**Highlights**

- **pepATTRACT** is a new fully blind peptide-protein docking protocol.
- It performs a global search of the protein surface and simultaneous peptide modeling.
- **pepATTRACT** yielded high-quality models when tested on a large variety of complexes.
- **pepATTRACT**-local outperforms two of the best current local docking protocols.

**Authors**

Christina E.M. Schindler, Sjoerd J. de Vries, Martin Zacharias

**Correspondence**

martin.zacharias@ph.tum.de

**In Brief**

Schindler et al. present the fully blind peptide-protein docking protocol **pepATTRACT**. **pepATTRACT** predicts both the binding site and the bound peptide conformation to high precision simultaneously. It has the potential for proteome-wide applications.
Fully Blind Peptide-Protein Docking with pepATTRACT

Christina E.M. Schindler, Sjoerd J. de Vries, and Martin Zacharias

INTRODUCTION

Peptide-mediated interactions play a dominant role in cellular processes and account for about 40% of all protein-protein interactions (Petsalaki and Russell, 2008). Peptide-protein complexes are involved in many signaling and regulatory pathways as well as the DNA replication machinery. A range of pathological disorders is related to peptide-protein interactions (Naider and Anglister, 2009), making them interesting leads for protein drug design. However, for rational design of peptidic drugs, a thorough understanding of the atomistic structural knowledge of a peptide-protein complex is necessary (Rubinstein and Niv, 2009; Vanhee et al., 2011). A number of structures have been resolved experimentally and have provided important insight into the nature of peptide-mediated interactions (Della et al., 2008; London et al., 2010; Petsalaki and Russell, 2008; Stein and Aloy, 2010). However, a large number of complexes is still lacking to date. Computational peptide-protein docking methods can complement experiments by providing models for the bound complex structure.

For proteome-wide applications, a peptide-protein docking method has to be fully blind, meaning that it should be based solely on the unbound (apo) structure of the protein and the peptide sequence. In other words, such an approach should predict both the peptide binding site (global search of the protein surface) and the bound peptide conformation to high precision simultaneously. A number of peptide-protein docking and binding site prediction tools have been developed to date (Antes, 2010; Bordiner and Abagyan, 2006; Dagliyan et al., 2011; Donsky and Wolfson, 2011; Dundas et al., 2006; Hetényi and van der Spoel, 2002; Lavi et al., 2013; Luitz and Zacharias, 2014; Niv and Weinstein, 2005; Petsalaki et al., 2009; Raveh et al., 2011; Rosenfeld et al., 1995; Saladin et al., 2014; Staneva and Wallin, 2009; Trabuco et al., 2012; Unal et al., 2010; Verschueren et al., 2013). Global docking and binding site prediction methods (Ben-Shimon and Eisenstein, 2010; Dagliyan et al., 2011; Dundas et al., 2006; Lavi et al., 2013; Petsalaki et al., 2009; Saladin et al., 2014; Trabuco et al., 2012; Verschueren et al., 2013) often identify the correct binding site but do not yield high-quality models for the peptide conformation (London et al., 2013b). The ligand docking approach Autodock was adapted to fully blind peptide docking, but is limited to short fragments (Hetényi and van der Spoel, 2002; London et al., 2013b; Unal et al., 2010).

