot residues in 77/103 (75%) of the peptides. We note that when a more stringent threshold is applied to define hot spots (ΔΔG > 1.5 kcal/mol), the number of predicted hot spots decreases, but still more than 60% of the predicted decrease in binding free energy is due to hot spots in 60% of the peptides. This correlates nicely with results from

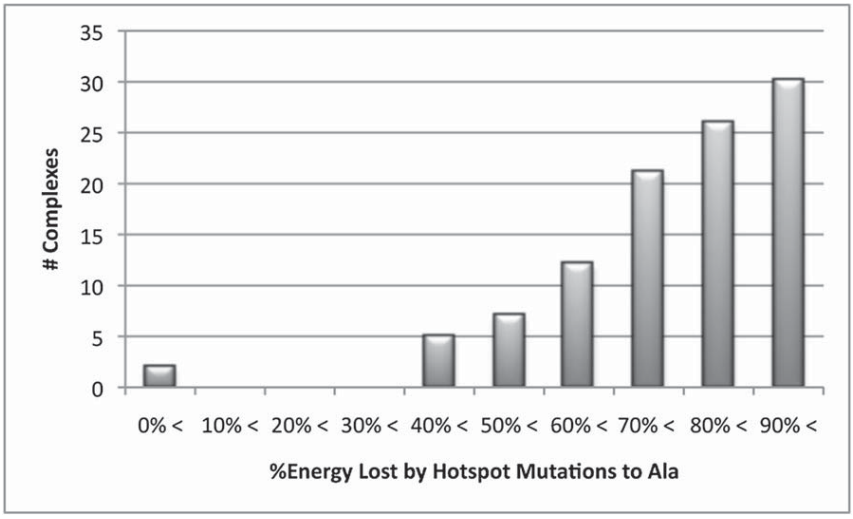

Figure 3. Peptide Hot Spot Residues Contribute the Major Part of the Binding Energy

The binding energy is not uniformly distributed among all peptide residues, rather it concentrates on a few hot spots. This histogram shows the percentage of the total binding energy that is contributed by these hot spots in each of the peptide-protein complexes of peptiDB. The contribution to binding energy by different residues is evaluated by calculating the predicted effect of mutating each residue to alanine (see Experimental Procedures). For example, in 30 complexes, 90%–100% of the binding energy is contributed by peptide hot spots, whereas only in 5 complexes 40%–50% of the binding energy origins from peptide hot spots.

another study that claimed that most of the energy in peptide binding is due to the binding of a core motif (Stein and Aloy, 2008). As an example, the androgen receptor is known to bind ligands that contain the FxxLF motif (He et al., 2002). Our hot spot analysis on a complex of this receptor bound to a decamer peptide SSRFESLFAG (Protein Data Bank [PDB] ID 1T7R) indeed predicts that only the known motif residues are binding hot spots (i.e., positions F4, L7, and F8 result in an energy loss of $\Delta\Delta G_{bind}$ = 3.0, 1.0, and 3.2 kcal/mol in Rosetta energy units, respectively). Our present analysis, however, goes beyond known binding motifs: it suggests that peptide hot spots are a general feature that can be used to characterize also peptides with no known binding motif.

We observed a slight dependency of the number of predicted hot spots on the length of the binding peptide (see Figure S3A; R^2 = 0.4; p_{val} < 0.005 in a one-tailed *t*-test. On average, there are about 2 hot spots in peptides of lengths 6-8 (2.1 ± 1) and about 3 hot spots in peptides of lengths 9–11 (2.9 ± 1.1). Hot spots do not show a tendency to occur in the terminal residues, rather, a certain preference for the central residues was observed (see Figure S3B). The same holds for a more stringent definition of hot spots ($\Delta\Delta G$ > 1.5 kcal/mol) with an average number of hot spots of 1.5 ± 1 for peptides of length 6–8 and 2 ± 1.4 for peptides of length 9–11.

Amino Acid Propensities of Peptide Residues and of Peptide Hot Spot Residues

In order to see which amino acids are overrepresented in peptide interface hot spots, we first calculated the amino acid frequencies in the peptides of our data set (n = 853; polyproline peptide were omitted here, to avoid a bias toward proline). Leucine shows a very high frequency among peptide residues, much higher than in protein-protein interfaces (Glaser et al., 2001) (11% versus 8%; p_{val} = 0.001 for χ^2-test; Figure 4A). We note that leucine has also been observed at a high frequency in IUPPI (Meszaros et al., 2007).

We next calculated the frequencies of each amino acid among the peptide interface hot spots. Figure 4B shows the overrepresentation of each amino acid as a peptide interface hot spot, normalized by its background frequency in the binding peptides. Peptide interface hot spot residues are most significantly enriched in phenylalanine (p_{val} = 6e-13), leucine (p_{val} = 1.8e-11), tryptophan (p_{val} = 1.7e-7), tyrosine (p_{val} = 1.3e-6), and isoleucine (p_{val} = 5e-5). The same residues are overrepresented even with the more stringent hot spot definition. Thus, in addition to the high relative frequency of aromatic residues among known interface hot spot residues in protein-protein interfaces (Glaser et al., 2001), peptide interface hot spots tend to be enriched in leucine and isoleucine as well. Interestingly, even when taking into account its high frequency in peptides, leucine is still overrepresented in hot spot positions. This might be explained by the overrepresentation of Retinoid-X-receptor domain binding peptides in the data set, which bare the LxxLL motif (ten entries). However, the results were also reproduced on the smaller, fold nonredundant set: leucine still prevails. This is in contrast to experimental data regarding protein-protein interfaces in general, where according to ASEdb (Bogan and Thorn, 1998) only 1 of 11 leucines in protein-protein interfaces is a hot spot that contributes significantly to binding. However, interface leucines are underrepresented in that database (with a frequency of just 3.6%) and the definition of a hot spot is different as well, complicating this comparison.

