## Title
Modeling beta-sheet peptide-protein interactions: Rosetta FlexPepDock in CAPRI rounds 38-45



## Journal
Proteins Structure Function and Bioinformatics, 88(8)

## ISSN
0887-3585

## Authors
Khramushin, Alisa
Marcu, Orly
Alam, Nawsad
et al.

## Publication Date
2020-08-01

## DOI
10.1002/prot.25871

Peer reviewed

HHS Public Access

Author manuscript

*Proteins*. Author manuscript; available in PMC 2020 October 07.

Published in final edited form as:

*Proteins*. 2020 August ; 88(8): 1037–1049. doi:10.1002/prot.25871.

# Modeling beta-sheet peptide-protein interactions: Rosetta FlexPepDock in CAPRI rounds 38-45

**Alisa Khramushin**[1], **Orly Marcu**[1], **Nawsad Alam**[1], **Orly Shimony**[1], **Dzmitry Padhorny**[2,3], **Emiliano Brini**[3], **Ken A. Dill**[3,4,5], **Sandor Vajda**[6,7], **Dima Kozakov**[2,3], **Ora Schueler-Furman**[1]

[1]Department of Microbiologyand Molecular Genetics, Institute for Medical Research Israel-Canada, Faculty of Medicine, The Hebrew University, Jerusalem, Israel

[2]Department of Applied Mathematics and Statistics, Stony Brook University, New York, New York

[3]Laufer Center for Physical and Quantitative Biology, Stony Brook University, New York, New York

[4]Department of Physics and Astronomy, Stony Brook University, New York, New York

[5]Department of Chemistry, Stony Brook University, New York, New York

[6]Department of Biomedical Engineering, Boston University, Boston, Massachusetts

[7]Department of Chemistry, Boston University, Boston, Massachusetts

## Abstract

Peptide-protein docking is challenging due to the considerable conformational freedom of the peptide. CAPRI rounds 38-45 included two peptide-protein interactions, both characterized by a peptide forming an additional beta strand of a beta sheet in the receptor. Using the *Rosetta FlexPepDock* peptide docking protocol we generated top-performing, high-accuracy models for targets 134 and 135, involving an interaction between a peptide derived from L-MAG with DLC8. In addition, we were able to generate the only medium-accuracy models for a particularly challenging target, T121. In contrast to the classical peptide-mediated interaction, in which receptor side chains contact both peptide backbone and side chains, beta-sheet complementation involves a major contribution to binding by hydrogen bonds between main chain atoms. To establish how binding affinity and specificity are established in this special class of peptide-protein interactions, we extracted *PeptiDBeta*, a benchmark of solved structures of different protein domains that are bound by peptides via beta-sheet complementation, and tested our protocol for global peptide-docking *PIPER-FlexPepDock* on this dataset. We find that the beta-strand part of the peptide is sufficient to generate approximate and even high resolution models of many interactions, but inclusion of adjacent motif residues often provides additional information necessary to achieve high resolution model quality.

Correspondence: Ora Schueler-Furman, Department of Microbiology and Molecular Genetics, Institute for Medical Research Israel-Canada, Faculty of Medicine, The Hebrew University, Jerusalem, Israel., ora.furman-schueler@mail.huji.ac.il; Dima Kozakov, Department of Applied Mathematics and Statistics, Stony Brook University, New York, NY. midas@laufercenter.org.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**Keywords**

beta sheet interactions; CAPRI; FlexPepDock; high-resolution modeling; peptide docking; peptide-protein interactions; Rosetta

## 1 | INTRODUCTION

Protein-peptide interactions are abundant in the cell and participate in multiple cellular processes such as regulation and cell-signaling. As these interactions are often transient and weak they are difficult targets for crystallography and NMR. Therefore, detailed modeling of those interactions is crucial for our understanding of different biological mechanisms. However, these interactions are also particularly difficult to model since the peptide often does not form a stable defined structure prior to binding, making it necessary to sample considerable conformational space of both internal degrees of freedom of the peptide, as well as its rigid body orientation relative to the receptor to find the correct conformation. A variety of approaches have been developed to address these challenges efficiently.[1]

We have previously developed a series of protocols that allow to generate highly accurate models of peptide-mediated interactions. *Rosetta FlexPepDock* was the first protocol to explicitly model full conformational freedom of the peptide backbone as well as side chains, allowing for the accurate refinement of approximate models of a peptide-protein interaction (generated, for example, from a homolog template, or based on experimental constraints).[2] These models can be used to identify substrates for a given peptide binding receptor or peptide-modifying enzyme (as demonstrated for example, on histone deacetylases[3] and more). In ab initio *Rosetta FlexPepDock* sampling of the peptide conformation space is increased by use of fragment libraries, allowing to generate a peptide-protein structure starting from a peptide of arbitrary conformation positioned within a given binding site.[4] While these protocols have significantly impacted the atom-accuracy modeling of peptide-protein interactions for which information about the binding site on the receptor is known, further development was necessary to achieve successful fully blind peptide docking, where only the peptide sequence and the receptor structure (but not the binding site) are given. A breakthrough came from the observation that a structure similar to that of the peptide can often be found among fragments extracted from protein monomer structures, based on similarity in sequence motifs and (predicted) secondary structure preference.[5,6] Therefore, similar to ab initio folding, that can be simplified by combining such fragments into a final structure, followed by high resolution refinement,[7] global peptide docking can be achieved by rigid body docking of these fragments, followed by high resolution peptide docking. We developed two protocols for global docking using *PIPER/CLUSPRO*[8] for the rigid body docking step: (a) *PeptiDock*[5] consists of a first step in which a peptide motif is extended and pruned according to predefined rules, until it produces a couple of hundred fragment hits in the Protein Data Bank (PDB[9,10]). These fragments are then clustered, docked, and minimized, producing routinely structures within 4 Å RMSD of the native peptide conformation; (b) *PIPER-FlexPepDock*[6] consists of three steps: First, a fragment library is compiled using a modified version of the Rosetta fragment picker,[11] second, these fragments are rigid body docked using *PIPER/CLUSPRO*,[12] and third, the top-scoring models are

further refined by *Rosetta FlexPepDock*,[2] resulting in near-native protein complex structures within 2-3 Å RMSD in the predominant number benchmarked cases. Notably, for both approaches performance is optimal when the peptide binding segment, or motif, is known, or correctly identified.

Peptide-protein docking is a sub-category of the more mature protein-protein docking field. Aiming to assess the performance of different computational approaches for modeling protein-protein interactions, CAPRI (Critical Assessment of PRediction of Interactions) releases regularly information about to be published NMR and X-ray structures of protein-protein complexes, to be used as targets for blind prediction. Models submitted by the participants are assessed and categorized as high, medium or acceptable accuracy according to defined CAPRI criteria[13,14] (note that for peptide-protein interactions, slightly different criteria apply[15]). Since its inception in 2002, the CAPRI experiment has significantly impacted the docking field. It has engaged tens of different groups, involved a wide variety of protocols, and encouraged exchange of ideas, approaches and experience, accelerating the advance of ever improving docking protocols.[13] Different types of protein interactions have been added along the years, including binding to nucleotides,[16] sugars,[14] and peptides.[17]

CAPRI rounds 38-45 included two peptide-protein interactions: The first target, T121 of round 38 involved a particularly challenging interaction with considerable receptor reorganization, for which we were the only groups to generate medium accuracy models (as defined by the CAPRI criteria[15]). However, since target information and structure are still unpublished, we will not go into details of this modeling challenge. The second set of targets, T134-135 of round 44, involved the interaction between a peptide derived from the cytoplasmic segment of the mouse myelin associated glycoprotein (L-MAG) bound to dynein light chain subunit 8 (DLC8),[18] for which we generated high-accuracy models. The round consisted of two stages: T134 involved the identification of the binding motif within the cytoplasmic segment of the MAG protein, while subsequent T135 involved modeling of the interaction with the given binding motif. For both targets, we were able to generate the best models among all CAPRI submissions. We describe here the strategies that we applied to successfully model these peptide-protein targets using *Rosetta FlexPepDock* (a summary scheme is shown in Figure 1).

Since both targets involve an interaction in which a peptide contributes an additional strand of a beta sheet in the receptor, we extended our study to a general analysis of beta-sheet complementing peptide-protein interactions. For this, we extracted *PeptiDBeta*, a benchmark of peptide-protein complex structures for which both the bound and free receptor structure are available. The dataset was generated using an automatically updated version of *PeptiDB*[19] (*AutoPeptiDB;* unpublished results).

A long-lasting question about interactions mediated by short linear motifs (SLiMs) is how binding specificity is achieved, given the low information content of many of these motifs and their frequency in protein sequences. In turn, many competing binding sites might be available on the receptor surface. Flanking regions around the peptide motif play an important role for the determination of binding specificity.[20,21] In this study, we use *PeptiDBeta* and our blind peptide docking protocol *PIPER- FlexPepDock* to investigate how

binding preference for a specific site on the receptor surface is achieved for beta-sheet complementing peptides. We find that for most cases the short, beta strand forming peptide stretch provides enough information to recognize the binding site on the receptor, and that additional residues often allow generation of atom-resolution, high accuracy models of an interaction (within 2-2.5 Å RMSD, similar to general performance of *PIPER-FlexPepDock*).[6]

## 2 | METHODS

### 2.1 | Peptide docking using *Rosetta FlexPepDock*

Peptide docking simulations were performed using the *Rosetta FlexPepDock* protocol, unless indicated otherwise. Detailed information about the runline commands and versions used are provided in the Supporting Information. Rosetta version 3.9, and energy function ref2015[22] were used.