In contrast to global prediction methods, local docking approaches sample peptide conformations at a known binding site only (and therefore are not fully blind). Local docking approaches can often yield high-quality models when tested on peptide-protein docking benchmarks (London et al., 2013b). Local methods include ligand docking based approaches (Tubert-Brohman et al., 2013) and several target-specific approaches (e.g., for MHC and PDZ domains) (Antes et al., 2006; Bordiner and Abagyan, 2006; Niv and Weinstein, 2005; Rosenfeld et al., 1995; Staneva and Wallin, 2009). Several protocols are based on protein-protein docking, which is also reflected by the recent addition of peptide-protein targets to the community-wide docking challenge CAPRI (Lensink et al., 2007; Lensink and Wodak, 2010b, 2013). The Rosetta FlexPepDock ab initio approach combines local docking at the binding site with peptide folding using a backbone structure library (Raveh et al., 2010, 2011). Trellet et al. (2013) developed a peptide docking protocol in the data-driven docking program, HADDOCK, which combines the principles of conformational selection and induced fit through an ensemble of peptide conformations and flexible refinement stages. They discovered that using only three distinct peptide conformations yields very good predictive performance (Trellet et al., 2013), supporting earlier observations on the frequency of peptide conformations found in peptide-protein complexes (London et al., 2010). The DynaDock method uses
a soft-core molecular dynamics-based refinement (Antes, 2010), whereas PepCrawler employs a fast RRT-based algorithm for local sampling of peptide conformations (Donsky and Wolfson, 2011). Recently a molecular dynamics-based approach was developed in our group, which employs Hamiltonian replica exchange simulations with a variation of soft-core potentials along the replicas. This method showed promising results with respect to local peptide-protein docking (Luitz and Zacharias, 2014).

Here, we present pepATTRACT, a flexible peptide-protein docking approach in ATTRACT (May and Zacharias, 2008; Schindler et al., 2015; Zacharias, 2003). pepATTRACT is rapid: it performs a coarse-grained ab initio docking search within minutes, followed by atomistic refinement of only the most favorable solutions. More importantly, pepATTRACT is fully blind: it requires knowledge of neither the binding site nor the peptide conformation. pepATTRACT yields high-quality models of the complex, comparable with the state-of-the-art local docking protocols Rosetta FlexPepDock ab initio (Raveh et al., 2011) and HADDOCK peptide docking (Trellet et al., 2013). We also combined pepATTRACT with ambiguous interaction restraints (Dominguez et al., 2003; Nilges, 1993) that define the peptide binding site, as used before with HADDOCK. The resulting pepATTRACT-local protocol outperformed both HADDOCK and Rosetta FlexPepDock ab initio by a significant margin on a large variety of peptide-protein interactions.

RESULTS

In this work, we have developed a fully blind peptide-protein docking protocol (pepATTRACT) and embedded it in the ATTRACT docking engine (Figure 1). This protocol was tested on 80 peptide-protein complexes from the peptiDB benchmark (London et al., 2010; Trellet et al., 2013) for which the unbound protein structures are available. Initially, peptide models were generated from the peptide sequence (Tien et al., 2013) yielding three distinct idealized conformations: an extended, an α-helical, and a polyproline-II conformation. This ensemble of peptide structures was then first rigidly docked to the protein partner using a coarse-grained representation of the partner molecules (Zacharias, 2003). The rigid body docking models were ranked by their ATTRACT scores, and the best 1,000 models were selected for atomistic refinement using the recently developed flexible interface refinement method iATTRACT (Schindler et al., 2015). Subsequently these 1,000 models were refined in a molecular dynamics simulation with AMBER 14 (Case et al., 2014) using a Generalized Born implicit solvent model (see Experimental Procedures for details). The final models were clustered by the fraction of common residue contacts (Rodrigues et al., 2012) and ranked by the average energy of the top four ranking members (Trellet et al., 2013).

Bound-Bound Rigid Body Docking

The coarse-grained ATTRACT force field (Zacharias, 2003) has been used successfully to predict protein-protein complex structures in the past. Although good performance was found when using ATTRACT for peptide binding site prediction (Saladin et al., 2014), it has not yet been applied systematically to peptide-protein complexes. To test the performance of the ATTRACT force field with regard to sampling and scoring peptide-protein complexes, we first performed bound-bound rigid body docking for all cases yielding a theoretical limit for the performance of unbound-unbound docking. In terms of sampling, we obtained an overall success rate of 97% with only two failed cases (Figure S1). In both failed cases the peptide is “threaded” through a cavity in the protein and the binding site is not well
accessible. 92% of the successful cases yielded models of sub-angstrom accuracy. Ranking the rigid body docking solutions by their ATTRACT score gave a success rate of 86% in the top 1,000 models and 41% for the top-ranked model. These results gave us confidence to use the coarse-grained ATTRACT force field for rigid body sampling and scoring in the initial stage of the peptide-protein docking protocol (Figure S1).