Peptides Adapt Themselves to the Unbound Structure of the Protein

A previous analysis of the binding of MoRFs (molecular recognition features) to proteins reported specific examples of various degrees of structural change (Vacic et al., 2007). We assess here the conformational changes upon peptide binding on a large set of proteins. Toward this goal, we constructed a data set of corresponding protein structures solved without the peptide (unbound set). Eighty-five entries in peptiDB involve one protein chain that binds the peptide (in the remaining entries, the protein partner is composed of two chains), and for 78 an unbound monomer structure was found (or a close homolog, see

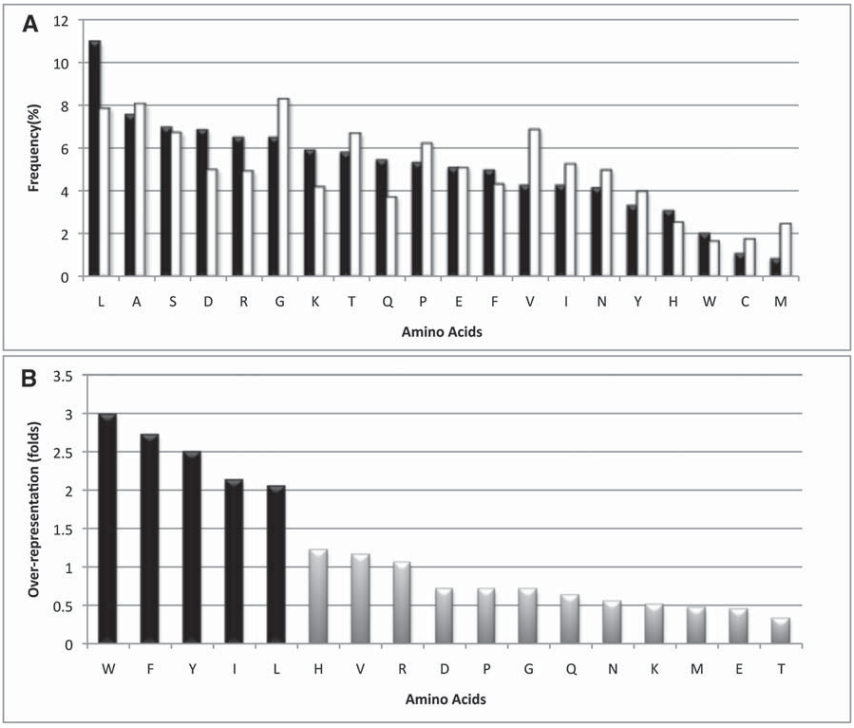

Figure 4. Aromatic Amino Acids, as Well as Leucine and Isoleucine, Are Overrepresented in Peptide Hotspot Residues

(A) The frequency of different amino acids in binding peptides (in peptiDB, black), compared with protein-protein interfaces (as calculated by Glaser et al. [2001], white).

(B) Amino acid overrepresentation in peptide hot spot residues: the plot shows the overrepresentation of each amino acid type in hot spots, relative to its frequency in binding peptides (shown in A). Hot spot residues are enriched significantly with amino acids W, F, Y, I, and L. Leucine is overrepresented in peptide hotspots, even when normalized against its already high background frequency in peptides.

Experimental Procedures). By comparing the structure of the protein prior to its association with the peptide to its structure after binding of the peptide, we could assess the degree of conformational change induced by the peptide.

The results of this analysis are striking (Table 2): 67 proteins in the data set (i.e., 86% of the unbound data set) do not change substantially: we calculated an average value of 0.83 Å root-mean-square deviation (rmsd) on interface Cα atoms, and 1.48 Å rmsd over all interface atoms (including side chains; see Figure 5 for an example). Using a more stringent set (33 proteins that include only unbound protein structures solved by X-ray crystallography, with a resolution better than 2 Å and more than 90% sequence identity), the results are even more pronounced: the average rmsd value over all interface atoms dropped to 1.06 Å (Table 2). Proteins that bind peptides in a β strand conformation tend to undergo slightly larger conformational changes than proteins that bind peptides α-helical or coiled conformations: their average Cα-atom rmsd value between the bound and free conformations is 1.17Å (average all-atom rmsd = 1.91 Å). In these cases, both sides undergo conformational changes when they extend a β sheet.

For the remaining peptide-protein interactions in our data set, we observed rather singular binding strategies. In four cases, the peptide displaces a part of the unbound monomer, be it a strand within a β sheet (in PDB IDs 1H6W and 3BFQ), an α-helix (2B1Z), or a coil (1VZQ). In two other instances, a loop at the binding site moves to accommodate the peptide (in 1KL3 and 2BBA), while in another three cases a disordered region at the interface becomes structured upon peptide binding (in 2QOS, 2P1K, and 3BU3). Finally, in one last case a helix breaks down to rearrange and become a part of the peptide binding interface (in 2P1T).

For 7 proteins (out of the 85 complexes that involve one protein chain, see above), no unbound structure without bound peptide is available, maybe because they are not stable without the peptide. This includes MHC proteins that have been reported to form a stable structure only upon binding of the antigenic peptide (Springer et al., 1998; Zarutskie et al., 1999), and complexes in which the peptide participates in the formation of a β sheet (e.g., 1N12, 2AI4, and 3BQF).

