

HPEPDOCK: a web server for blind peptide–protein docking based on a hierarchical algorithm

Pei Zhou, Bowen Jin, Hao Li and Sheng-You Huang*

Institute of Biophysics, School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

Received January 26, 2018; Revised April 21, 2018; Editorial Decision April 23, 2018; Accepted April 30, 2018

ABSTRACT

Protein–peptide interactions are crucial in many cellular functions. Therefore, determining the structure of protein–peptide complexes is important for understanding the molecular mechanism of related biological processes and developing peptide drugs. HPEPDOCK is a novel web server for blind protein–peptide docking through a hierarchical algorithm. Instead of running lengthy simulations to refine peptide conformations, HPEPDOCK considers the peptide flexibility through an ensemble of peptide conformations generated by our MODPEP program. For blind global peptide docking, HPEPDOCK obtained a success rate of 33.3% in binding mode prediction on a benchmark of 57 unbound cases when the top 10 models were considered, compared to 21.1% for pepATTRACT server. HPEPDOCK also performed well in docking against homology models and obtained a success rate of 29.8% within top 10 predictions. For local peptide docking, HPEPDOCK achieved a high success rate of 72.6% on a benchmark of 62 unbound cases within top 10 predictions, compared to 45.2% for HADDOCK peptide protocol. Our HPEPDOCK server is computationally efficient and consumed an average of 29.8 mins for a global peptide docking job and 14.2 mins for a local peptide docking job. The HPEPDOCK web server is available at <http://huanglab.phys.hust.edu.cn/hpepdock/>.

INTRODUCTION

Peptide–protein interactions play an important role in a variety of biological processes such as signal transduction, immune responses, and cellular regulation (1,2). It has been found that nearly 40% of the protein–protein interactions are mediated by short peptides (1). Therefore, determining the structure of protein–peptide complexes involved in these interactions is crucial for understanding the molecular mechanism and thus modulating the protein–protein

interactions for therapeutic purpose (3,4). However, compared to the large number of identified protein–peptide interactions, only a limited number of protein–peptide complex structures were experimentally determined due to the high cost and technical difficulties in experimental methods. Therefore, computational modeling like molecular docking has played an important role in the determination of protein–peptide complex structures. Starting from a protein structure and a peptide sequence, protein–peptide docking predicts the complex structure by sampling possible peptide binding conformations and ranking the putative protein–peptide complexes with an energy scoring function (2).

Compared to protein–ligand and protein–protein docking, protein–peptide docking faces two challenges. First, unlike protein–ligand docking in which the binding site is normally known, the information of the binding site for peptide is not available in many cases. Therefore, protein–peptide docking often requires a global search around the whole protein for putative binding modes. Second, compared to small compounds and proteins, peptides are much more flexible and do not have a stable conformation before binding to a receptor. It is computationally expensive to fully sample the peptide conformations in protein–peptide docking. To address these challenges, a variety of state-of-the-art algorithms have been developed for protein–peptide docking in the past few years (5–16).

Based on whether the information about the binding site is available, current protein–peptide docking algorithms can be grouped into two categories: local peptide docking and global peptide docking. When the information about the binding site is known or the binding sites are predicted by certain algorithms like PepSite (17) or PEP-SiteFinder (18), etc. (19), one can sample peptide binding conformations around a specific binding site through local peptide docking. The Rosetta FlexPepDock (13), PepCrawler (14) and HADDOCK peptide protocol (11) are well-known algorithms for local peptide-docking, where the protein–peptide binding modes were obtained through a high-resolution docking refinement within the binding site. When the binding site is unknown, global peptide docking is required to blindly searching for putative peptide binding conformations around the whole protein. Due to the relatively larger

*To whom correspondence should be addressed. Tel: +86 27 87543881; Fax: +86 27 87556576; Email: huangsy@hust.edu.cn

search space, global peptide docking is much more challenging than local peptide docking. Recently, several global peptide docking algorithms such as AnchorDock (12), CABS-dock (10), pepATTRACT (8) and MDockPep (7) have been developed for the blind prediction of protein–peptide complexes, among which CABS-dock is available as a web server and pepATTRACT has a web version for its rigid docking protocol (20). However, all of these blind peptide-docking algorithms suffer from lengthy simulations in peptide binding refinement and often take hours on a GPU or multi-core CPUs for docking a peptide. Although the pepATTRACT web server is fast, the peptide flexibility is not sufficiently considered because it only considers three idealized conformations for a peptide (20).