**Unbound-Unbound Flexible Docking**

We then turned to the real challenge of blind unbound-unbound docking using a flexible docking approach with a coarse-grained rigid body docking, an atomistic flexible interface refinement, and a final molecular dynamics refinement stage (see Experimental Procedures for details). Note that this protocol requires neither knowledge of the binding site nor of the peptide conformation and thus represents a “worst case” scenario.

Figure 2 shows the results for the docking success rates after the different docking stages. Overall, our protocol generated a near-native model for 70% of the 80 peptide-protein complexes (i.e., 56 complexes) when evaluating the top 1,000 final docking models (see Experimental Procedures for details). 29% of these 56 successful cases yielded even sub-angstrom predictions. After clustering and ranking the clusters by the average energy of their top four ranked members, the top ten clusters contained at least one near-native cluster for 68% of the successful docking cases (48% of all cases) and the top-ranked cluster was found to be near-native in 29% of the successful cases (Figure 3; Table S1). Figure 4 illustrates docking models from these near-native top-ranking clusters. For the cases with sub-angstrom accuracy of the protein main chain, close agreement of predicted side-chain structure with the native bound complex was also observed (Figure S2).

**The Effect of Refinement**

We wanted to analyze the effect of the different refinement stages considering sampling and scoring separately. When it comes to sampling, we took for each complex the full set of refined structures and computed the interface-root-mean-square-deviation (IRMSD) before and after refinement, without regard to any ranking. iATTRACT refinement increased the total success rate of the protocol by 10%. It succeeded both in refining structures to sub-angstrom precision as well as generating additional near-native solutions (Figure 2), and also helped to resolve minor clashes in transitioning from a coarse-grained to a full atomistic force field. This sampling improvement during iATTRACT refinement is also reflected by an average change in IRMSD of the structures by $0.10\ \AA$. Note that iATTRACT refinement also allowed changes in the peptide main chain dihedral angles (Figure S5). Compared with the results after iATTRACT refinement, AMBER refinement generated one additional successful docking case and improved the IRMSD of the structures on average by $0.44\ \AA$. This clearly demonstrates that the additional flexibility and sampling of the MD refinement played a positive role, although structures which were already close to the bound form showed only little further improvement.

To capture only the effects of scoring, we ranked the final AMBER-refined docking models by different scores and then calculated for each ranking the best IRMSD of the top ten ranked structures for each complex. When comparing the ranking from before (ATTRACT score) and after refinement (AMBER score),
we found an average improvement of 0.32 Å for the AMBER-based ranking. This was connected to a 50% increase in the top ten success rate with the AMBER-based ranking compared with the ATTRACT-based ranking (Figure S2). Hence, the refinement yielded a significant improvement in terms of scoring.

**Sampling the Bound Peptide Conformation**

On average, the backbone of the idealized peptide conformations deviated at least by 2.3 Å from the crystallized form. We thus wanted to determine whether peptide structures moved closer to the bound form during docking. Figure 5 displays the IRMSD versus the change in peptide backbone RMSD for the final near-native docking models. Interestingly, for subangstrom models, there was a clear tendency for the peptide structure to move closer to the bound form. A similar result was found for the RMSD calculated on all heavy atoms including side chains (Figure S4). However, for models of only near-native quality, on average no improvement was found. This can be partly explained by the large amount of flexibility inherent to peptides and the fact that the interface does not restrict the conformation of all residues (Raveh et al., 2011; Schindler et al., 2015). Nevertheless, these results might indicate that more extensive peptide conformational sampling could be beneficial for some cases.