In principle, structural changes can also propagate to other regions of the protein, for example in cases where allosteric effects are involved in transferring the signal of the peptide to other functional regions in the protein (e.g., Lockless and

Table 2. Conformational Changes Induced by Binding of the Peptide

Data Set	Size	No. of Cα Atoms[a]	Cα Atom Rmsd[b]	No. of Atoms[c]	All Atom Rmsd	Resolution[b]	% Sequence Identity
Full data set	67	15.73 ± 4.75	0.83 ± 0.7	128.1 ± 37.45	1.48 ± 0.86	1.82 ± 0.49[d]	85.2 ± 21.7
Stringent set[e]	33	15.73 ± 4.5	0.47 ± 0.33	130.35 ± 37.17	1.06 ± 0.5	1.54 ± 0.37	99.3 ± 1.7

Average values and standard deviations are indicated.

[a] Number of protein Cα atoms at the interface (see Experimental Procedures for definition of interface).

[b] Rmsd and resolution are given in angstroms.

[c] Total number of protein atoms at the interface.

[d] The average resolution was calculated for the subset of structures determined by X-ray crystallography (12 structures were solved by nuclear magnetic resonance).

[e] This data set contains structures with good resolution (<2 Å) and high sequence identity (>90%) only. See text for more details.

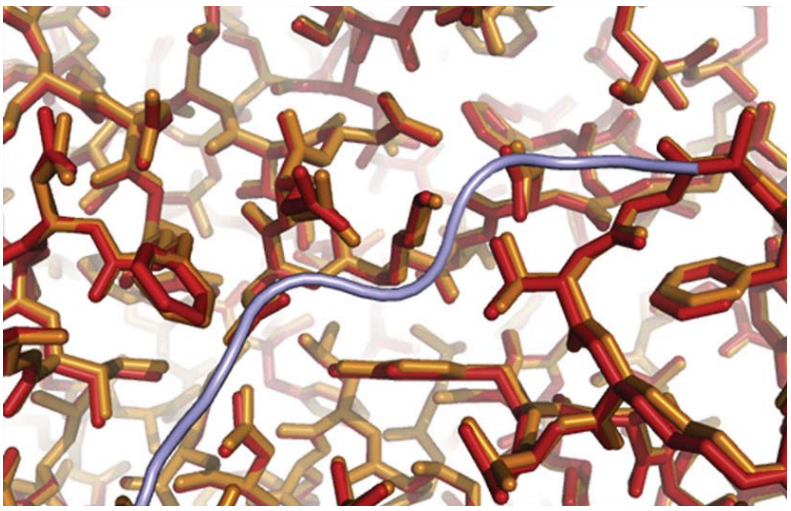

Figure 5. Protein Side Chains Are Prearranged to Accommodate the Peptide

The unbound human protein kinase PIM-1 (in red; PDB ID 2J2I:B) shows perfect agreement with the bound structure (in orange, bound to the PIM-1 consensus peptide in purple; PDB ID 2C3I): Note the exact placement of side chains, and a very low interface atom rmsd value of 0.29 Å.

Ranganathan, 1999). We confined ourselves to the immediate interface, to increase our confidence that the observed changes are indeed due to the peptide binding. However, we did not observe cases of substantial structural changes in areas remote from the peptide binding site (the average rmsd between bound and unbound proteins was 1.26 ± 0.9, calculated over an average of 88 ± 9% of the sequence).

It is also interesting to note that B-factor analysis of unbound structures showed no consistent tendency for interface residues. In the bound structures, predictably, the interface residues were colder than the protein average while the peptide was warmer than the average (data not shown).

Peptides Tend to Bind in the Largest Pocket on the Protein Surface

Where do peptides bind on the protein surface? Are there favored regions that will optimize binding cost? Specifically, are pockets preferred? We used the CASTp server (Dundas et al., 2006) to define pockets on the protein structure and evaluated whether peptides indeed bind in those regions. On the subset of 85 complexes in which the protein monomer consists of one chain only (see Figure 1C), CASTp detected an average of 15 ± 10 pockets on each protein.

We detected two main binding strategies regarding the utilization of pockets (see Figure 6A). Twenty-two peptides (26%) bind to a very large pocket (pocket ASA > 100 Å^2, calculated using the algorithm of Lee and Richards (1971), in most cases to the largest pocket available on the protein surface (18/22; see Figure 6B and an example for such a complex in Figure 6C). Forty peptides (47%) were found to bind to a small pocket (pocket area < 100 Å^2); in these cases, one of the peptide's side chains is buried in this pocket in a knob-hole fashion (see Figure 6D for an example). Interestingly, even when the peptide latches onto a small pocket, this is still in general the largest pocket available on the protein (29/40; see Figure 6B). Peptides that bind through the knob-hole strategy display a relatively small overall ASA of up to 600 Å^2, while peptides that bind in large pockets can display larger ASA of up to 1000 Å^2. Pocket parameters for these two groups are summarized in Table 3. α-Helical peptides tend to bind using the knob-hole strategy, whereas β strand peptides prefer pockets.

Of the peptides that did not bind to any kind of pocket on the protein surface (detected by CASTp), nine are polyproline peptides. Indeed, extended polyproline conformation might not fit into a compact pocket. Alternatively, in absence of a suitable binding environment, incorporation of prolines may be favorable for peptide binding, because this will restrict the backbone flexibility and thus limit the configurational entropy loss. Indeed, in a recent study that reported significantly lower configurational entropy for known peptide inhibitors, polyproline peptides were among those with lowest entropy values (Unal et al., 2009). Lastly, two peptides were bound in a cavity (not accessible to solvent) inside the binding proteins (PDB IDs 1XOC and 2D5W) and two more peptides seem to bind in a pocket undetected by CASTp (PDB IDs 1CZYand 1KL3).