To efficiently consider the peptide flexibility in peptide docking, we have developed a hierarchical algorithm for blind and flexible peptide docking by fast modeling of peptide conformations and sequent global sampling of binding orientations, which is referred to as HPEPDOCK. In our docking algorithm, the peptide flexibility is considered by generating an ensemble of peptide conformations with our efficient MODPEP program (21). Then, the sampled peptide conformations are globally docked against the whole protein using the rigid docking protocol of MDock (22,23). Compared to pepATTRACT which only accounts for limited peptide flexibility through three idealized conformations of a peptide, HPEPDOCK is able to sufficiently consider the peptide flexibility by generating a large number of peptide conformations (up to 1000). HPEPDOCK is also efficient for local peptide docking if information about the binding site is available. We have participated in recent CAPRI experiments and achieved a good performance in the peptide-docking challenges (24–26). Here, we propose a web server of our peptide-docking algorithm that was used in our CAPRI experiments. Compared to current peptide docking servers, our HPEPDOCK server accepts not only structures but also sequences as input for the protein and can automatically integrate the available peptide binding information from the Protein Data Bank (PDB) (27). In addition, HPEPDOCK may also be used for nucleic acids–peptide docking. The docking process is fully automatic and the results are presented interactively to users through a web page.

MATERIALS AND METHODS

The HPEPDOCK docking protocol

HPEPDOCK is a hierarchical flexible-peptide docking protocol that integrates our MODPEP program for peptide conformational sampling (21), a modified version of MDock for peptide docking (22,23), several third-party programs including HHSuite (28), FASTA (29), ClustalW (30,31), and MODELLER (32) for structure modeling, and a set of tools developed in our group. The docking pipeline implemented in HPEPDOCK is illustrated in Figure 1.

The first step of the docking protocol is to provide inputs for the receptor and the peptide. The server accepts both sequences and structures as input for the protein and the peptide. Users are also given an option to provide the information about binding site and the number of binding modes to be output.

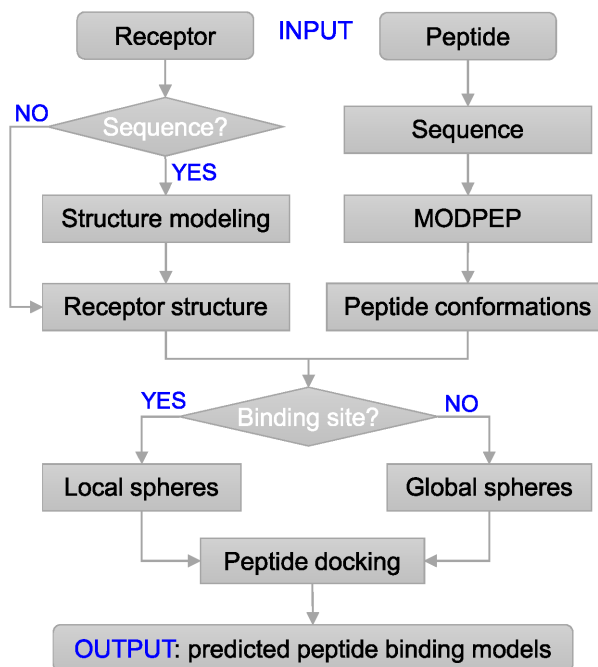


Figure 1. The workflow of HPEPDOCK web server for hierarchical flexible peptide docking that supports both local and global docking.