**Binding Site Prediction**

Several groups have proposed that contact analysis of docking models can be used to predict the interface of the proteins (interface post-prediction) (Fernández-Reco et al., 2004; Hwang et al., 2010b; Lensink and Wodak, 2010a; Sacquin-Mora et al., 2008; Saladin et al., 2014; de Vries and Bonvin, 2011). We thus also wanted to evaluate how well the binding site was predicted regardless of the peptide conformation. We analyzed the interface contacts of the top ten ranking final docking models and found that at least one true protein interface residue was identified in 99% of the docking cases. Furthermore, for 85% of the cases at least 50% of the correct interface residues could be recovered by these top ten ranking models. We compared our data with the recently published peptide binding site prediction tool PEP-SiteFinder. PEP-SiteFinder is based on the PTOOLS implementation of ATTRACT (Fiorucci and Zacharias, 2010; Saladin et al., 2009; Zacharias, 2003) and performs rigid body docking with peptide structures generated by the PEP-FOLD method (Maupetit et al., 2010); it was benchmarked on 41 unbound-unbound complexes from the peptiDB set. PEP-SiteFinder identified at least 50% of the correct interface residues in the top ten poses in 71% of these cases (Saladin et al., 2014), whereas our protocols was able to achieve this for 85% of these 41 complexes. In sum, the pepATTRACT protocol performed very well in interface post-prediction.

**Comparison with Other Methods**

It is interesting to compare our results with those of previously published methods. Here, we present a fully blind docking approach that includes a global search of the entire protein surface. Prior global methods were either limited to short peptide fragments (Hetényi and van der Spoel, 2002) or did not yield high-quality models of the peptide conformation (London et al., 2013b). However, we can compare the performance of...
pepATTRACT with published local docking methods. In local docking, the position of the peptide is restrained toward its native binding site. The Rosetta FlexPepDock ab initio protocol was tested on 14 unbound-unbound docking cases from our data set and achieved a docking success of 50% for the top ten ranking clusters (7 of 14) (Raveh et. al., 2011). Evaluating the same set of complexes by the same criteria, we found a docking success rate of 57% (8 of 14) using pepATTRACT. The HADDOCK peptide docking protocol reported an overall success rate of 69% for unbound-unbound docking on 62 complexes (Trellet et al., 2013). When analyzing the data for this subset, we found an overall success rate of 73% among all final pepATTRACT models. We should note, however, that we did not achieve the same scoring performance: the top-ranked cluster was near-native in only 33% of the successful cases in contrast to 50% reported for the HADDOCK protocol (Trellet et al., 2013). Still, rather surprisingly, in terms of sampling pepATTRACT yielded results similar to or slightly better than the most successful local docking methods.

Local Docking

While our blind docking results are in the range of success rates reported for the best local docking protocols FlexPepDock ab initio and HADDOCK, we also wanted to make a direct comparison by only performing local docking. The pepATTRACT-local protocol included a set of ambiguous distance restraints (Dominguez et al., 2003) toward the binding site to restrict the sampling exactly matching the conditions used in the HADDOCK protocol (Trellet et al., 2013). We applied the restraints both during the rigid body sampling stage and the flexible interface refinement. For technical reasons, the ambiguous distance restraints were not used in the AMBER refinement and the final scoring, which might have slightly deteriorated the results. Using pepATTRACT-local, we were able to generate a near-native solution in the top five clusters for 13 of the 14 cases tested in the published Rosetta FlexPepDock ab initio protocol. The FlexPepDock ab initio protocol itself could only achieve this result for four cases (Raveh et al., 2011). When limiting our data set to the 62 complexes used by Trellet et al. (2013), we obtained an overall success rate of 79% (49 of 62) with pepATTRACT-local, and 37% of these successful cases yielded sub-angstrom models. This compares very favorably with the results from the HADDOCK peptide-protein docking protocol, which achieved a 69% overall success rate (43 of 62) and only 23% sub-angstrom models among the successful cases (Trellet et al., 2013). Cluster-based scoring identified a near-native cluster at the top in 57% of the successful cases, which is comparable with the 50% achieved by the HADDOCK protocol (Trellet et al., 2013). The improved success rates, and especially the improvements in scoring for pepATTRACT-local, demonstrate the benefit of including additional information about the native binding site.