DISCUSSION

Binding of peptides to proteins is met by challenges that are common to, but also distinct from, those encountered by proteins interacting with other proteins. Peptides are much shorter than proteins, and still they are able to bind to proteins in a very efficient way, thereby allowing integration of communication networks by the use of simple and small linear motifs. How are these interactions facilitated? How are they different from interactions between two globular proteins? Here we have analyzed for the first time a representative data set of atom-resolution peptide-protein complex structures, and compared it with the well-characterized structural properties of protein complexes in general. By analyzing a range of different parameters, we were able to define the prominent differences between these two approaches to communication between molecules in cells.

Ways to Optimize Peptide Binding to Proteins

In contrast to stable, folded protein domains, peptides are usually flexible molecules in their free state (e.g., Ho and Dill, 2006) and lose a large amount of configurational entropy upon association (e.g., Killian et al., 2009). How then do they pay the price of the reduction in this entropy upon binding? Our analysis suggests several possible answers to this question.

Proteins are surprisingly rigid regarding their association with peptides. For most of the peptide-protein interactions, we

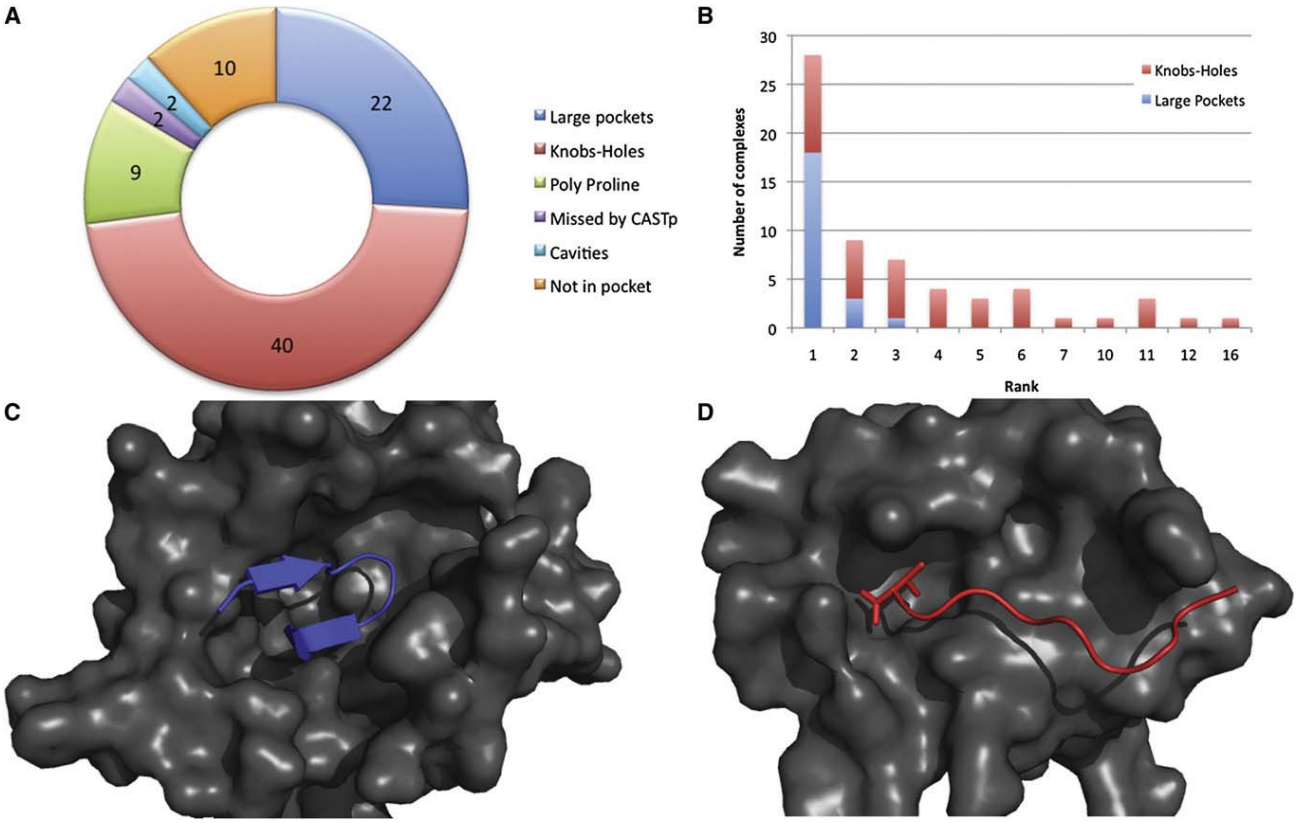

Figure 6. Peptides Tend to Bind in the Largest Pocket on the Protein Surface

(A) Distribution of peptide pocket binding strategies. Most peptides bind either to a large pocket, or latch on to a hole that accommodates one side chain.

(B) Whether binding to a large pocket or binding via a knob-hole fashion interaction, peptides tend to bind to the largest pocket on the surface.

(C) Complement protein C8 in complex with a peptide baring the C8 binding site (PDB ID 2QOS). The β hairpin peptide is bound in the largest pocket on the protein surface.

(D) DnaK substrate binding PDZ domain bound to a substrate peptide (PDB ID 1MFG). The peptide is anchored to the protein via its C′-terminal valine knob that fits into its corresponding hole in the PDZ domain.

could find only very small conformational changes at the protein interface. Importantly, in roughly half of the complexes, at least one bridging water molecule is bound already in the free protein monomer, as part of the preformed complex. Thus, some of the configurational entropic cost has been paid in advance for the binding protein, as well as for water molecules at the interface. This "prepaid strategy" might imply that peptide binding will not induce a substantial reduction in overall configurational entropy, thereby minimizing the cost of association.