With the inputs, the next step of HPEPDOCK is to prepare the receptor and peptide structures. For the receptor, if the input is a structure, the server will directly use the structure for docking. If the input for the protein is a sequence, its three dimensional (3D) structure will be constructed through homology modeling as follows. Given a protein sequence, a sequence similarity search is conducted against the PDB sequence database to find the homologous templates for the target protein. Here, the HHSuite package is used for sequence search (28), due to its efficient detection of remote homologs. If multiple templates are available, the one with the highest sequence coverage, the highest sequence similarity, and the highest resolution is selected. With the selected template, the 3D structure of the protein is built using MODELLER (32), in which the sequence alignment is conducted by ClustalW (30,31). For the peptide, if the input is a structure, it will be converted to a sequence. Our MODPEP program (21) is then used to generate an ensemble of protein-bound peptide conformations from the sequence converted from the structure or submitted by users. The number of generated peptide conformations is set to 1000 by default in the web server.

With the receptor structure and an ensemble of peptide conformations, a modified version of MDock is used to dock the ensemble of peptide conformations against the whole receptor (22,23). Here, only the rigid-protein docking protocol of MDock is used for protein–peptide docking, in which a single protein structure is used during the docking calculation. The peptide flexibility is considered by docking an ensemble of multiple peptide conformations generated with our MODPEP program. Specifically, the molecular surface for the receptor is first calculated using the DMS program (33). Next, the sphere points that repre-

sent the negative images of the molecular shape are generated by SPHGEN (34). If the information about the binding site is not provided by users, the sphere points, which cover the whole receptor, will be clustered and used to represent the possible peptide binding positions on the receptor. Then, the putative peptide binding modes are globally sampled by matching the peptide atoms with the sphere points (34). If information about the binding site is provided by users, the sphere points around the binding site are selected and the HPEPDOCK server will perform a local peptide docking. The putative peptide binding orientations will be generated by matching the peptide onto the sphere points (34). To adapt MDock for protein-peptide docking, we have modified MDock in two aspects. First, we used a reduced model of peptide during the generation of putative peptide binding orientations. Namely, each residue is represented by two pseudo atoms corresponding to the CA atom and the center of mass for the rest atoms of the residue, so as to reduce the memory consumption and accelerate the sampling process. After the peptide orientations are generated with the reduced model, they are replaced by the all-atom models for the evaluation of binding energy scores. Second, we replaced the original scoring function for protein-ligand interactions with an iterative knowledge-based scoring function for protein-protein interactions (35). The sampled binding orientations are optimized by a SIMPLEX minimization algorithm guided by their binding energy scores, in which both the protein and the peptide are treated as rigid bodies. The final docking models are interactively provided to users through a web page. On the result page, users can download the docking results and view the top 10 binding models through a Jmol web interface (36).

For the computational efficiency of structure modeling, a local copy of the PDB database is maintained on our web server and updated monthly. It should be noted that although users can submit either sequences or structures as inputs for the protein, there are some differences between them in the HPEPDOCK pipeline. With a sequence input, the server will search for its homologous templates and then build the 3D structure based on the template. If no template is found, docking will not be conducted. Nevertheless, given the capability of HHSuite and FASTA in detecting remote homology, sequence inputs are applicable in most of practical applications (28,29).

Input

The inputs for the receptor and the peptide are required for HPEPDOCK to perform a peptide docking job. For protein and peptide inputs, both structure and sequence are supported. For users' convenience, the server accepts four types of input for protein and peptide, two for structure and two for sequence, as follows:

- Upload a pdb file in PDB format.
- Provide a structure by PDB ID:ChainID (e.g. 3BFW:A).
- Copy and paste a protein/peptide sequence in FASTA format.
- Upload a protein/peptide sequence file in FASTA format

Only one type of input is needed. For the protein, the maximum number of amino acids is set to 2000 and the maximum number of atoms is set to 20 000. Modified amino acids are not supported. For the peptide, there is no limit for the length, but sequences of <30 amino acids are recommended for the sake of accuracy. For structure input, users can upload their own pdb files or provide a structure by PDB: chain ID(s). Since the server is now designed to model single-chain protein from sequence, users are recommended to upload their own structure if the protein contains multiple chains.