**FlexPepDock refinement:** *FlexPepDock* refinement has been described previously: In brief, starting from an approximate conformation of a peptide-protein complex, the peptide is optimized by iterative sampling of rigid body and internal backbone peptide degrees of freedom, with periodic repacking of all interface side chains, starting with almost exclusively attractive forces and gradually ramping back repulsion, to allow a smooth transition to nearby energy minima.[2]

**FlexPepBind threading:** *FlexPepBind*[3] uses a template structure of a peptide-protein interaction to thread a list of peptides each onto the template and refine each peptide using *FlexPepDock.*[2] The structure of the DLC8 protein solved in complex with the Nek9 peptide (PDBid 3zke[23]) was used as template. The 57 residue C-terminal part of the MAG-L mouse protein (sequence: KYESEKRLGSERRLLGLRGES PELDLSYSHSDLGKRPTKDSYTLTEELAEYAEIRVK) was split to all overlapping 11mer peptides. Each 11mer was threaded onto the Nek9 peptide conformation, using the *Rosetta fixbb design*[24] protocol and then minimized using the Rosetta *FlexPepDock* application with the “minimization only” option (ie, no random perturbations were applied).

### 2.2 | Global blind peptide docking using *PIPER-FlexPepDock*

Global docking with no prior information about neither binding site nor peptide conformation was performed using the *PIPER-FlexPepDock* protocol.[6] In brief, the peptide conformation is represented as an ensemble of fragments extracted from the PDB, based on sequence and (predicted) secondary structure using the Rosetta Fragment picker (with the vall 2011 fragment library).[11] These fragments are mutated to the target peptide sequence. Fifty fragments are rigid body docked onto the receptor protein using the *PIPER* rigid body docking program.[12] The top 250 models for each fragment are then further refined using the *Rosetta FlexPepDock* protocol, including receptor backbone minimization, and top-scoring models are clustered. The protocol is freely available for noncommercial use as an online server: https://piperfpd.furmanlab.cs.huji.ac.il.

### 2.3 | Modeling round 38 target T121 using refinement with modeling employing limited data (MELD) molecular dynamics simulations

The Kozakov-Vajda-Dill groups applied for the modeling of T121 also an additional strategy which involved the refinement of the docked peptide-protein complex structures using MELD × MD simulations. MELD has been previously discussed in detail.[25,26] Briefly, MELD uses sparse, ambiguous, and uncertain information to reduce the search space of physics based simulations. This is done using a Bayesian approach implemented using "smart" flat-bottom potential springs, which do not bias the energy if the information is satisfied. Relative populations of MELD structures can serve as a proxy of relative free energies. To allow MD to overcome the high energy barrier between different zones of the phase space that the smart springs create, a Hamiltonian and temperature replica exchange protocol is used. Previously, MELD has been successfully applied to the problems of protein folding,[27–29] peptide docking,[30,31] and protein dimer structure predictions.[32] More details are provided in the Supporting Information.

### 2.4 | Determination of interface hotspots using computational alanine scanning

Computational alanine scanning was performed using the *Robetta* alanine scanning protocol, as described by Kortemme et al.[33] In this protocol, after the protein-peptide interface is defined, every residue at the interface is mutated to alanine one by one and the energy gaps caused by the mutations are estimated. No backbone or side-chain perturbations are allowed upon mutation, as it has been shown that for mutations to alanine, a conservative protocol performs best.[34,35] We applied a local version of the protocol implemented in the *Robetta* alanine scanning server.[36] To remove clashes, crystal structures were minimized prior to the calculation.

### 2.5 | *PeptiDBeta*, a benchmark for beta-strand peptide-protein interactions

The *PeptiDB* set of structures of peptide-protein complexes has been used widely to calibrate peptide docking protocols.[19,37–40] In order to streamline the tedious process of dataset curation (e.g., to remove interactions where the peptide conformation is significantly influenced by crystal contacts), we developed *AutoPeptiDB*, an automatic scheme that generates periodic updates of *PeptiDB* (in preparation). *AutoPeptiDB* (version 6/2018) was the starting point for the extraction of *PeptiDBeta*, a dataset of protein-peptide interactions, in which peptides extend existing protein beta-sheets. To prevent potential bias for overrepresented structures in the PDB, we compiled a domain nonredundant set defined according to the classification in the database of evolutionary classification of protein domains (ECOD).[41] Only single chain receptor—peptide interactions were retained, and complexes in which peptide temperature factor was greater than 35, as well as covalently bound peptides were removed. To isolate beta-complementing complexes from this dataset, we filtered out complexes with ECODs corresponding to all-helical proteins, and retained only complexes in which peptides adopt beta-strand conformation, as estimated by the *Define Secondary Structure of Proteins* algorithm (DSSP).[42] Protein complexes with beta-strand peptide longer than 10 amino acids were removed. Out of 75 structures remaining after the filtering, we generated the final set of 14 peptide-protein complexes for which also the free receptor structure has been solved (Table 4). We also include a set of 5 peptide-PDZ

domain complexes to assess the variability of our results among complexes involving the same domain (Table S2).

### 2.6 | Delineation of the span of the peptide segment to be docked

The peptide motif to be modeled for the *PeptiDBeta* benchmark was defined using two different criteria: (a) Definition of the beta-strand segment of the peptide, based on DSSP calculation of backbone hydrogen bonds, and (b) definition of extended motif, based on the *PeptiDock* motif definition approach, as described previously.[5] In brief, a starting peptide motif is extended/restricted according to nearby amino acids that are frequently involved in known binding motifs, until a search in the PDB using the motif results in a couple of hundreds of fragments.

## 3 | RESULTS

Figure 1 summarizes the main approaches applied in this study to peptide docking in CAPRI rounds 38-45. These rounds included three peptide-protein complex targets: Targets T121 in round 38 and T134 and 135 in round 44. We describe round 44 in detail and only briefly mention T121, as the structure has not been published yet, and therefore no details can be revealed. We then present an in depth analysis of global peptide docking on *PeptiDBeta*, a benchmark for beta-sheet complementing peptide-protein interactions, focusing on the influence of peptide length and motif definition on binding specificity.

### 3.1 | Round 44–Template based peptide threading and docking refinement

T134/T135 consist of the interaction between DLC8 (a dimer) with a 12-residue peptide isolated from the 57-residue C-terminal cytoplasmic segment of mouse L-MAG.[18] The complex was first crystallized with DLC8 bound to a 57-residue c-terminal segment of L-MAG (PDB id 6gzj), and subsequently with DLC8 bound to a 12-residue peptide extracted from the same above segment (PDB id 6gzl), with the resulting complexes adopting very similar structures. The challenge for T134 was to predict which 12-residue peptide within the longer segment binds to DLC8, as well as to model the structure of the resulting complex. For T135, the predictors were given the sequence of the 12 residue peptide for which the DLC8-bound structure had been determined. The main challenge for this target was to produce a highly accurate model for the protein-peptide complex.

Our modeling strategy involved the use of a solved structure of DLC8 bound to a different peptide as template, assuming that the MAG-derived peptide would bind in the same site. Application of the *Rosetta FlexPepBind*[3] threading protocol identified the binding peptide within the 57-residue segment (Figure 2A). Using our protocol for high-resolution peptide docking *Rosetta FlexPepDock* on the predicted binding motif, we were able to generate the top-ranking, high accuracy models for these targets (Table 1).

The suggested motif and the binding site were subsequently validated with an unbiased simulation performed with our global peptide docking protocol *PIPER-FlexPepDock*, for which no information about binding site or peptide conformation was provided. The simulation resulted in top-scoring models of high accuracy (Figure 2B,C and Table 2).

### 3.2 | Target 134: Identifying the peptide binding motif in the 57-residue C-terminal tail of MAG

**3.2.1 | Template selection and binding motif identification—**In order to proceed with template-based modeling of the complex, we considered 19 structures of DLC8 solved in complex with different peptides. One of the structures had been solved with the Nek9 peptide (PDB id: 3zke,[23] sequence: VGMH**S**KG**TQT**A) which contains a motif in accordance with the consensus sequence discussed in *Bodor* et al[43]: $[DS]_{-4}\underline{K}_{-3}X_{-2}[TVI]_{-1}\underline{Q}_{o}[TV]_{1}[DE]_{2}$. In the 57-residue fragment provided for prediction we noticed a similar motif (**D**S**T**L**T**), suggesting possible binding at the same site. Based on this observation we chose this structure as a template for the docking procedure. We noted however that our sequence did not confer to the reported consensus that contains a glutamine rather than a leucine amino acid in between the two threonines and a serine instead of lysine at position—3 (underlined residues).