In other cases, the free protein structure is known to be unstable and is stabilized only upon binding to the peptide. Here, stabilization of the complete structure is apparently achieved by a large enthalpic contribution. Often nonlocal, long-range effects are known to be coupled with peptide-protein association. The binding of antigenic peptides to MHC molecules, for example, leads to conformational changes that allow association and dissociation of additional proteins to the molecule, such as of β_2-microglobulin to the α chain of the MHC molecule, and others (Springer et al., 1998; Stern and Wiley, 1992; Zarutskie et al., 1999).

The rigidity of proteins upon binding is also observed in many cases of protein-protein interactions: Docking Benchmark 3.0 (Hwang et al., 2008) contains a predominant group of "easy" docking targets (88/124) that do not change significantly upon binding (an average of 0.83 Å interface Cα rmsd). A total of 50 of these 88 complexes are either enzyme-inhibitor complexes or antibody-antigens complexes—both optimized by evolution for tight binding.

An additional way to optimize binding can be seen in the significant over-representation of hydrogen bonds in peptide-protein interactions—although peptides bury a much smaller interface area, they form a number of hydrogen bonds across the interface that is comparable to protein-protein interfaces. This is accomplished by recruiting in addition to the side chains, also the backbone hydrogen bond donors and acceptors—after all, whereas the protein backbone is often rigid and embedded in a regular secondary structure, peptide backbones are flexible and amenable to more interactions with the partner.

When comparing the quality of packing at peptide-protein interfaces to that of protein-protein complexes (either by

Table 3. Peptide Binding Preference to Surface Pockets

Binding Mode	Size	Average Rank[a]	Median Rank[a]	No. of Pockets	Pocket Area[b]	Pocket Volume[c]
Large pockets	22	1.2 ± 0.5	1	17 ± 8	343 ± 196	340 ± 278
Knob-holes	40	4.3 ± 3.6	3	16 ± 8	40 ± 25	21 ± 16

Average values and standard deviations are indicated.
[a] Rank of peptide binding pocket on the protein monomer surface as ranked by the CASTp server).
[b] Pocket solvent-accessible surface area (Å^2) as calculated by the Lee and Richards algorithm (Lee and Richards, 1971).
[c] Pocket volume (Å^3) as calculated by the Lee and Richards algorithm.

a knowledge based potential, or by measuring the volume of voids at the interface normalized by the buried surface area), we again observe that peptides optimize the packing to avoid the entropic cost of voids.

Recognition of Binding Site on the Protein Monomer

Following our previous point, it seems that the protein interface is predefined and ready to accommodate the binding peptide. The peptide, however, has to "scan" the protein surface in an efficient way. We propose that binding involves a two-step mechanism for the recognition and binding of the peptide to its target site, similar to what has been proposed for protein-protein (Frisch et al., 2001; London and Schueler-Furman, 2008) and protein-DNA recognition (Slutsky et al., 2004). In a first step, an encounter complex is formed where the peptide searches for a large enough pocket on the protein surface into which it can bind. We showed that most of the peptides indeed bind in one of the largest pockets available on the protein's surface. Then, in a second step, the peptide is latched on by binding of a small number of hot spot residues that anchor it to the protein. Not surprisingly, the number of these hot spots is roughly correlated with the length of the peptide: for each 3 residues, another anchoring hot spot is added to provide enough binding energy (see Figure S3A). A previous study aimed at increasing the binding affinity of a peptide to its partner by designing a longer peptide (Sood and Baker, 2006). Although a longer peptide could indeed be designed, it was difficult to improve its overall binding affinity. This is in agreement with the observation that with increased length, more hot spot residues are needed to allow binding and to overcome a larger configurational entropic cost. Additional affinity that would surpass the increased configurational entropy of a longer peptide chain might be possible within the context of a preoptimized peptide-binding groove that would allow the introduction of several adjacent hotspots.

Implication for Computational Modeling

Finally, the present analysis is a stepping-stone for the development of a protocol for the computational structural modeling and manipulation of peptide-protein interactions. The parameters assessed in this study characterize existing peptide-protein interactions and can be used to evaluate the quality of models. In addition, this study offers a straight-forward protocol for computational modeling of peptide-protein complexes: (1) The finding that peptide bind in one of the largest pockets restricts the search space to a small number of well defined pockets—as opposed to protein-protein interactions, in which the potential binding site is elusive for prediction and the search space for initial encounter complexes is huge. This might improve on current approaches (Petsalaki et al., 2009) for predicting peptides binding regions. (2) The finding that the protein partner does not substantially change its structure upon binding can focus our computational resources toward the modeling of the flexibility of the peptide partner within pockets on the protein surface. We are currently developing such a modeling protocol that incorporates insights from the present study. The ultimate goal would be to predict binding peptides on a large scale based on structure, and by this to curate and extend interaction networks.

EXPERIMENTAL PROCEDURES

Creation of the PeptiDB Database

The PDB (version of 08.08.2008) (Berman et al., 2000) was queried for structures answering the following constraints: (1) the structure contains more than one chain, (2) one of the chains is between 5 and 15 amino acids long, (3) the structure does not contain DNA or RNA, and (4) it is a crystal structure with resolution $\leq$ 2.0 Å. This query extracted 866 putative complexes. These were filtered for complexes that do not include any heteroatom at the interface (defined as a nonwater heteroatom within 4 Å of the peptide), and are not covalently linked, reducing the size to 387 complexes. We clustered these complexes according to sequence similarity of the protein-binding partner, and created a set of nonredundant protein-peptide complexes with pairwise sequence identity below 70%. From each cluster, the structure presenting the best resolution was chosen. In the few cases where these structures contain clashes at the peptide-protein interface, they were replaced by another representative from the same cluster. The final database contains 103 complexes, of which 85 are composed of a single-chained protein partner, while the remaining 18 complexes involve a protein partner that consists of two chains.

