

# The pepATTRACT web server for blind, large-scale peptide–protein docking

Sjoerd J. de Vries<sup>1</sup>, Julien Rey<sup>1</sup>, Christina E. M. Schindler<sup>2</sup>, Martin Zacharias<sup>2</sup> and Pierre Tuffery<sup>1,\*</sup>

<sup>1</sup>INSERM UMR-S 973/Université Paris Diderot/Sorbonne Paris Cité/RPBS, Paris 75205, France and <sup>2</sup>Physik T38, Technische Universität München, Garching 85748, Germany

Received February 23, 2017; Revised April 07, 2017; Editorial Decision April 14, 2017; Accepted April 18, 2017

## ABSTRACT

Peptide–protein interactions are ubiquitous in the cell and form an important part of the interactome. Computational docking methods can complement experimental characterization of these complexes, but current protocols are not applicable on the proteome scale. pepATTRACT is a novel docking protocol that is fully blind, i.e. it does not require any information about the binding site. In various stages of its development, pepATTRACT has participated in CAPRI, making successful predictions for five out of seven protein–peptide targets. Its performance is similar or better than state-of-the-art local docking protocols that do require binding site information. Here we present a novel web server that carries out the rigid-body stage of pepATTRACT. On the pepTiDB benchmark, the web server generates a correct model in the top 50 in 34% of the cases. Compared to the full pepATTRACT protocol, this leads to some loss of performance, but the computation time is reduced from ~18 h to ~10 min. Combined with the fact that it is fully blind, this makes the web server well-suited for large-scale *in silico* protein–peptide docking experiments. The rigid-body pepATTRACT server is freely available at <http://bioserv.rpbs.univ-paris-diderot.fr/services/pepATTRACT>.

## INTRODUCTION

Peptide–protein interactions are ubiquitous in the cell. They are estimated to account for 15–40% of the interactome (1) and more than 1.8 million putative peptide sequences have been identified in prokaryotes alone (2). Computational docking methods can complement experimental characterization of these complexes. For example, the Rosetta FlexPepDock *ab initio* protocol (3) and the protein–peptide protocol of HADDOCK (4) are well-known methods that can do *local protein–peptide docking*: the prediction of the struc-

ture of a protein–peptide complex, given the structure of the unbound protein, the sequence of the peptide and information about the binding site. There are now several web servers that can do local protein–peptide docking, including GalaxyPepDock (5) (based on comparative modeling) and PEP-FOLD3 (6). These can be used in synergy with web servers that can predict the binding site of a protein–peptide complex (7,8), refine the protein–complex structure (9) or optimize the peptide sequence (10,11).

We have previously introduced pepATTRACT (12), a novel docking protocol that is *fully blind*, i.e. it requires only the experimental protein structure and the peptide sequence and no information about the binding site. Briefly, pepATTRACT rigidly docks three idealized peptide conformations onto the receptor protein using ATTRACT (13,14), followed by a two-step flexible refinement: first using iATTRACT (15), and then by molecular dynamics with AMBER (16). Although fully blind, pepATTRACT performs similarly to the local docking methods FlexPepDock and HADDOCK. A version that does use binding site information, pepATTRACT-local, significantly outperforms both. In various stages of its development, pepATTRACT has participated in CAPRI, making successful predictions for five out of seven protein–peptide targets (however, for several targets, homology models of the complex were used) (17). pepATTRACT is currently available as a web interface that generates an ATTRACT protocol for execution on the user's own computer.

Given the large number of peptides and peptide–protein interactions, it would be beneficial to have a peptide–protein docking method that is applicable on a large-scale. To process thousands of peptide–peptide interactions without experimental data on the binding site, such a method would have to be fully blind, and run to completion within a few minutes, preferably as a web server. Recently, several new fully blind protein–peptide docking methods have been published: AnchorDock (18), CABS-dock (19) and MDockPeP (20). Of these, only CABS-dock is available as a web server. Unfortunately, current fully blind methods are too slow for large-scale applications. pepATTRACT, An-

\*To whom correspondence should be addressed. Tel: +331 57 27 83 86; Fax: +331 57 27 83 72; Email: pierre.tuffery@univ-paris-diderot.fr

chorDock and CABS-dock are all based on lengthy molecular dynamics simulations, and MDockPeP is based on an iterative cycle of docking runs. All four methods take several hours to run, even on a modern GPU or multi-core CPU.