With this structure as template, we set out to identify the binding peptide within the provided 57 residue c-terminal fragment. Using *Rosetta FlexPepBind*, we threaded each overlapping window peptide of 11 residues onto the Nek9 peptide: Side chains were mutated while keeping the peptide backbone fixed, using the *Rosetta fixbb design* application[24] (details of this and following simulations are described in Methods, and command line parameters are provided in the Supporting Information). After threading, a short minimization was performed using *Rosetta FlexPepDock*, allowing for optimization of all peptide atoms, as well as receptor side chain atoms. The resulting energies of the different peptide sequences were used to identify the motif. Both Rosetta reweighted (Figure 2A) and interface scores (-Figure S1) identified the same motif: RPTK**D**SY**TLT**E.

### 3.3 | High-resolution model refinement

Since the peptide in the template that we chose (3zke) is only 11 residues long, while we were asked to model a 12 amino acid peptide, we added one residue at each terminus, resulting in the two peptides **K**RPTKDSYTLTE and RPTKDSYTLTE**E**. For both of these complexes we performed three different simulations: (a) *FlexPepDock* refinement, (b) ab initio *FlexPepDock* with fragments for the first and last 3mer, respectively, and (c) full ab initio *FlexPepDock*, resulting in a total of six distinct simulations. We selected the models for submission from the top-scoring resulting models from these runs (See Table 1A). Overall these models converged, with an average backbone RMSD between the top 10 submitted models of 0.86 Å. Six were high and the rest medium accuracy models.

For the ranking of the different models, we complemented interface and reweighted scores with additional parameters calculated using the *Rosetta Interface analyzer* application.[44] We looked for a high number of hydrogen bonds between the partners, a low number of unsatisfied buried hydrogen bond acceptors and donors, and favorable packing (as represented by the *packstat* measure that penalizes for voids smaller than a water molecule within the protein complex). However, none of these measures proved to be helpful for the distinction of high-accuracy models from the rest in a posterior analysis (data not shown).

### 3.4 | Binding site confirmation with fully blind docking protocol

To further strengthen confidence in the predicted binding mode, we also performed an unbiased docking simulation of the peptide identified by threading (RPTKDSY**TLT**E), using our blind peptide docking protocol *PIPER-FlexPepDock*,[6] without including any information about the binding site on the receptor, nor about the peptide conformation. The top-scoring models based on both reweighted and interface scores recovered the binding mode with high accuracy (backbone RMSD 0.58 Å; Figure 2B,C).

### 3.5 | Target 135

For T135 the sequence of the 12mer was given: PTKDSYTLTEEL. Starting with our top submitted model for T134, we added the leucine residue at the C-terminal end of the model peptide and removed the N-terminal arginine (RPTKDSYTLTEE**L**). We then performed two *Rosetta FlexPepDock* ab initio runs, including information about side-chain rotamers of either the top ranked submitted model, or the initial template (3zke). The models showed a wide variety of C-terminal peptide conformations (Figure 3A), but overall fell into two clusters, in which the C-terminal leucine pointed up or down. When leucine points upwards it creates hydrophobic interactions with the leucine of the TLT motif. Leucine pointing downwards interacts with the hydrophobic pocket on the receptor surface. Seven of our models were of high quality (Table 1B).

### 3.6 | Analysis of the results

For target 134 we identified the correct binding motif and generated high-accuracy models. However, the crystal structure did not include the N-terminal arginine that we included in the motif, but included an additional leucine in its C-terminus. Most of our models for T135 do not converge in the C-terminal region, unlike the rest of the peptide, for which the hydrogen-bonding pattern is conserved in all the models. This observation lead us to the assumption that this part of the peptide might be disordered. As the crystal structure was published, we analyzed the environment of the structure in the crystal and found that indeed the only interaction in which this C-terminal leucine participates is an artificial crystal contact (Figure 3B).

An accurate structural model can provide important information about an interaction, by identifying interface hotspots—residues that contribute critically to binding.[45] This information may be used to generate mutations that will abolish an interaction, and thereby can serve for the experimental assessment of the functional importance of that interaction. For a model to be useful, we would expect it to predict the same interface hotspots as would an experimentally solved structure. While the MAG peptide-DLC8 interaction is mostly stabilized by backbone interactions that are not affected directly by mutation to alanine, there are a number of peptide positions whose side chains participate in hydrogen bonding with the receptor. We used *Robetta alanine scanning*[33,36] to identify interface hotspot residues both on the solved as well as modeled complex structures (Figure 3C, D; summarized in Table S1). On the peptide side, the crystal structure suggests three hotspots: $Y_6$.$L_8T_9$. On the receptor side, residues $Y_{65}$ & $T_{67}$ on the neighboring beta strand engage with the peptide $D_4$ side chain, while residues $Y_{75}$ & $Y_{77}$ of an underlying sheet engage with the peptide backbone at the motif positions $L_8T_9$. However, none of the residues located in

the helix of the second monomer that interacts with $L_8$ are classified as hotspots in the crystal structure. In contrast, our models predominantly highlight $D_4…L_8T_9$ as hotspots, overlapping better with the previously reported motif ($D_4Kx_6TQ_8T_9E$, numbered accordingly), and helical residues $I_{34}$ and $K_{36}$ are now highlighted as hotspots. Thus, critical leucine residue $L_8$ from the TLT motif that interacts with the helix of the second DLC8 monomer in the dimer, as well as the following threonine $T_9$ were predominantly classified as major contributors to binding, while the side chain of tyrosine residue $Y_6$ was misplaced in our models and therefore only the crystal structure showed substantial contribution to binding for this residue.

### 3.7 | Determinants of binding site specificity

Interestingly, the blind docking simulation generates significantly better models when the docked peptide includes the N-terminal arginine residue $R_0$ (Table 2). Comparison between the energy landscapes of the **R**PTKDSYT**L**TE and PTKDSYT**L**TE peptides (Figure 2B and Figure S2A) reveals more alternative low-energy regions in the energy landscape for the 10mer. In our model $R_0$ creates a strong ionic bond with receptor $E_{71}$ (Figure 3E), which might be critical for the peptide's site recognition. Moreover, the docking simulation of PTKDSYT**L**TE**EL** solved in the crystal structure that includes the C-terminal residues $E_{11}L_{12}$, but does not include the N-terminal arginine, did not succeed to recognize the near-native conformation, probably due to reduced fragment quality (Table 2 and Figure S2B). Finally, while the TLT leucine $L_8$ is clearly a "hotspot" and is considered to be part of the motif, global docking of a peptide that does not include this key residue, RPTKDSYT, still successfully recovers a near-native structure (Figure S2C). The lack of clear parameters that could bode for successful global docking suggests emerging guidelines that include the docking of different peptide segments in search for the correct orientation, where additional residues may help increase specificity, but also reduce the quality of representation of the fragment set used in the rigid docking step.

### 3.8 | Round 38, T121—A challenging target involving considerable rearrangement

The second peptide-protein interaction to be modeled in CAPRI rounds 38-45, round 38 target T121 challenged us due to considerable conformational rearrangement of the receptor upon binding of the peptide. We were the only groups to identify the correct orientation of the beta-sheet extending peptide in this target, and thus our models were the only ones to be scored as medium accuracy (Table 3). Unfortunately, the study describing the experimental NMR structure is still not published and the structure has not been released yet. We therefore only briefly outline the two approaches that were successful for this target, without providing details.

The approach taken by the Furman lab generated the top-quality, medium-accuracy model among all T121 submissions. It involved first the definition of a short binding motif within the peptide partner, based on homologous interactions. Global docking using *PIPER-FlexPepDock* of this peptide on a homolog bound structure positioned the peptide next to the suspected beta sheet on the receptor. This initial conformation was copied onto the unbound structure of the target receptor, and extensive optimization of the receptor structure was performed using the *Rosetta FastRelax*[46] protocol, in the presence of the short motif peptide,

to open up the binding pocket for extension. The 10 top-scoring relaxed structures were then used as receptors for another *PIPER-FlexPepDock* global docking run, in which surprisingly the peptide orientation was inverted. This new structure was further refined using *Rosetta FlexPepDock.* The four top-scoring docking models generated by *PIPER-FlexPepDock* on the relaxed structures all contained peptides in the reverse orientation, thereby emphasizing that this orientation could have been identified based on energy criteria (had we not manually selected other models with a canonical peptide orientation instead, and ranked the correct orientation as last submission).

The approach taken by the Kozakov-Vajda-Dill labs generated medium accuracy models using an alternative receptor starting structure: Instead of extensive relaxation of the receptor structure, a homology model of the bound protein conformation was generated starting from the bound homolog template structure (using MODELER[47]). This receptor conformation was used as input to *ClusPro-Peptidock* global peptide docking (freely available for noncommercial use as https://peptidock.cluspro.org[5]), which resulted in medium accuracy models of the interactions. Unfortunately, these server-generated models were excluded from the final CAPRI ranking since the peptide was too short, violating the CAPRI sequence identity criterion (the docked peptide segment was defined based on a motif identified from sequence analysis of homologous sequences of the protein from which the peptide was derived). *PIPER-FlexPepDock* global peptide docking using this receptor structure generated top-scoring conformations similar to the top-performing medium accuracy model described above. These were not submitted, but provided useful information for the subsequent refinement described next.