A threshold of 70% sequence identity will still include entries of structurally similar proteins. Indeed, the main overrepresented folds in the data set are Retinoid-X-receptor (10 complexes), antibody-antigen (9), Sh3 domain (9), PDZ domain (5), and BIR domain (4). We therefore repeated our analysis on a truly nonredundant subset of 61 peptide-protein complexes that does not contain any protein pair from the same fold (defined based on CATH; Orengo et al., 1997; see Table S1).

We note that our data set differs from previous compilations of interactions that involve unstructured regions by the high resolution of the structures (e.g., no cutoff was applied in Meszaros et al., 2007), as well as the small size of the included peptides (partners of sizes up to 70 residues have been included in studies by Mohan et al., 2006 and Vacic et al., 2007).

We compared features in peptiDB with a representative data set of protein-protein interactions taken from Mintseris and Weng, 2003 (PPI), as well as a data set of IUP-protein interactions taken from Meszaros et al., 2007 (IUPPI). Because in these two data sets a very loose resolution threshold was applied (<3.25 Å or none), structures of low resolution can bias the analysis. We therefore repeated our analysis on a subset of PPI that contains structures with a similar resolution threshold as in peptiDB (64 out of 205).

Creation of Unbound Data Set

We used a BLAST (Altschul et al., 1990) search against the sequences in the PDB (pdbaanr of 09.11.2008; Wang and Dunbrack, 2003) to extract the

unbound monomer structures of the protein-binding partners in the peptiDB data set. In cases where the unbound monomer structure contains more than one chain, manual inspection validated that the structure does not bind a peptide at the same interface. The unbound monomer structure with the best resolution was selected to represent the complex in the unbound data set. If no free structure of the protein could be found, we also included structures of homologs (with an average of 85% and a minimum of 32% sequence identity). This allowed us to increase the number of cases where we could evaluate the difference between the bound and free monomer conformations. In principle, the difference in sequence may also introduce structural differences that are not necessarily related to the binding of the peptide, but this effect was observed to be marginal (see Results).

Extraction of Parameters in the Unbound Monomer Structure

The protein interface residues with the peptide were defined as residues within 4.5 Å of any peptide atom. For each pair of bound and unbound monomers, the interface residues were mapped onto the unbound structure (based on a global sequence alignment [Needleman and Wunsch, 1970]). The subset of interface residues was then used to superimpose the two structures and to calculate rmsd values using the *lsqman* program (Kleywegt, 1996).

Characterization of Protein-Peptide and Protein-Protein Interfaces

We used a range of different measures to characterize protein-protein and protein-peptide interfaces. The surface area buried upon binding (ASA); interface residues; interface atoms; distribution of polar, nonpolar, and neutral atoms at the interface; planarity; and eccentricity were calculated using the ProtorP server (Reynolds et al., 2009). In the following, a short description of each of these parameters is provided.

- *ASA* is the mean solvent-accessible surface area buried by the two chains (e.g., Jones et al., 2000).
- *Planarity* is the rmsd of all interface atoms to a best-fit plane through all interface atoms.
- *Eccentricity* measures the deviation of an ellipse from a circle.

$$Eccentricity = \sqrt{1 - \left(\frac{b^2}{l^2}\right)},$$

where b and l are the breadth and length of the ellipse, respectively. Eccentricity varies between 0 (perfect circle) and 1 (straight line).

- RosettaHoles score: RosettaHoles is a knowledge-based potential representing the packing environment of atoms in native structures (Sheffler and Baker, 2009). We calculated the average score over all interface atoms (defined as the atoms for which the packing score changes when the peptide is removed).
- *Hydrogen bonds* were calculated using the HBplus software with default parameters (McDonald and Thornton, 1994).
- *Salt bridges* are defined as opposite charges across the interface, within a distance cutoff of 6.0Å.
- *Bridging water molecules* are defined as water molecules that form a hydrogen bond with both chains at the interface.

Hot Spot Analysis

Hot spot residues were predicted by a computational alanine scan on each of the complexes in the peptiDB data set, using the Rosetta software, as reported by Kortemme et al. (2004). Hot spots were defined as residues that upon mutation to alanine are predicted to significantly decrease the binding energy ($\Delta\Delta G_{bind}$ > 1 kcal/mol; measured in Rosetta energy units). This threshold was verified against the ability to recover sequence motifs deposited in the ELM database (Puntervoll et al., 2003, and data not shown). To assure robustness, the analysis was repeated with a more stringent threshold of $\Delta\Delta G_{bind}$ > 1.5 kcal/mol, which yielded similar results (differences are indicated in the text). In addition, we verified that the results are independent of the computational alanine scanning procedure used: an analysis using the FoldX software (Guerois et al., 2002) produced similar results (computed changes in binding energy are correlated with R^2 = 0.6, whereby FoldX slightly overestimates $\Delta\Delta G_{bind}$ in comparison to Rosetta; Figure S3C).

Surface Pocket Analysis

Pockets on the protein monomers were defined by the CASTp server (Dundas et al., 2006), and ranked by their size (the peptide was removed for this analysis). The output of CASTp consists of a list of atoms participating in each pocket. The binding pocket was defined as the pocket that has the most contacts with the peptide (based on a 4.5 Å threshold). This analysis was performed on the bound protein structures, and similar results can be expected for the unbound structures, since these are very similar (see Results).