DISCUSSION

Peptide-protein interactions constitute a large fraction of all protein-protein interactions, but due to their abundance and the inherent flexibility many complexes have eluded experimental characterization. The high level of flexibility and the small size of the interface have proved to be obstacles for peptide-protein docking, and to date many methods only perform local docking, relying on information about the peptide binding site. Previous global methods were either limited to short fragments or did not yield precise predictions for the peptide conformation (London et al., 2013b). To our knowledge, the pepATTRACT approach is one of the first fully blind flexible peptide-protein docking protocols for peptides of length scales typically found in peptide-protein complexes (London et al., 2010). pepATTRACT allows for global searches of the entire protein surface given the protein structure and the peptide sequence. It identifies the binding site and simultaneously predicts the bound peptide conformation for a large variety of complexes. This is in contrast to the previously developed binding site prediction method PEP-SiteFinder, which also includes a global docking search using the PTOOLS/ATTRACT program (Saladin et al., 2014). PEP-SiteFinder only predicts the binding site but does not return structures of the peptide-protein complex. Applied to a large benchmark set of peptide-protein complexes, the pepATTRACT protocol yielded near-native models for 70% of the docking cases in a fully blind prediction manner. Its performance as a fully blind prediction method is comparable with two of the most successful local docking methods, Rosetta FlexPepDock ab initio (Raveh et al., 2011) and HADDOCK peptide docking (Trellet et al., 2013). pepATTRACT also gives very good results in interface post-prediction when compared with a state-of-the-art peptide binding site prediction tool (Saladin et al., 2014). The method could be useful for large-scale studies and the design of peptide-based inhibitors for modulating protein-protein interactions. In addition, interaction of globular proteins with disordered peptide or protein segments could also be modeled with this approach. Several peptides in the pepDB benchmark are actually derived from disordered protein regions, e.g., the cytoplasmic region of the group 1 metabotropic glutamate receptors for docking case
restraints (Dominguez et al., 2003; Nilges, 1993) to restrict the sampling toward a known binding site. The performance of pepATTRACT-local clearly surpassed that of Rosetta FlexPepDock ab initio (Raveh et al., 2011) and HADDOCK (Trellet et al., 2013) for a large number of peptide–protein complexes. We envision two application scenarios for pepATTRACT-local. Information about the native binding site can be obtained from experiments (Acharya et al., 2014; Clarke et al., 2011) and easily included during the docking process to generate high-quality complex structures. If experimental data are unavailable, bioinformatic prediction tools could be used to identify possible binding sites (Ben-Shimon and Eisenstein, 2010; Dundas et al., 2006; Lavi et al., 2013; Petsalaki et al., 2009; Saladin et al., 2014; Trabuco et al., 2012; Verschueren et al., 2013). As a special case of bioinformatic prediction, the contacts from the best pepATTRACT models can be extracted as an interface post-prediction (see the section on Binding Site Prediction). These predicted interface residues can then be used to restrict the sampling in a subsequent run with pepATTRACT-local and thus improve the results in terms of sampling and scoring (see the section on Local Docking).