Here we present a novel web server that corresponds to a simplified version of the full pepATTRACT protocol. The server carries out only the rigid-body stage, performing docking runs in about 10 min. In addition, an analysis is performed of the most frequent protein residues in the interface of the docking models. The docking models are clustered and the top 50 models are shown in the browser. Combined with the fact that it is fully blind, this makes the web server well-suited for large-scale *in silico* protein–peptide docking experiments. It is possible to run the server in batch mode, submitting multiple docking jobs via a script. See the pepATTRACT main page for more details.

## MATERIALS AND METHODS

### Benchmark

The rigid-body pepATTRACT web server was tested on all complexes from the peptiDB benchmark (21) where unbound receptor protein structures are available, including several additions of unbound structures by Trellet *et al.* (4). For all complexes, the unbound form of the receptor protein structure was used. This corresponds to the same data set as used in the original pepATTRACT paper (12).

### Peptide–protein docking protocol

A reduced version of the pepATTRACT protocol (12) is performed, using only the rigid-body stage. Three idealized peptide model structures are generated from sequence using the Python library PeptideBuilder (22). Missing side chain atoms in the protein are completed using the program PDB2PQR (23). The protein and peptide structures are converted to the ATTRACT coarse-grained atom type representation (13). Global rigid body docking is performed with ATTRACT (13,14), using 100 000 random starting positions per peptide model. The rigid body docking solutions are ranked by ATTRACT score. The top 10,000 docking solutions are clustered by pairwise ligand RMSD using a cutoff of 1 Å.

### Interface analysis and prediction

The top 50 docking models (before clustering) are selected, and all protein–peptide contacts between heavy atoms within 5 Å are computed and pooled for all models. For every protein residue, the interface propensity is computed, defined by the number of contacts in which this residue participates, divided by 50, the number of models. The residues are sorted by interface propensity and presented to the user. In this study, the top N residues with the highest interface propensity are predicted as ‘interface’, where N is the number of residues that makes at least one contact in the model structure of the complex. Interface predictions using PepSite (7) and PEP-SiteFinder (8) were made using their respective web servers. For PEP-SiteFinder, the top N residues were selected in the same way (except for 2JAM, for which prediction failed). For PepSite, all predicted patches on the protein receptor were pooled.

## IMPLEMENTATION

### Web server input

The web server requires:

The structure of the protein receptor in PDB format. If the receptor consists of multiple chains, they are concatenated. In case of nuclear magnetic resonance structures, the first model is selected. Only PDB ATOM records are considered. Co-factors (HETATM records) are ignored. Missing side chains are modeled using the program PDB2PQR (23). Modified amino acid side chains are replaced by dummy atoms. The maximum size of the protein receptor is 10 000 heavy atoms.

The sequence of the peptide (one-letter code). Modified amino acids are not supported. During the docking, the peptide will be modeled as an ensemble of three idealized peptide conformations. In theory, pepATTRACT enforces no particular maximum length of the peptide sequence, but note that as the length of the peptide grows, the probability decreases that its structure is well approximated by one of the three idealized peptide conformations that are used in the docking. For this reason, the server does not presently accept sequences of more than 20 amino acids.

### Web server output

The server returns 50 atomic models, with each model being the lowest-energy structure of a docking cluster. These models can be downloaded and they are also visualized directly in the browser. The interactive display relies on the PV—JavaScript Protein Viewer (<https://biasmv.github.io/pv/>). The interface propensity of every residue is also given (see below). Note that all protein residues have been renumbered from 1, their residue number may be different than in the original PDB.

Finally, the ATTRACT force field energy of each model is shown in a table. Note that this energy should be used only to identify the correct model, it cannot be used to predict binding affinities.