The successful submission of the Kozakov group as a human team involved refinement of full complexes using the MELD methodology developed by the Dill lab[25,26] (see Methods and Supporting Information for more details). The starting point were top-scoring models obtained from *Piper-FlexPepDock* global docking onto the MODELER-generated homology model. The centroid of the second most populated cluster (cluster #1) from the MELD0 protocol was the only other model of medium accuracy submitted for T121 by modeling groups. The fact that the structure is the second but not the most populated cluster indicates that either the simulations did not converge and might need to be longer, and/or that the clustering metric and protocols need to be improved. Of note, a naive approach, in which the template-based model was extended to a longer peptide and cominimized using a CHARMM-based energy function did not provide any acceptable models.

We note that the medium accuracy models generated by both groups were submitted as models 10 and 9, respectively, even though unbiased energy and cluster density criteria would have top-ranked these models. Manual intervention resulted in their ranking after wrong models.

### 3.9 | *PeptiDBeta*—Assessment of global docking of beta-sheet complementing peptide-protein interactions

It is widely known that nonpaired beta-strand edges can lead to nonspecific interactions and protein aggregation, and Nature uses negative design to prevent such aggregation.[48] In order to investigate what reinforces specificity in the beta-sheet complementing type of peptide-

protein interactions, and prevents nonspecific interactions, we compiled *PeptiDBeta*, a nonredundant benchmark of protein-peptide complexes in which the peptide extends a beta sheet in the receptor (see Methods). Our dataset consists of crystallographically solved protein-peptide complexes ("bound" structures) as well as of corresponding free ("unbound" structures) receptors, where each complex represents a different fold (according to ECOD classification[41]) (Table 4). We applied *PIPER-FlexPepDock* to model those interactions, and to study the relative contributions of the backbone hydrogen-bond network *vs* single-residue "hotspots." The high resolution *FlexPepDock* refinement step included receptor backbone minimization to account for minor backbone rearrangements of the receptor upon binding to the peptide ligand.

A critical step in blind global peptide docking is the definition of the peptide to be docked, since often the critical motif spans only part of the available peptide sequence. If the docked peptide is too short, the correct interaction might be missed due to nonspecific binding to alternative sites on the protein. In turn, if it is too long, the fragment library might not include an adequate representation of the full peptide, and wrong flanking regions might prevent the identification of the critical interaction pattern. We applied two different approaches to define the peptide for global docking: First, we restricted the peptide to the part that forms beta sheet hydrogen bonds (as defined by DSSP[42]). This allows to assess whether binding to a specific site is identified when only the beta sheet part is included. For many of the interactions, this resulted in very short peptides (many only three residues long). In our second round, we extended the peptides to include nearby motif residues (as defined by the *PeptiDock* procedure,[5] see Methods). This resulted in longer peptides to be docked (Table 4). Comparison of these two runs allows to estimate the contribution of residues beyond the central beta sheet to binding specificity. The results of the docking simulations of the DSSP-derived peptides on bound and unbound receptor structures are summarized in Table 4 and Figure 4 (blue and orange lines). Compared to the highly accurate results for bound docking, unbound docking performance was reduced. However, extending the motifs led to significant improvement in protocol performance (see Table 4), in particular for high-accuracy models within 2.5 Å RMSD (Figure 4, green line). Overall, our study highlights the importance of accurate definition of the peptide binding motif for obtaining high accuracy models. Development of robust peptide docking protocol necessitates therefore careful definition of the motif region to be docked at the initial, blind global docking step.

## 4 | DISCUSSION

We have here described the strategies that we used to model peptide-protein docking targets in CAPRI rounds 38 and 44. For both targets, our approach was able to generate the top-performing models for the CAPRI challenge (Tables 1 and 3 and Figures 1 and 2). While for targets T134 and T135 of round 44 an accurate receptor template structure was available, the challenging modeling of T121 of round 38 involved considerable rearrangement of the receptor upon peptide binding. Still, thanks to accurate definition of the peptide binding motif, we were able to identify the correct binding groove, and the correct orientation of the peptide within that groove, via iterations of optimization of the peptide-receptor interaction (with docking) and receptor flexibility (using either Monte Carlo sampling with *Rosetta FastRelax*, or Molecular Dynamics with *MELD*). In order to further assess our ability to

model this type of interactions, we generated a benchmark for beta sheet—peptide interactions, *PeptiDBeta* and assessed the performance of *PIPER-FlexPepDock* blind peptide docking on this set (Table 4 and Figure 4).

### 4.1 | Accurate modeling is possible when either peptide motif or protein binding sites are well defined

Application of *Rosetta FlexPepBind* to thread the cytoplasmic tail of MAG to DLC8 in T134 allowed clear discrimination of the binding 12mer from other 12mers within this 57-residue segment (Figure 2A). The reliability of this method has also been demonstrated previously at the proteome level, where *FlexPepBind* was used to for example, identify novel substrates of HDAC8.[3] In turn, application of *PIPER-FlexPepDock* blind docking with a defined peptide motif allows to generate high resolution-atom level models of an interaction (Figure 2B). This highlights the accuracy of the Rosetta energy function and the efficiency and adequacy of sampling of our protocols. When less information is available, for example, when the motif is not known, or/and the receptor structure moves significantly (beyond small backbone moves that can be modeled by minimization), then sampling, and with it scoring, remain challenging.

### 4.2 | Using bound (homolog) structures significantly improves global-docking prediction

As shown for both CAPRI Rounds, the use of a bound template can significantly improve model quality. This holds not only for the cases where a crystal structure for the complex of interest exists, but also for homologous structures or when the structure was crystallized with a different ligand.[49] For T121, the binding site of the peptide was identified using an ortholog bound structure, making it possible to model the binding site within the unbound structure, which was necessary to be able to accommodate the inverse peptide orientation in a subsequent docking run. In general, receptor conformational changes may severely affect the blind docking performance, but as was shown on the T121 example, such changes can be dealt with when the binding site is known, making it possible to focus on localized receptor flexibility.

The improved performance on bound structures is also evident in our *PeptiDBeta* benchmark (Table 4 and Figure 4). This reinforces the increasing appreciation of the role of template-based docking in general, as well for peptide-protein docking in particular.[50,51] With the continued addition of solved structures of protein complexes, template-based docking will become an option for the modeling of more and more interactions.

### 4.3 | Influence of peptide length and sequence on docking performance in beta-sheet complementing peptide-protein interactions

While a significant part of the binding energy in beta-sheet complementing peptide-protein interactions is contributed by a network of hydrogen bonds formed by backbone atoms, there are also side chains, and residues in “flanking regions” of the peptide that will contribute to the binding energy and, more importantly, to binding specificity. Indeed, for many interactions, docking quality improves significantly, to high resolution, when the peptide motif is extended and refined, unless it is too long to be represented by a fragment ensemble and precludes successful initial rigid body docking (for 5e0l, docking of the full 11-mer

peptide failed to produce an accurate model; best model backbone RMSD = 10.45 Å, based on initial fragments with minimal RMSD = 2.03 Å). For example, many PDZ domains bind a c-terminal carboxylate of their peptide/protein ligand. In the example of PDBid 1n7f, the DSSP defined motif (RTYS) was missing the critical C-terminal cysteine residue (Figure S3B and Table S2), precluding accurate results. Motif extension with the terminal cysteine dramatically improved the results from 14.78 to 0.53 Å RMSD (Table S2). However, not all hotspots need necessarily to be included in the docked peptide for blind docking to be successful, as we demonstrated for different peptide segments of T135 (Table 2 and Figure S2).

### 4.4. | The importance of the context of an interaction

Peptide binding specificity may also be dictated by an additional motif in the sequence (as exemplified by an additional helix in the interaction of 4rjf in Table 4). Thus, these two motifs can bind in two binding pockets on the receptor surface, leading to possible mutual binding dependencies. It is indeed a widely-used strategy to achieve strong binding by the use of a number of weak interactions, as this also provides possibilities of context-dependent switching of interactions.[52,53] Modeling of wrapping interactions that consist of several concatenated motifs, each binding to a distinct site on a protein (or on a multiprotein complex) will involve the combination of several individually docked peptides. If the context determines the binding specificity of such a motif, meaning that the specificity will be achieved only when the binding events of the adjacent motifs occur together, the identification of such interactions can be challenging due to missing information for each isolated motif.[54]

### 4.5 | Impact of improved peptide docking protocols on biological research

With the improvement of docking approaches, spurred by community wide efforts such as the CAPRI experiment, more peptide-mediated interactions will be accessible to high-resolution modeling. It remains to be seen whether these protocols—that have been calibrated on interactions that can be solved by experiment—will also provide meaningful models of the many interactions that are inaccessible to X-ray crystallography, NMR or Cryo-EM. Several databases that compile information on interactions that do not necessarily involve a defined bound structure can provide initial directions towards the extension of the rather static picture of protein communication that emerges from current protein docking approaches and assessments.[55–58] That being said, the progress made in the field of computational molecular modeling, as well as the abundance of the available experimental results, make it now possible to proceed to modeling of more complex, irregular interactions, requiring a combination of different computational approaches as well as integration of experimental data.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

This work was supported, in whole or in part, by the Israel Science Foundation (ISF) funded by the Israel Academy of Science and Humanities [717/17] (to O.S.F.), the USA-Israel Binational Science Foundation (BSF) [2015207] (to O.S.F. and D.K.), the National Science Foundation (NSF) [DBI175927; AF1645512] (to D.K.), and the National Institute of General Medical Sciences [R35GM118078; R21GM127952] (to D.K.).