Statistical Tests

All statistical evaluations were performed with the R package (Ihaka and Gentleman, 1996). The Bonferroni correction was applied to account for multiple testing.

A two-sample *t*-test was applied to assess the statistical significance of the difference between feature values extracted from peptide-protein interfaces and protein-protein, or IUP-protein interfaces, respectively.

A χ^2 goodness-of-fit test was applied to evaluate the statistical significance of the difference between the number of hydrogen bonds and amino acid in protein-protein and peptide-protein interfaces.

The hypergeometric distribution was used to evaluate the statistical significance of the enrichment in peptide hot spots for specific amino acid types.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and two tables and can be found with this article online at doi:10.1016/j.str.2009.11.012.

ACKNOWLEDGMENTS

We would like to thank Joe Dundas for assisting with the CASTp analysis, and William Sheffler for the creation and assistance with RosettaHoles. We thank Monika Fuxreiter for sharing with us the amino acid composition data. All structural images have been produced using PyMOL software (DeLano, 2002). This work was supported by the Israel Science Foundation, founded by the Israel Academy of Science and Humanities (grant number 306/6), and the German-Israeli Foundation for Scientific Research and Development (Young Scientists Program 2159-1709.3/2006).

Received: July 21, 2009
Revised: November 8, 2009
Accepted: November 11, 2009
Published: February 9, 2010

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Bahadur, R.P., Chakrabarti, P., Rodier, F., and Janin, J. (2004). A dissection of specific and non-specific protein-protein interfaces. J. Mol. Biol. *336*, 943–955.

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.

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

Ceol, A., Chatraryamontri, A., Santonico, E., Sacco, R., Castagnoli, L., and Cesareni, G. (2007). DOMINO: a database of domain-peptide interactions. Nucleic Acids Res. *35*, D557–D560.

Clackson, T., and Wells, J.A. (1995). A hot spot of binding energy in a hormone-receptor interface. Science *267*, 383–386.

Cohen, M., Reichmann, D., Neuvirth, H., and Schreiber, G. (2008). Similar chemistry, but different bond preferences in inter versus intra-protein interactions. Proteins *72*, 741–753.

Dall'Acqua, W., Goldman, E.R., Eisenstein, E., and Mariuzza, R.A. (1996). A mutational analysis of the binding of two different proteins to the same antibody. Biochemistry *35*, 9667–9676.

DeLano, W.L. (2002). The PyMOL molecular graphics system (http://www.pymol.org).

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. *34*, W116–W118.

Fong, J.H., Shoemaker, B.A., Garbuzynskiy, S.O., Lobanov, M.Y., Galzitskaya, O.V., and Panchenko, A.R. (2009). Intrinsic disorder in protein interactions: insights from a comprehensive structural analysis. PLoS Comput. Biol. *5*, e1000316.

Frisch, C., Fersht, A.R., and Schreiber, G. (2001). Experimental assignment of the structure of the transition state for the association of barnase and barstar. J. Mol. Biol. *308*, 69–77.

Fuxreiter, M., Tompa, P., and Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. Bioinformatics *23*, 950–956.

Glaser, F., Steinberg, D.M., Vakser, I.A., and Ben-Tal, N. (2001). Residue frequencies and pairing preferences at protein-protein interfaces. Proteins *43*, 89–102.

Guerois, R., Nielsen, J.E., and Serrano, L. (2002). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J. Mol. Biol. *320*, 369–387.

Hayouka, Z., Rosenbluh, J., Levin, A., Loya, S., Lebendiker, M., Veprintsev, D., Kotler, M., Hizi, A., Loyter, A., and Friedler, A. (2007). Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium. Proc. Natl. Acad. Sci. USA *104*, 8316–8321.

He, B., Minges, J.T., Lee, L.W., and Wilson, E.M. (2002). The FXXLF motif mediates androgen receptor-specific interactions with coregulators. J. Biol. Chem. *277*, 10226–10235.

Ho, B.K., and Dill, K.A. (2006). Folding very short peptides using molecular dynamics. PLoS Comput. Biol. *2*, e27.

Hwang, H., Pierce, B., Mintseris, J., Janin, J., and Weng, Z. (2008). Protein-protein docking benchmark version 3.0. Proteins *73*, 705–709.

Ihaka, R., and Gentleman, R. (1996). R: A language for data analysis and graphics. J. Comput. Graph. Stat. *5*, 299–314.

Jones, S., Marin, A., and Thornton, J.M. (2000). Protein domain interfaces: characterization and comparison with oligomeric protein interfaces. Protein Eng. *13*, 77–82.

Killian, B.J., Kravitz, J.Y., Somani, S., Dasgupta, P., Pang, Y.P., and Gilson, M.K. (2009). Configurational entropy in protein-peptide binding: computational study of Tsg101 ubiquitin E2 variant domain with an HIV-derived PTAP nonapeptide. J. Mol. Biol. *389*, 315–335.

Kleywegt, G.J. (1996). Use of non-crystallographic symmetry in protein structure refinement. Acta Crystallogr. D Biol. Crystallogr. *52*, 842–857.

Kortemme, T., Kim, D.E., and Baker, D. (2004). Computational alanine scanning of protein-protein interfaces. Sci. STKE *2004*, pl2.

Lee, B., and Richards, F.M. (1971). The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. *55*, 379–380.

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

Lockless, S.W., and Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. Science *286*, 295–299.

London, N., and Schueler-Furman, O. (2008). Funnel hunting in a rough terrain: learning and discriminating native energy funnels. Structure *16*, 269–279.

McDonald, I.K., and Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. *238*, 777–793.