In the first stage of the pepATTRACT protocol the coarse-grained ATTRACT force field was used, which has been previously parameterized for protein–protein complexes (Fiorucci and Zacharias, 2010). The high success rates already obtained in the rigid body sampling stage indicate that the force field is also applicable to model peptide binding. Vanhee et al. (2009) found that many of the conformations adopted by peptides in complexes are also found in monomeric proteins. Recently, London et al. (2013a) suggested that a large number of protein–protein interactions is dominated by the contributions of short binding motifs, so-called hot segments. These recurrent interface design principles and, thus, the similarity between protein–protein and peptide–protein complexes could explain the success in applying the ATTRACT force field to the peptide–protein docking problem.

The success of the pepATTRACT protocol is based on an efficient combination of different flexibility mechanisms in the ATTRACT engine. The protocol employs a coarse-grained force field, ensemble docking, flexible interface refinement, and a final molecular dynamics refinement to model protein and peptide flexibility. This versatile combination allows a high level of detail and accuracy in the final stages but at the same time is computationally efficient enough to screen 300,000 initial positions in a matter of minutes in the initial search stage. The large sampling in the rigid body phase provides placements at the native binding site even of non-optimal peptide structures, which were then relaxed to near-native models in the subsequent flexible refinement stages. Identifying many good initial global placements of the peptide and refining these is possibly more efficient than trying to sample all degrees of freedoms of the peptides right from the start, due to the fact that it is easy to get stuck in local minima of the rugged docking energy landscape. Using a smoother coarse-grained energy function is certainly also helpful in this context. The coarse-grained representation of the peptide also partly compensates for inaccuracies in the initial peptide conformation.

While the overall success rate for pepATTRACT is highly encouraging, docking success still strongly depends on the quality of the peptide modeling and the range of conformational changes on the protein (Figure 2). For the 31 easy benchmark cases, we only had one case whereby we could not sample any near-native solution. Nearly half of these successful easy cases yielded subangstrom predictions. For docking cases of medium difficulty, we still obtained a good success rate of 69% for generating near-native solutions; however, for the hard docking cases this rate dropped to 15% (2 of 13). For cases in which the best peptide model deviated by more than 5 Å backbone RMSD from the bound form, we were unable to sample any correct solution, also when using the local docking protocol. The current peptide docking protocol uses only three idealized peptide conformations, and thus a very limited subset of the peptide conformational phase space. It does not take the sequence of the peptides into account, e.g., disulfide bridges and preferred backbone dihedral angles for certain residue combinations. More extensive peptide modeling could include statistical approaches (Thévenet et al., 2012), peptide backbone libraries (Gront et al., 2011), or even ab initio folding in molecular dynamics simulations (Ho and Dill, 2006; Patmanidis and Glykos, 2013). Using more diverse peptide conformations may help to improve the sampling but also bears the risk of increasing the number of false-positive solutions. It is also worth noting that there were only four cases in which the deviation of the bound peptide was greater than 5 Å backbone RMSD from the idealized conformations. This demonstrates that the idealized peptide conformations capture the main features of the bound form well. Furthermore, the correct binding site could be identified even with non-optimal peptide conformations (see Figure S5 and the section on Binding Site Prediction).

In contrast, when examining the IRMSD between bound and unbound protein for the 24 failed docking cases, 14 cases display an IRMSD of >1 Å. To investigate the influence of the protein conformational change on docking success, we performed ab initio rigid body docking using the unbound structure of the protein partner and the bound form of the peptide. We found a success rate for the top 1,000 ranked models of 63% (Figure S1), which is equal to the success rate found for unbound-unbound docking after the rigid body stage (Figure 2) and significantly lower than for bound-bound docking (86%). Using the unbound protein structure prevents sampling of near-native conformations completely for 12 docking cases, compared with two for bound-bound rigid body docking (Figure S1). In addition, the scoring performance deteriorated, with only 74% of the successful cases ranked in the top 1,000, compared with 89% in bound-bound docking (Figure S1). Hence, the conformational change on the protein side seems to be a greater limitation to docking success than the accuracy of the peptide modeling. For considering receptor flexibility, the current protocol could be easily extended to include multiple conformations for the receptor in an ensemble docking approach or to approximately describe global backbone flexibility using pre-calculated normal modes (May and Zacharias, 2008). However, for the hard docking cases, which include also partial refolding of the protein receptor, such a semi-rigid docking approach might not be sufficient.