## RESULTS

### Docking performance

The performance of the rigid-body pepATTRACT web server was tested on the peptiDB benchmark (21). For 27/80 complexes, at least one of the 50 models had an interface RMSD (iRMSD) of better than 2 Å. Among the top 10 models, this was achieved for 14 complexes. The full pepATTRACT protocol achieves this for 38 cases, but takes many hours to complete. Thus, increasing the number of models to 50 affects the overall performance in a limited manner, but above all, it makes possible to largely reduce the execution time, so as to provide a routine tool for the user. The docking of all 80 complexes of the peptiDB took ~11 h on one computing node. Docking of multiple targets may however proceed simultaneously on the RPBS cluster if nodes are available.

Supplementary Table S1 shows the result of each individual complex. Failures are typically caused by a scoring

problem, i.e. correct structures are generated but not ranked high enough.

### Interface prediction performance

The web server also returns an analysis of the protein residues that are most prevalent in protein–peptide contacts among the top 50 docking models (before clustering). The interface propensity (the average number of protein–peptide contacts) of each residue is visualized in a table. In the docking model visualization, the interface propensity can be projected onto the receptor protein. When the top  $N$  residues with the highest interface propensity are selected (with  $N$  being the total number of interface residues), the specificity (precision) and sensitivity of the interface prediction are both 37.2%. This is considerably better than two existing web servers that we tested. PEP-SiteFinder (8) achieved a sensitivity/specificity of 27.3% on the same dataset. PepSite (7), which only accepts peptides up to 10 amino acids, achieved 13.4% sensitivity and 26.6% specificity on that subset (38.3% for rigid-body pepATTRACT). For rigid-body pepATTRACT, in 89% of the cases, at least one predicted residue was correct, compared to 65% for PEP-SiteFinder and 53% for PepSite.

We anticipate this performance makes the service particularly well tuned to assist users in the identification of the binding site, and the design of further experiments in the wet lab to probe peptide–protein interactions.

## EXAMPLES

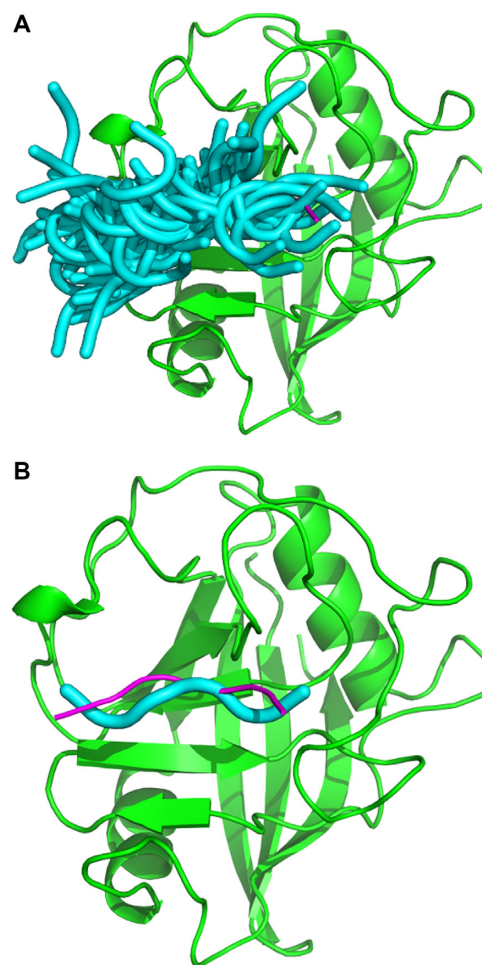
### Examples of web server output

*Example 1.* In demonstration mode, the web server performs a docking between cyclophilin A (unbound crystal structure; PDB code: 2ALF) and the HAGPIA peptide from the HIV-1 capsid protein. Figure 1 shows the results of the docking. On the left, the 50 best peptide poses are depicted. One sees that the peptide tends to always interact with the same region on the receptor surface. Among the top 50 models, the third model has an interface RMSD of 0.83 Å toward the crystal structure of the cyclophilin A–peptide–protein complex (PDB code: 1AWR) (right—the experimental peptide conformation is in magenta).

*Example 2.* The WD40 domain of WDR5 represents a class of histone methyl-lysine recognition domains that is important for recruiting H3K4 methyltransferases to K4-dimethylated histone H3 tail as well as for global and gene-specific K4 trimethylation. It is able to bind histone H3K4 peptides (PDB code: 2H14). Starting from the conformation of the unbound protein (PDB code: 2H9M) pepATTRACT identifies two candidate regions for peptide interaction, and best pose 5 has an interface RMSD of 1.35 Å toward the experimental peptide pose (Figure 2).