Meszaros, B., Tompa, P., Simon, I., and Dosztanyi, Z. (2007). Molecular principles of the interactions of disordered proteins. J. Mol. Biol. *372*, 549–561.

Mintseris, J., and Weng, Z. (2003). Atomic contact vectors in protein-protein recognition. Proteins *53*, 629–639.

Mohan, A., Oldfield, C.J., Radivojac, P., Vacic, V., Cortese, M.S., Dunker, A.K., and Uversky, V.N. (2006). Analysis of molecular recognition features (MoRFs). J. Mol. Biol. *362*, 1043–1059.

Neduva, V., and Russell, R.B. (2006). Peptides mediating interaction networks: new leads at last. Curr. Opin. Biotechnol. *17*, 465–471.

Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. *48*, 443–453.

Nooren, I.M., and Thornton, J.M. (2003). Structural characterisation and functional significance of transient protein-protein interactions. J. Mol. Biol. *325*, 991–1018.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. (1997). CATH–a hierarchic classification of protein domain structures. Structure *5*, 1093–1108.

Parthasarathi, L., Casey, F., Stein, A., Aloy, P., and Shields, D.C. (2008). Approved drug mimics of short peptide ligands from protein interaction motifs. J. Chem. Inf. Model. *48*, 1943–1948.

Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. Science *300*, 445–452.

Petsalaki, E., and Russell, R.B. (2008). Peptide-mediated interactions in biological systems: new discoveries and applications. Curr. Opin. Biotechnol. *19*, 344–350.

Petsalaki, E., Stark, A., Garcia-Urdiales, E., and Russell, R.B. (2009). Accurate prediction of peptide binding sites on protein surfaces. PLoS Comput. Biol. *5*, e1000335.

Puntervoll, P., Linding, R., Gemund, C., Chabanis-Davidson, S., Mattingsdal, M., Cameron, S., Martin, D.M., Ausiello, G., Brannetti, B., Costantini, A., et al. (2003). ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. Nucleic Acids Res. *31*, 3625–3630.

Reynolds, C., Damerell, D., and Jones, S. (2009). ProtorP: a protein-protein interaction analysis server. Bioinformatics *25*, 413–414.

Rodier, F., Bahadur, R.P., Chakrabarti, P., and Janin, J. (2005). Hydration of protein-protein interfaces. Proteins *60*, 36–45.

Sheffler, W., and Baker, D. (2009). RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. Protein Sci. *18*, 229–239.

Slutsky, M., Kardar, M., and Mirny, L.A. (2004). Diffusion in correlated random potentials, with applications to DNA. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *69*, 061903.

Sood, V.D., and Baker, D. (2006). Recapitulation and design of protein binding peptide structures and sequences. J. Mol. Biol. *357*, 917–927.

Springer, S., Doring, K., Skipper, J.C., Townsend, A.R., and Cerundolo, V. (1998). Fast association rates suggest a conformational change in the MHC class I molecule H-2Db upon peptide binding. Biochemistry *37*, 3001–3012.

Stanfield, R.L., and Wilson, I.A. (1995). Protein-peptide interactions. Curr. Opin. Struct. Biol. *5*, 103–113.

Stein, A., and Aloy, P. (2008). Contextual specificity in peptide-mediated protein interactions. PLoS ONE *3*, e2524.

Stern, L.J., and Wiley, D.C. (1992). The human class II MHC protein HLA-DR1 assembles as empty alpha beta heterodimers in the absence of antigenic peptide. Cell *68*, 465–477.

Unal, E.B., Gursoy, A., and Erman, B. (2009). Conformational energies and entropies of peptides, and the peptide-protein binding problem. Phys. Biol. *6*, 36014.

Vacic, V., Oldfield, C.J., Mohan, A., Radivojac, P., Cortese, M.S., Uversky, V.N., and Dunker, A.K. (2007). Characterization of molecular recognition features, MoRFs, and their binding partners. J. Proteome Res. *6*, 2351–2366.

Wang, G., and Dunbrack, R.L., Jr. (2003). PISCES: a protein sequence culling server. Bioinformatics *19*, 1589–1591.

Xu, D., Tsai, C.J., and Nussinov, R. (1997). Hydrogen bonds and salt bridges across protein-protein interfaces. Protein Eng. *10*, 999–1012.

Zarutskie, J.A., Sato, A.K., Rushe, M.M., Chan, I.C., Lomakin, A., Benedek, G.B., and Stern, L.J. (1999). A conformational change in the human major histocompatibility complex protein HLA-DR1 induced by peptide binding. Biochemistry *38*, 5878–5887.

Zhao, L., and Chmielewski, J. (2005). Inhibiting protein-protein interactions using designed molecules. Curr. Opin. Struct. Biol. *15*, 31–34.

Zvelebil, M.J., and Thornton, J.M. (1993). Peptide-protein interactions: an overview. Q. Rev. Biophys. *26*, 333–363.

Note Added in Proof

Two recent additional relevant studies by Vanhee et al. have just been published about the structural basis of peptide-protein interactions. One study describes PepX, a structural database of peptide-protein complexes (Vanhee, P., Reumers, J., Stricher, F., Baeten, L., Serrano, L., Schymkovitz, J., and Rousseau, F. [2010]. PepX: a structural database of non-redundant protein-peptide complexes. Nucleic Acids Res. *38,* D545-D551), and another study demonstrates that peptide-protein complexes use the same set of structural motifs as protein monomer structures (Vanhee, P., Stricher, F., Baeten, L., Verschueren, E., Lenaerts, T., Serrano, L., Rousseau, F., and Schymkovitz, J. [2009]. Protein-peptide interactions adopt the same structural motifs as monomeric protein folds. Structure *17*, 1128-1136).