To make the pepATTRACT docking protocol easily accessible to the scientific community, an extension of our previously presented protein–protein docking web interface has been
developed (de Vries et al., 2015). The web interface helps the user set up a script, which performs the rigid body sampling stage and the flexible interface refinement starting from the structure of the unbound protein and the peptide sequence. It also provides the option to specify residues for ambiguous interaction restraints (Dominguez et al., 2003; Nilges, 1993) (pepATTRACT-local). Note that the script does not contain commands for the AMBER refinement stage and the clustering. Scripts for these two stages are available from the authors on request. The docking can then be performed on the user’s machine with a local installation of the ATTRACT program. A typical docking run takes around 1–2 hr on a single (four-core) processor with the main computer time used in the flexible interface refinement. AMBER refinement of 1,000 structures, which is not included in the docking script, typically takes around 16 hr on a consumer graphics processing unit for an average-sized complex. Hence, the whole protocol as described in this work can be run overnight on a standard desktop PC for many complexes. This compares favorably with Rosetta FlexPepDock ab initio, which requires 24 hr for 50,000 models on 120 processors (translating to ≈57.6 hr for the refinement of 1,000 models per processor) (Raveh et al., 2011). The docking script generated by the web interface provides an easy entry point for non-expert users into fast peptide-protein docking in ATTRACT. The web interface is available at http://www.attract.ph.tum.de/peptide.html.

**EXPERIMENTAL PROCEDURES**

The fully blind peptide-protein docking protocol pepATTRACT consists of the following steps (Figure 1). First, peptide model structures are generated from sequence (Tien et al., 2013). Next, global rigid body docking with ATTRACT using a coarse-grained force field is performed (May and Zacharias, 2008; Zacharias, 2003). The rigid body docking solutions are ranked by ATTRACT score and the best 1,000 ranked models are selected for a subsequent atomistic refinement stage with ATTRACT (Schindler et al., 2015). The structures were then finally refined in a molecular dynamics simulation with AMBER 14 (Case et al., 2014).

**Peptide Structures**

For each peptide we generated three conformations from its sequence using the Python library PeptideBuilder (Tien et al., 2013). We chose backbone dihedral angles to represent α-helical (φ = −57°, ψ = −47°), extended (φ = −139°, ψ = −135°), and polyproline conformations (φ = −78°, ψ = 149°). This conformational selection approach is based on Trellet et al. (2013).

**ATTRACT Rigid Body Docking**

The protein and peptide structures were converted to the ATTRACT atom type representation (Zacharias, 2003) with the ATTRACT tool reduce. The empirical coarse-grained force field in ATTRACT represents the amino acid side chains by one or two beads and takes all backbone atoms without hydrogens into account. Interactions are based on Lennard-Jones (LJ)-type potentials and can be either attractive or repulsive (using a saddle point instead of an energy minimum in the LJ potential) (Fiorucci and Zacharias, 2010). In addition, interactions between charged residues are calculated via a Coulomb term with a distance-dependent dielectric constant (ε = 154). Starting points for the docking were generated by choosing random positions and orientations for the protein partners. For bound-bound docking we used 100,000 starting points, and tripled this number for unbound-unbound docking to account for the three possible peptide conformations. The starting structures are subjected to rigid body optimizations in a potential energy minimization of 1,000 minimization steps with the ATTRACT metric minimizer (May and Zacharias, 2005, 2008). Energy calculation was accelerated using a pre-calculated grid (de Vries et al., 2015), and an additional harmonic potential was applied on the center of mass of the proteins to draw them toward each other (“gravity”). A subsequent potential energy minimization of 1,000 minimization steps was applied without this gravity potential. All peptide conformations were docked separately (ensemble docking). The complete docking run takes approximately 10 min to 1 hr depending on the size of peptide and protein partner. Finally, the docking candidates were ranked by ATTRACT energy evaluated within a squared cutoff of 50 Å².