Funding information

Israel Science Foundation, Grant/Award Number: 717/2017; United States-Israel Binational Science Foundation, Grant/Award Number: 2015207; National Institute of General Medical Sciences, Grant/Award Numbers: R21GM127952, R35GM118078; National Science Foundation, Grant/Award Numbers: AF1645512, DBI175927

## REFERENCES

1. Schueler-Furman O, London N Modeling Peptide-Protein Interactions. Methods in Molecular Biology. Springer, Humana Press, New York, NY; 2017.
2. Raveh B, London N, Schueler-Furman O. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. Proteins Struct Funct Bioinform. 2010;78:2029–2040.
3. Alam N, Zimmerman L, Wolfson NA, Joseph CG, Fierke CA, Schueler-Furman O. Structure-based identification of HDAC8 non-histone substrates. Structure. 2016;24:458–468. [PubMed: 26933971]
4. Raveh B, London N, Zimmerman L, Schueler-Furman O. Rosetta FlexPepDockab-initio: simultaneous folding, docking and refinement of peptides onto their receptors. PLoS One. 2011;6: e18934. [PubMed: 21572516]
5. Porter KA et al. ClusPro PeptiDock: efficient global docking of peptide recognition motifs using FFT. Bioinformatics. 2017;33:3299–3301. [PubMed: 28430871]
6. Alam N et al. High-resolution global peptide-protein docking using fragments-based PIPER-FlexPepDock. PLoS Comput Biol. 2017;13 (12):e1005905. [PubMed: 29281622]
7. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol. 1997; 268:209–225. [PubMed: 9149153]
8. Kozakov D et al. The ClusPro web server for protein-protein docking. Nat Protoc. 2017;12:255–278. [PubMed: 28079879]
9. Burley SK et al. Protein data Bank (PDB): the single global macromolecular structure archive Methods in Molecular Biology. Vol 1607 Humana Press Inc., New York, NY; 2017:627–641. [PubMed: 28573592]
10. Berman HM. The Protein Data Bank. Nucleic Acids Res. 2000;28: 235–242. [PubMed: 10592235]
11. Gront D, Kulp DW, Vernon RM, Strauss CEM, Baker D. Generalized fragment picking in Rosetta: design, protocols and applications. PLoS One. 2011;6:e23294. [PubMed: 21887241]
12. Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: an FFT-based protein docking program with pairwise potentials. Proteins Struct Funct Genet. 2006;65:392–406. [PubMed: 16933295]
13. Méndez R, Leplae R, Lensink MF, Wodak SJ. Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. Proteins: Struct Funct Bioinformat. 2005; 60:150–169.
14. Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. Proteins Struct Funct Bioinfor. 2013;81:2082–2095.
15. Marcu O et al. FlexPepDock lessons from CAPRI peptide–protein rounds and suggested new criteria for assessment of model quality and utility. Proteins Struct Funct Bioinform. 2017; 85:445–462.
16. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. Proteins Struct. Funct. Bioinform 2010;78:3073–3084.
17. Lensink MF, Velankar S, Wodak SJ. Modeling protein–protein and protein–peptide complexes: CAPRI 6th edition. Proteins Struct. Funct. Bioinform 2017;85:359–377.
18. Myllykoski M et al. High-affinity heterotetramer formation between the large myelin-associated glycoprotein and the dynein light chain DYNLL1. J Neurochem. 2018;147:764–783. [PubMed: 30261098]

19. London N, Movshovitz-Attias D, Schueler-Furman O. The structural basis of peptide-protein binding strategies. Structure. 2010;18: 188–199. [PubMed: 20159464]
20. Frappier V, Duran M, Keating AE. PixelDB: protein–peptide complexes annotated with structural conservation of the peptide binding mode. Protein Sci. 2018;27:276–285. [PubMed: 29024246]
21. Stein A, Aloy P. Contextual specificity in peptide-mediated protein interactions. PLoS One. 2008;3:e2524. [PubMed: 18596940]
22. Alford RF et al. The Rosetta all-atom energy function for macromolecular modeling and design. J Chem Theory Comput. 2017;13:3031–3048. [PubMed: 28430426]
23. Gallego P, Velazquez-Campoy A, Regue L, Roig J, Reverter D. Structural analysis of the regulation of the DYNLL/LC8 binding to Nek9 by phosphorylation. J Biol Chem. 2013;288:12283–12294. [PubMed: 23482567]
24. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci USA. 2000;97:10383–10388. [PubMed: 10984534]
25. MacCallum JL, Perez A, Dill KA. Determining protein structures by combining semireliable data with atomistic physical models by Bayesian inference. Proc Natl Acad Sci USA. 2015;112:6985–6990. [PubMed: 26038552]
26. Perez A, MacCallum JL, Dill KA. Accelerating molecular simulations of proteins using Bayesian inference on weak information. Proc Natl Acad Sci USA. 2015;112:11846–11851. [PubMed: 26351667]
27. Perez A, Morrone JA, Brini E, MacCallum JL, Dill KA. Blind protein structure prediction using accelerated free-energy simulations. Sci Adv. 2016;2:e1601274. [PubMed: 27847872]
28. Robertson JC, Perez A, Dill KA. MELD × MD folds Nonthreadables, giving native structures and populations. J Chem Theory Comput. 2018;14:6734–6740. [PubMed: 30407805]
29. Robertson JC et al. NMR-assisted protein structure prediction with MELD×MD. Proteins Struct Funct Bioinform. 2019;87: 1333–1340.
30. Morrone JA, Perez A, MacCallum J, Dill KA. Computed binding of peptides to proteins with MELD-accelerated molecular dynamics. J Chem Theory Comput. 2017;13:870–876. [PubMed: 28042966]
31. Morrone JA et al. Molecular simulations identify binding poses and approximate affinities of stapled α-helical peptides to MDM2 and MDMX. J Chem Theory Comput. 2017;13:863–869. [PubMed: 28042965]
32. Brini E, Kozakov D, Dill KA. Predicting protein dimer structures using MELD × MD. J Chem Theory Comput. 2019;15:3381–3389. [PubMed: 30908034]
33. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. Proc Natl Acad Sci USA. 2002; 99:14116–14121. [PubMed: 12381794]
34. Barlow KA et al. Flex ddG: rosetta ensemble-based estimation of changes in protein–protein binding affinity upon mutation. J Phys Chem B. 2018;122:5389–5399. [PubMed: 29401388]
35. Kellogg EH, Leaver-Fay A, Baker D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins Struct Funct Bioinform. 2011;79:830–838.
36. Kortemme T, Kim DE, Baker D. Computational alanine scanning of protein-protein interfaces. Sci STKE. 2004;2004:p12.
37. Trellet M, Melquiond ASJ, Bonvin AMJJ. A unified conformational selection and induced fit approach to protein-peptide docking. PLoS One. 2013;8:e58769. [PubMed: 23516555]
38. Lee H, Heo L, Lee MS, Seok C. GalaxyPepDock: a protein-peptide docking tool based on interaction similarity and energy optimization. Nucleic Acids Res. 2015;43:W431–W435. [PubMed: 25969449]
39. Schindler CEM, De Vries SJ, Zacharias M. Fully blind peptide-protein docking with pepATTRACT. Structure. 2015;23:1507–1515. [PubMed: 26146186]
40. Ben-Shimon A, Niv MY. AnchorDock: blind and flexible anchor-driven peptide docking. Structure. 2015;23:929–940. [PubMed: 25914054]
41. Cheng H et al. ECOD: an evolutionary classification of protein domains. PLoS Comput Biol. 2014;10:e1003926. [PubMed: 25474468]

42. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22:2577–2637. [PubMed: 6667333]
43. Bodor A et al. DYNLL2 dynein light chain binds to an extended linear motif of myosin 5a tail that has structural plasticity. Biochemistry. 2014;53:7107–7122. [PubMed: 25312846]
44. Benjamin Stranges P, Kuhlman B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. Protein Sci. 2013;22:74–82. [PubMed: 23139141]
45. Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature. 2007;450:1001–1009. [PubMed: 18075579]
46. Tyka MD et al. Alternate states of proteins revealed by detailed energy landscape mapping. J Mol Biol. 2011;405:607–618. [PubMed: 21073878]
47. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993;234:779–815. [PubMed: 8254673]
48. Richardson JS, Richardson DC. Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc Natl Acad Sci USA. 2002;99:2754–2759. [PubMed: 11880627]
49. Movshovitz-Attias D, London N, Schueler-Furman O. On the use of structural templates for high-resolution docking. Proteins. 2010;78: 1939–1949. [PubMed: 20408170]
50. Lee H, Seok C. Template-based prediction of protein-peptide interactions by using galaxypepdock Methods in Molecular Biology. Vol 1561 Humana Press Inc., New York, NY; 2017:37–47. [PubMed: 28236232]
51. Kundrotas PJ, Zhu Z, Janin J, Vakser IA. Templates are available to model nearly all complexes of structurally characterized proteins. Proc Natl Acad Sci USA. 2012;109:9438–9441. [PubMed: 22645367]
52. Ivarsson Y, Jemth P. Affinity and specificity of motif-based protein– protein interactions. Curr Opin Struct Biol. 2018;54:26–33. [PubMed: 30368054]
53. Matthews JM, Potts JR. The tandem β-zipper: modular binding of tandem domains and linear motifs. FEBS Lett. 2013;587:1164–1171. [PubMed: 23333654]
54. Peterson LX, Roy A, Christoffer C, Terashi G, Kihara D. Modeling disordered protein interactions from biophysical principles. PLoS Comput Biol. 2017;13(4):e1005485. [PubMed: 28394890]
55. Fuxreiter M Fuzziness in protein interactions—a historical perspective. J Mol Biol. 2018;430:2278–2287. [PubMed: 29477337]
56. Sickmeier M et al. DisProt: the database of disordered proteins. Nucleic Acids Res. 2007;35:D786. [PubMed: 17145717]
57. Fichó E, Reményi I, Simon I, Mészáros B. MFIB: a repository of protein complexes with mutual folding induced by binding. Bioinformatics. 2017;33:3682–3684. [PubMed: 29036655]
58. Mészáros B et al. PhaSePro: the database of proteins driving liquid–liquid phase separation. Nucleic Acids Res. 2019; gkz848, 10.1093/nar/gkz848.

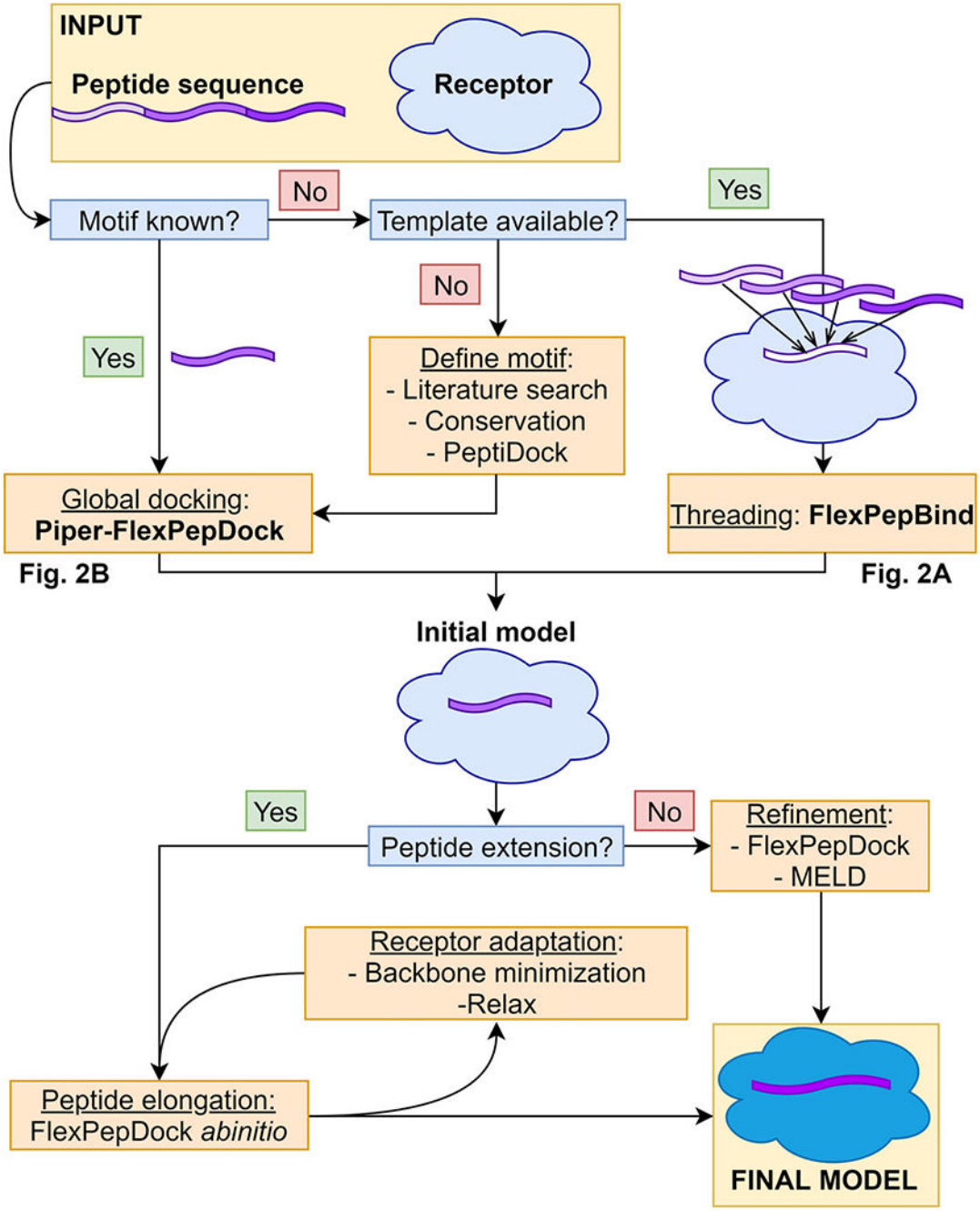


**FIGURE 1.**

Summary of the different approaches used to model the peptide-protein complexes in this study. Results from specific steps that are presented in Figure 2 are labeled in the scheme. In this study, *FlexPepBind* threading was performed by fast minimization of the interface of each candidate peptide-protein complex structure, rather than by extensive refinement. See text for more details