## CONCLUSION

Here we present a novel pepATTRACT web server for fully blind peptide–protein docking. By eliminating the flexible refinement stages, the computation time is reduced from



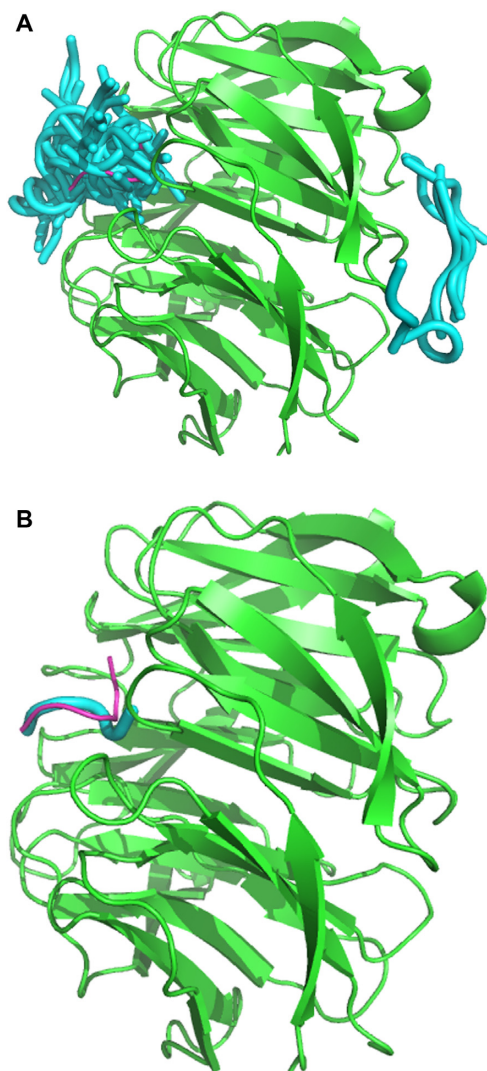
**Figure 1.** Left: experimental complex structure of the unbound conformation of the receptor (PDB code: 2ALF) with the 50 peptide best poses. green: protein; cyan: peptide. Right: Peptide pose 3 (iRMSD 0.83 Å). magenta: experimental peptide pose (PDB code: 1AWR).

~18 h to ~10 min. However, compared to the full pepATTRACT protocol, the performance is somewhat reduced: from having a correct model in the top 10 in 38/80 cases (48%), to having a correct model in the top 50 in 27/80 cases (34%). If more precision is required, the user may opt instead for the existing web interface at <http://www.attract.ph.tum.de/services/ATTRACT/peptide.html>, which sets up the computationally intensive full pepATTRACT protocol to be run locally by the user.

Several improvements of the service are under consideration. At first, the service presently runs on CPUs, when some part of the service could be ported to GPUs, still improving its speed. Second, an obvious limitation is that the peptide is presently represented by only three idealized conformations, which may give a poor approximation of the bound peptide conformation, especially for longer peptides. A perspective is to integrate a peptide structure prediction program, such as PEP-FOLD3 (6), into the pepATTRACT protocol.

However, combined with the fact that it is fully blind, the short running time makes the pepATTRACT web server





**Figure 2.** Left: experimental complex structure of the unbound conformation of the receptor (PDB code: 2H14) with the 50 peptide best poses. green: protein; cyan: peptide. Right: Peptide pose 5 (iRMSD 1.35 Å). magenta: experimental peptide pose (PDB code: 2H9M).

well-suited for large-scale *in silico* protein-peptide docking experiments, and the performances in the identification of the receptor interacting residues can provide a useful starting point to rationalize the design of further experiments in the wet lab.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

ANR [ANR-11-INBS-0013]; INSERM [UMR-S 973]; Ressource Parisienne en Bioinformatique Structurale. Funding for open access charge: INSERM [UMR-S 973]. *Conflict of interest statement.* None declared.