**pepATTRACT-local**

To perform local docking, we repeated the pepATTRACT protocol with additional restraints (pepATTRACT-local), recreating the conditions used in previous docking procedures (Trellet et al., 2013). We used ATTRACT with...
ambiguous distance restraints based on active and passive residues, following their original specification in the HADDOCK method (Dominguez et al., 2003; Nilges, 1993). The active residues on the protein were derived from the residue contacts in the bound complex structure within a cutoff of 5 Å. All peptide residues were treated as passive residues. The minimum distance was set to 3 Å during the coarse-grained docking simulations and to 2 Å for the atomistic refinement. For rigid body docking, an initial rotational sampling stage was added to the protocol, in which only the restraints are applied and the proteins can orient toward each other with the translational degrees of freedom fixed (Dominguez et al., 2003; Trellet et al., 2013). The rotational sampling phase applies a maximum of 1,000 minimization steps.

**Data Set**

Docking was performed on 80 peptide-protein complexes from peptiDB docking benchmark (London et al., 2010) for which the unbound protein structures were available, including several additions of unbound structures by Trellet et al. (2013). The protein structures were downloaded from the PDB. If necessary, residues were renumbered in the unbound structures to match the bound forms, parts in the unbound form that are not present in the bound form were removed (and vice versa), and point mutations were introduced to resolve minor differences in the protein sequences.

**Benchmark Classification**

To classify the benchmark, we aligned the unbound protein structure and the peptide models to the bound complex, and calculated the backbone IRMSD ($\text{IRMSD}_{\text{ab}}$) for all residues within a distance cutoff of 10 Å of the partner molecule. Cases were classified according to the minimal $\text{IRMSD}_{\text{ub}}$ accounting both for protein flexibility and peptide modeling quality, i.e., similarity of the peptide to one of the idealized conformations. This classification scheme is similar to the one used for the protein-protein docking benchmark (Hwang et al., 2010a). We chose the following criteria to characterize the docking cases:

- **Easy:** $\text{IRMSD}_{\text{ub}} < 1.5$ Å
- **Medium:** $1.5$ Å < $\text{IRMSD}_{\text{ub}}$ < $3$ Å
- **Hard:** $\text{IRMSD}_{\text{ub}}$ > $3$ Å

According to this classification, the benchmark contains 31 easy, 36 medium, and 13 hard cases.

**Evaluation Criteria**

The docking solutions were evaluated by IRMSD (Méndez et al., 2005). Since peptide-protein interfaces are typically smaller than protein-protein interfaces, we chose the following criteria (Trellet et al., 2013) to characterize the docking solutions:

- **Not acceptable:** IRMSD > 2 Å
- **Near-native:** IRMSD < 2 Å
- **Sub-angstrom:** IRMSD ≤ 1 Å

The IRMSD is calculated on the backbone atoms of both protein and peptide residues that are within 10 Å of the partner molecules (as defined based on the crystal structure of the complex).

We further refer to as “acceptable models” any near-native or better (sub-angstrom) predictions. For evaluating the sampling and the scoring performance, we calculated the percentage of successful docking cases. A docking case was deemed successful if at least one acceptable solution was found in the top N solutions.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes five figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.str.2015.05.021.

**ACKNOWLEDGMENTS**

C.S. acknowledges funding by the Center for Integrated Protein Science Munich (CIPSM). The calculations were performed on computational resources funded by a CIPSMWomen consumables grant and at Leibnitz Supercomputer Center (LRZ, grant: pr86l).

**REFERENCES**


Received: March 6, 2015

Revised: May 21, 2015

Accepted: May 25, 2015

Published: July 2, 2015