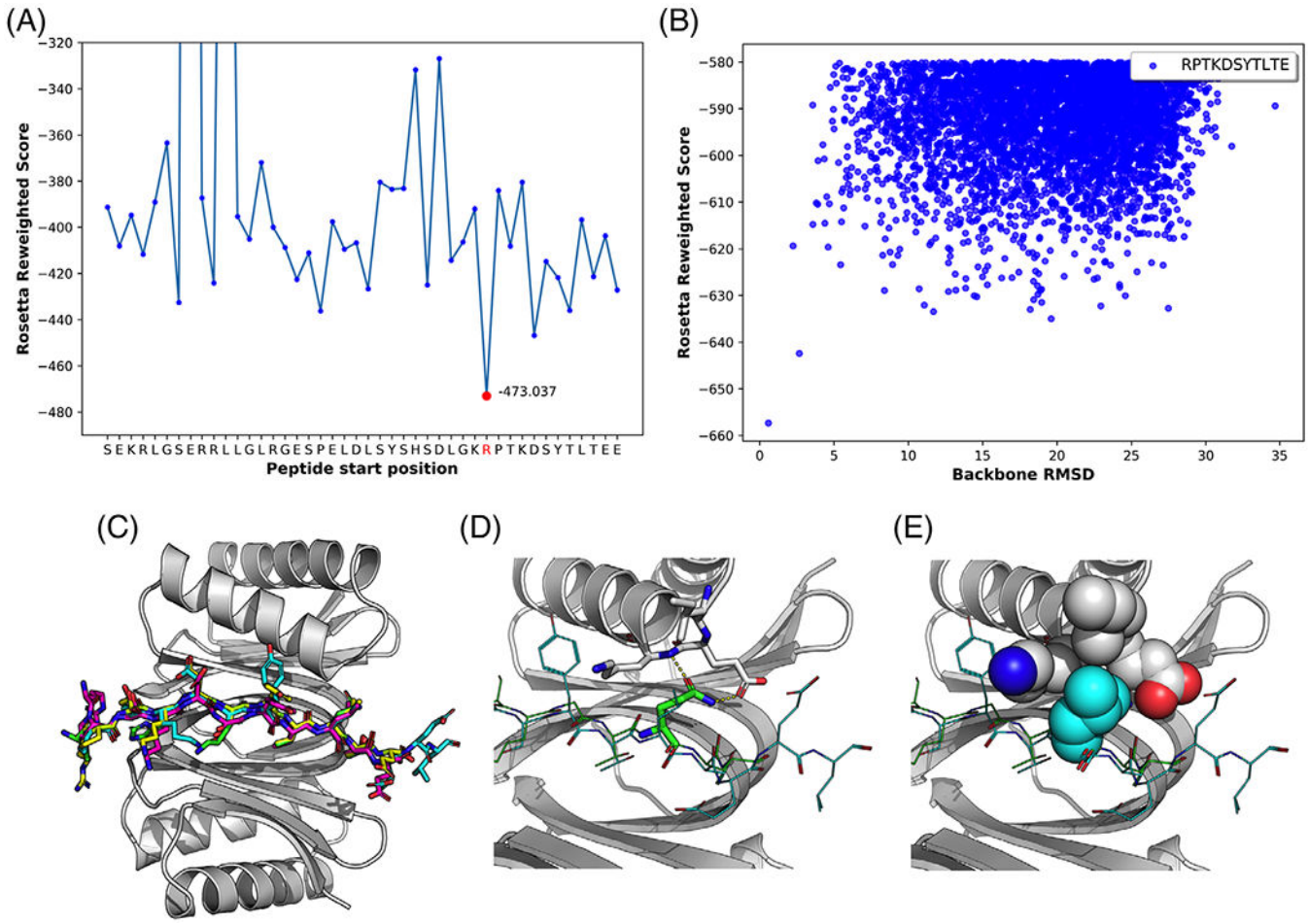


**FIGURE 2.**

Identification of the peptide binding motif using the *FlexPepBind* threading approach. A, Threading results. Plot of the estimated binding energy of each of the overlapping 11-mer peptides that were threaded onto the template structure (PDB id 3zke). The binding motif is highlighted by the best binding energy (red in the online version). Binding is estimated by the Reweighted Score (for similar results using the Interface score see Figure S1). B, Energy landscape of blind global peptide docking of RPTKDSYTLTE to DLC8 shows preferred near-native conformations. The simulation was performed using *PIPER-FlexPepDock* starting from the receptor structure and the peptide sequence, without providing any information about the receptor binding site or the peptide conformation. C, Comparison of models to the solved structure of the MAG peptide-DLC8 interaction. High accuracy models generated by *FlexPepBind* threading (RPTKDSYTLTE in yellow; rmsBB_CAPRI_if = 0.39 Å; rmsSC_CAPRI_if = 1.45 Å) and the top-scoring model generated by *PIPER-FlexPepDock* global blind docking (in magenta; rmsBB_CAPRI_if = 0.42 Å; rmsSC_CAPRI_if = 1.61 Å) are compared to the template structure (DLC8 bound to the Nek9 peptide; receptor—gray cartoon; peptide VGMHSKGTQTA—green sticks) and the solved crystal structure of the MAG peptide—DLC8 complex (PDBid: 6gzl—cyan sticks). D, E, Details of the interaction of the TQT and TLT motif with the second DLC8 monomer. Shown are, D, the original interaction (green sticks), and E, the Van der Waals interactions

formed by leucine (in cyan spheres) that compensate for the polar interactions formed by glutamine

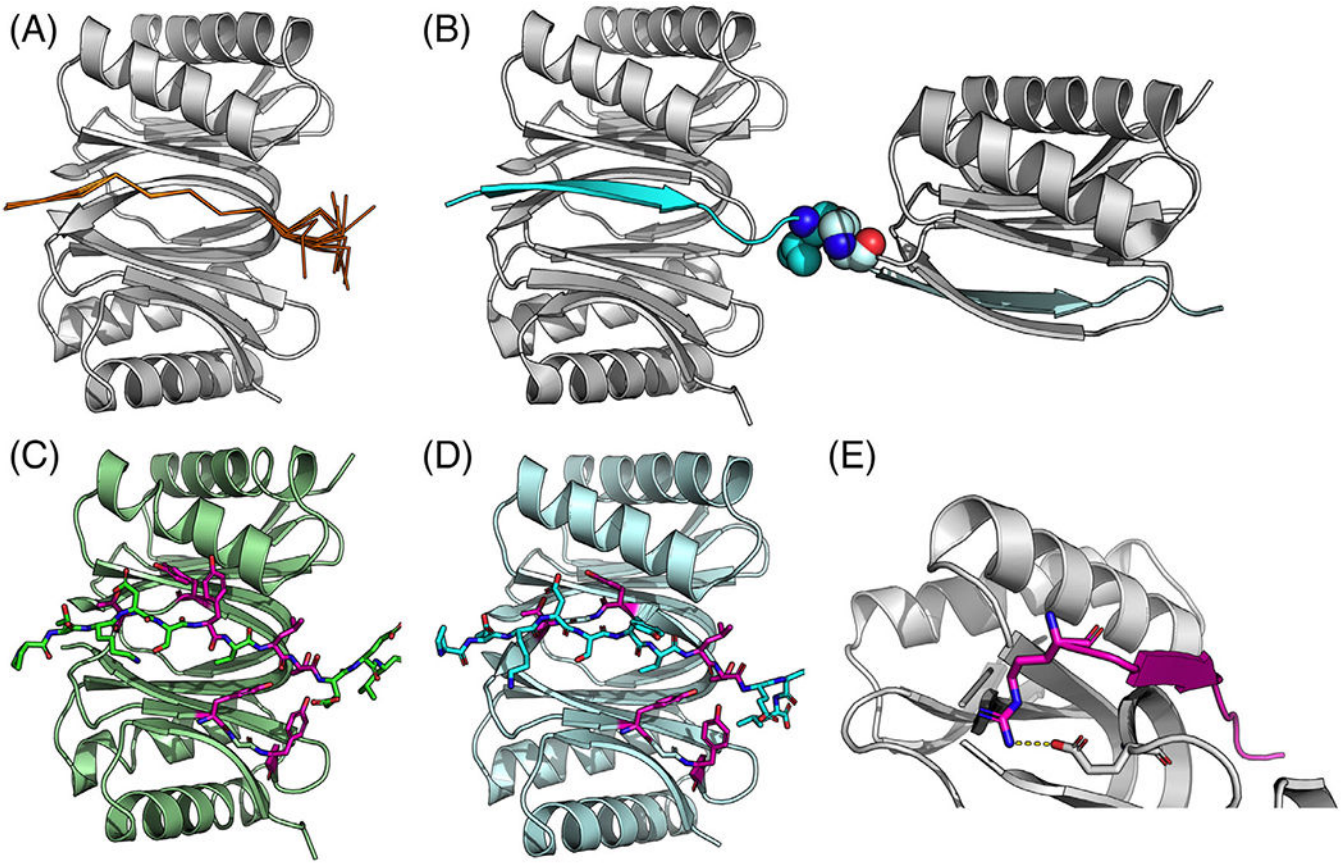


**FIGURE 3.**

Details of the MAG peptide—DLC8 interaction revealed by the models, compared to the crystal structure. A, B, The C-terminal leucine residue of the MAG peptide is stabilized by crystal contacts in the crystal structure, but not defined in the model. A, Models of the full peptide (including one additional leucine residue, PTKDSYTLTEEL) show different orientations of the C-terminus. B, The solved structure reveals that the C-terminus is mainly stabilized by a nonbiological crystal contact, indicating that in solution this residue does most probably not adapt any defined conformation, but rather remains unstructured. C, D, Interface hotspots identified by computational alanine scanning of, C, the crystal structure (green, 6glz), and, D, T135 model #1 (cyan). See text for more details, and Table S1. E, The top-scoring model from the docking simulation of the RPTKDSYTLTE peptide in T134 shows an ionic bond involving the N-terminal arginine, providing an explanation for the importance of that residue for successful docking, even though it is not resolved in the crystal structure