## REFERENCES

- Petsalaki, E. and Russell, R.B. (2008) Peptide-mediated interactions in biological systems: new discoveries and applications. *Curr. Opin. Biotechnol.*, **19**, 344–350.
- Rey, J., Deschavanne, P. and Tuffery, P. (2014) BactPepDB: a database of predicted peptides from a exhaustive survey of complete prokaryote genomes. *Database (Oxford)*, **2014**, doi:10.1093/database/bau106.
- Raveh, B., London, N., Zimmerman, L. and Schueler-Furman, O. (2011) Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors. *PLoS One*, **6**, e18934.
- Trellet, M., Melquiond, A.S.J. and Bonvin, A.M.J.J. (2013) A unified conformational selection and induced fit approach to protein-peptide docking. *PLoS One*, **8**, e58769.
- Lee, H., Heo, L., Lee, M.S. and Seok, C. (2015) GalaxyPepDock: a protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Res.*, **43**, W431–W435.
- Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derreumaux, P. and Tuffery, P. (2016) PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.*, **44**, W449–W454.
- Trabuco, L.G., Lise, S., Petsalaki, E. and Russell, R.B. (2012) PepSite: prediction of peptide-binding sites from protein surfaces. *Nucleic Acids Res.*, **40**, W423–W427.
- Saladin, A., Rey, J., Thévenet, P., Zacharias, M., Moroy, G. and Tuffery, P. (2014) PEP-SiteFinder: a tool for the blind identification of peptide binding sites on protein surfaces. *Nucleic Acids Res.*, **42**, W221–W226.
- London, N., Raveh, B., Cohen, E., Fathi, G. and Schueler-Furman, O. (2011) Rosetta FlexPepDock web server—high resolution modeling of peptide-protein interactions. *Nucleic Acids Res.*, **39**, W249–W253.
- Zaidman, D. and Wolfson, H.J. (2016) PinaColada: peptide-inhibitor ant colony ad-hoc design algorithm. *Bioinformatics*, **32**, 2289–2296.
- Obarska-Kosinska, A., Iacoangeli, A., Lepore, R. and Tramontano, A. (2016) PepComposer: computational design of peptides binding to a given protein surface. *Nucleic Acids Res.*, **44**, W522–W528.
- Schindler, C.E.M., de Vries, S.J. and Zacharias, M. (2015) Fully blind peptide-protein docking with pepATTRACT. *Structure*, **23**, 1507–1515.
- Zacharias, M. (2003) Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.*, **12**, 1271–1282.
- de Vries, S.J., Schindler, C.E.M., Chauvot de Beauchêne, I. and Zacharias, M. (2015) A web interface for easy flexible protein-protein docking with ATTRACT. *Biophys. J.*, **108**, 462–465.
- Schindler, C.E.M., de Vries, S.J. and Zacharias, M. (2015) iATTRACT: simultaneous global and local interface optimization for protein-protein docking refinement. *Proteins*, **83**, 248–258.
- Case, D.A., Darden, T.A., Cheatham, T.E. III, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M. et al. (2012) AMBER 12. University of California, San Francisco.
- Schindler, C.E.M., Chauvot de Beauchêne, I., de Vries, S.J. and Zacharias, M. (2016) Protein-protein and peptide-protein docking and refinement using ATTRACT in CAPRI. *Proteins*, **85**, 391–398.
- Ben-Shimon, A. and Niv, M.Y. (2015) AnchorDock: blind and flexible anchor-driven peptide docking. *Structure*, **23**, 929–940.
- Kurcinski, M., Jamroz, M., Blaszczyk, M., Kolinski, A. and Kmiecik, S. (2015) CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Res.*, **43**, W419–W424.
- Yan, C., Xu, X. and Zou, X. (2016) Fully blind docking at the atomic level for protein-peptide complex structure prediction. *Structure*, **24**, 1842–1853.
- London, N., Movshovitz-Attias, D. and Schueler-Furman, O. (2010) The structural basis of peptide-protein binding strategies. *Structure*, **18**, 188–199.
- Tien, M.Z., Sydykova, D.K., Meyer, A.G. and Wilke, C.O. (2013) PeptideBuilder: a simple Python library to generate model peptides. *PeerJ*, **1**, e80.
- Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. and Baker, N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.*, **32**, W665–W667.