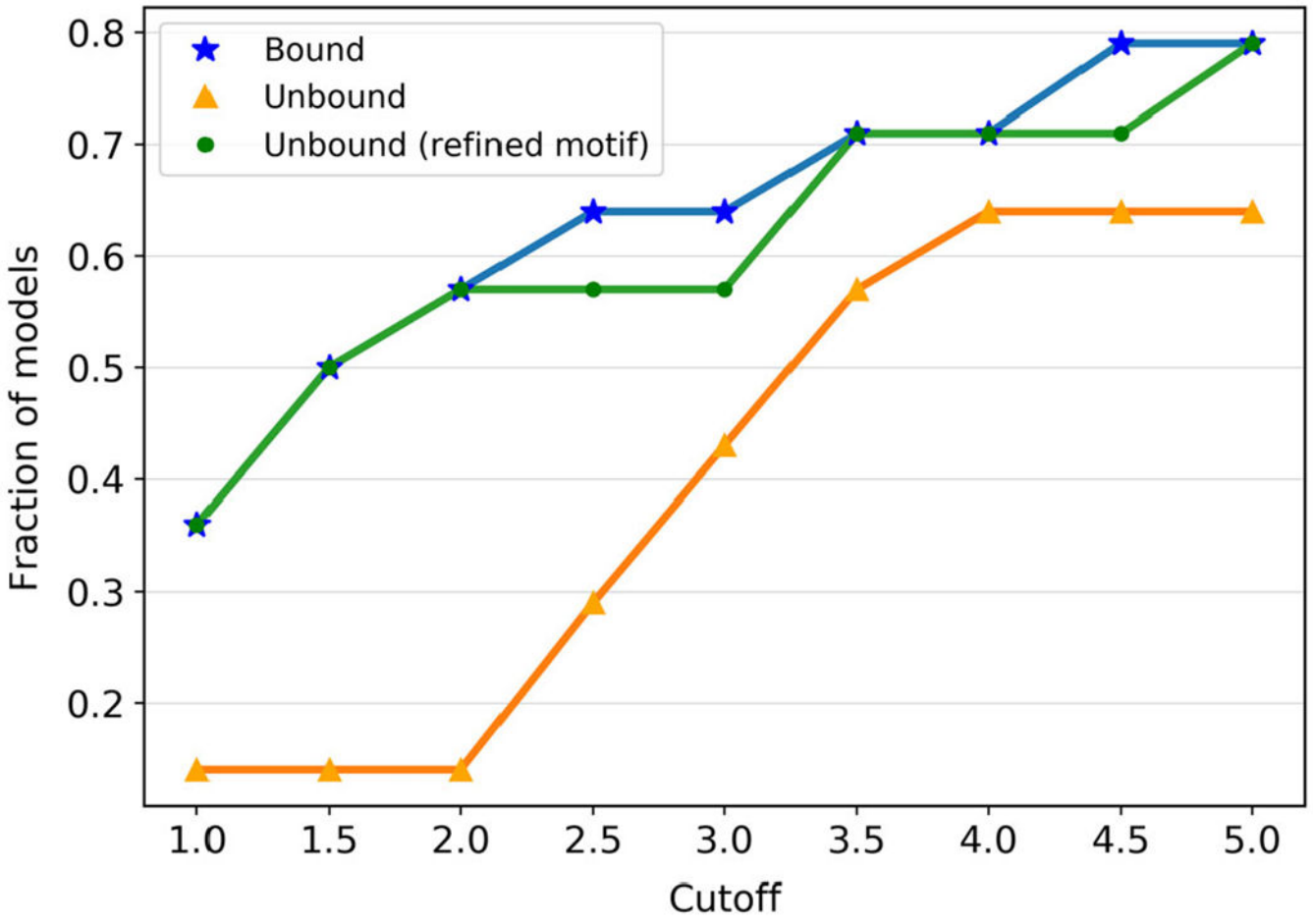


**FIGURE 4.**

*PIPER-FlexPepDock* performance on beta-sheet complementing peptide-protein complexes. The *y*-axis indicates the fraction of complexes (out of a total of n = 14 complexes in *PeptiDBeta*) for which the best interface bb RMSD model among the top10 clusters lies within the cutoff indicated on the *x*-axis. Blue/orange lines show results for peptides defined by the beta-sheet forming sequence (as defined by DSSP), docked on bound/unbound structures, respectively. The green line shows results for peptide motifs defined according to *PeptiDock* rules docked on unbound structures. While performance is not optimal when only the short beta-strand peptide is docked on free receptors, it improves for redefined peptide motifs, to levels similar to those reported for the original *PIPER-FlexPepDock* benchmark[6]

**TABLE 1**

Top 10 models submitted for round 44 (T134 and T135)

| **A** | | | | | |
|---|---|---|---|---|---|
| **T134** | | **Capri measures**[14] | | | |
| **#** | **N/C ext** | **Fnat** | **if_bb RMSD** | **if_sc RMSD** | **Model quality** |
| 1 | C-term | 0.84 | 0.68 | 2.32 | Medium |
| 2 | C-term | 0.82 | 0.77 | 2.18 | Medium |
| 3 | N-term | 0.90 | 0.33 | 1.55 | High |
| 4 | N-term | 0.90 | 0.30 | 1.60 | High |
| 5 | N-term | 0.92 | 0.30 | 1.49 | High |
| 6 | C-term | 0.90 | 0.32 | 1.67 | High |
| 7 | C-term | 0.84 | 0.72 | 2.39 | Medium |
| 8 | C-term | 0.79 | 0.76 | 2.38 | Medium |
| 9 | N-term | 0.86 | 0.33 | 1.57 | High |
| 10 | C-term | 0.87 | 0.40 | 1.83 | High |
| **B** | | | | | |
| **T135** | **Capri measures** | | | | |
| **#** | **Fnat** | **if_bb RMSD** | **if_sc RMSD** | **Model quality** | |
| 1 | 0.90 | 0.42 | 1.82 | High | |
| 2 | 0.90 | 0.37 | 1.83 | High | |
| 3 | 0.84 | 0.43 | 1.88 | High | |
| 4 | 0.92 | 0.40 | 1.98 | High | |
| 5 | 0.87 | 0.40 | 2.24 | High | |
| 6 | 0.82 | 0.62 | 2.05 | Medium | |
| 7 | 0.87 | 0.38 | 1.81 | High | |
| 8 | 0.92 | 0.38 | 1.71 | High | |
| 9 | 0.87 | 0.65 | 2.20 | Medium | |
| 10 | 0.79 | 0.75 | 2.55 | Medium | |

Abbreviations: Fnat, fraction of native contacts recovered; if_bb_RMSD/if_sc_RMSD, RMSD of the backbone/side chain atoms of interface residues, in Ångstroms.

Correction added on 06 February 2020, after first online publication: Original column 2 has been deleted from Table 1B.

## TABLE 2

Global docking accuracy for different peptide sequences of MAG (as measured by RMSD to solved crystal structure, 6gzl)

| Peptide sequence | if_bb_RMSD[a] | Peptide RMSD[b] |
|---|---|---|
| RPTKDSYT**L**TE | 0.58 (#1) | 1.91[c] |
| PTKDSYT**L**TE | 1.97 (#7) | 1.27 |
| PTKDSYT**L**TEEL | 5.00 (#7) | 2.06 |
| PTKDSYT | 1.91 (#1) | 1.13 |
| PTKDS | 20.26 (#6) | 1.20 |
| RPTKDS | 2.05 (#8) | 0.69[c] |

[a]Best interface backbone RMSD among top10 clusters, in Ångstroms (rank of model is indicated in parentheses).

[b]Best fragment RMSD within the fragment library used for global PIPER rigid body docking.

[c]RMSD was calculated excluding the N-terminal arginine residue that is not resolved in the crystal structure.

**TABLE 3**

Medium quality predictions submitted by the Furman and Kozakov/Vajda/Dill groups for T121

| Prediction | fnat | fnon_nat | Clashes | L_rmsd | I_rmsdbb | I_rmsdsc | Refinement strategy |
|---|---|---|---|---|---|---|---|
| T121_P27.M10 | 0.44 | 0.55 | 10 | 3.81 | 2.76 | 3.91 | *Rosetta FlexPepdock*[2] and *FastRelax*[46] |
| T121_P33.M09 | 0.39 | 0.6 | 7 | 4.52 | 3.04 | 4.97 | *MELD*[25] |

**TABLE 4**

*PeptiDBeta*, a benchmark for modeling of beta-sheet complementing peptide-protein interactions. Results of *PIPER FlexPepDock* simulations using different peptide spans are shown

| | PDBid | | Peptide | | Best RMSD (in top10 clusters) | | |
|---|---|---|---|---|---|---|---|
| **Domain** | **Bound** | **Unbound** | **Sequence**[a] | **Length** | **Bound**[b] | **Unbound**[c] | **Best fragment RMSD**[d] |
| SH2 | 1d4t | 1d1z | KSL***TIYA***QVQK | 4 | 1.49 | 3.16 | 0.38 |
| MATH | 1lb6 | 1lb4 | KQEP***QEID***F | 4 | 0.60 | 3.31 | 0.43 |
| SdrG | 1r17 | 2ral | G***FFFSARGH***RP | 8/7 | 0.76 | **2.17/1.23** | 1.34/1.16 |
| Pkinase_Tyr | 1zys | 3jvr | ***ASVS***A | 4 | 3.35 | 8.61 | 0.25 |
| Sina_3rd | 2a25 | 4c9z | K*P****AAVV****AP*I | 4/7 | 1.08 | 17.14/**0.87** | 0.43/0.72 |
| PHD | 2puy | 2puy | A***RTKQT***ARKS | 5 | 2.42 | **0.99** | 0.84 |
| Candida_ALS | 2y7l | 2y7m | AK***QAG***DV | 3/5 | 0.42 | 6.13/4.89 | 0.37/1.34 |
| BIR | 3d9t | 3mup | A***TPF*** | 3/4 | 1.68 | 2.88/**1.18** | 0.33/0.65 |
| MmlI | 3hds | 3hfk | ***ASWS***A | 4 | *7.64* | 16.86 | 0.40 |
| PCNA | 4rjf | 6fcm | KRRQTSMTDFF***HSKRRLI***FS | 7/6 | *9.85* | 5.61/7.39 | 0.55/0.70 |
| PDZ | 4uu5 | 4wsi | MWNLMPPPAM***ERL***I | 3/4 | 0.30 | 3.94/**0.53** | 0.26/0.66 |
| Rad60-SLD | 4wjq | 2las | SGEAEE***RIIVL***S | 3/5 | *4.16* | 2.93/**0.80** | 0.15/0.34 |
| Chromo | 4×3k | 5ejw | ***KAAR***KSA | 4 | 0.76 | **0.75** | 0.42 |
| Dynein_light | 5e0l | 3dvt | KAID***AATQ***TEE | 3/5 | *5.19* | **2.13/1.54** | 0.53/0.98 |

[a] Bold—peptide spans the beta-strand segment defined by DSSP; in italics - peptide spans the motif defined according to PeptiDock rules.[3]

[b] Failed bound runs are highlighted in italics.

[c] Unbound runs within 2.5 Å RMSD are highlighted in bold.

[d] Best similarity to bound peptide conformation among the fragments used during the rigid body docking step.