By default, the top 100 peptide binding models are output, while users can change the output number within the range from 1 to 500 when submitting their job.

In addition, users also have an option to provide the binding site information for the peptide by giving receptor residues in a text box or uploading a reference file. The binding site information, if provided, will be used as a reference to select the sphere points for local peptide docking. Users may also give a name to their job and provide a valid email address for notification of job completion.

Benchmarks

To evaluate the performance of our HPEPDOCK docking protocol, we have used two benchmarks of unbound/unbound protein-peptide complexes from peptiDB (37). One is for blind global peptide docking, and the other is for local peptide docking when the information about binding site is provided.

The test set for global peptide docking was constructed based on the 80 unbound/unbound protein-peptide complexes from the peptiDB docking benchmark (8,11,37) with two criteria. First, the stoichiometry of the biological unit is 1:1; Second, the unbound structure should have a sequence identity of >90% with the corresponding bound structure. Of 80 protein-peptide complexes, 57 cases met these two criteria (Supplementary Table S1). All the structures were downloaded from the PDB and the parts in the unbound structure that are not present in the bound structure were removed.

Since our HPEPDOCK server supports sequence input for protein, we have constructed a benchmark of protein model-peptide sequence test cases based on the same global docking benchmark of 57 protein-peptide complexes. Specifically, for each case in the benchmark, the structure of the protein was automatically constructed from its sequence by homology modeling, in which the native structure was excluded from the templates. As shown in Supplementary Table S1, the constructed homology models have a wide range of conformational changes and give an RMSD from the lowest 0.32 Å for targets 2CCH and 2VJ0 and the highest 10.93 Å for target 2P1T. Here, the RMSD was calculated based on the backbone atoms of the protein after optimal superimposition of the homology model with the bound structure. The average RMSD of homology models is 1.05 Å from the bound structures.

The benchmark for local peptide docking consists of 62 unbound-unbound protein-peptide complexes from peptiDB (Supplementary Table S2) (11,37). The benchmark has been used to evaluate the HADDOCK peptide docking

protocol for local peptide docking. Therefore, docking results are available for the HADDOCK peptide protocol on this benchmark, which can be used as a reference to evaluate the local peptide docking protocol of our HPEPDOCK algorithm.

Evaluation criteria

The quality of a predicted protein–peptide binding mode is measured by its interface RMSD (IRMSD) from the native structure. The IRMSD is calculated based on the backbone atoms of protein and peptide residues that are within 10 Å from the partner molecules in the native structure (38). If a binding mode has an IRMSD of ≤ 2.0 Å, it is defined as a successful prediction (or a hit). For evaluation on a benchmark, the docking performance is measured by the success rate, i.e. the percentage of the cases with at least one hit when a certain number of top predictions are considered.

RESULTS

HPEPDOCK server

The hardware supporting the HPEPDOCK server is a Linux cluster of two compute nodes, each of which includes two Intel(R) Xeon E5-2690 v4 2.60 GHz CPUs with 28 cores and 256 GB of Memory. The web server is based on Apache HTTP, HTML, PHP and the JSmol web applet for the docking pipeline and binding model visualization. The SLURM Workload Manager (39) is used as the job scheduler of HPEPDOCK server. For global peptide docking, four CPU cores are assigned to a job for parallel running (40), while for local peptide docking, one CPU core is assigned to a job. A maximum of 50 jobs can be running at the same time while hundreds of jobs can be queued in the background. The web service does not require registration and can be freely accessed.

After users submit their input data, the web interface will be redirected to a web page showing the job ID and running status. The job status including ‘QUEUED’, ‘RUNNING’, and ‘RESULTS’ will be updated every 10 s on the status page. The URL to the docking results is something like <http://huanglab.phys.hust.edu.cn/hpepdock/data/jobid>, where ‘jobid’ is a unique job ID. Users can bookmark the status page for access to the docking results at a later time. Users will also be notified by email when the job is finished if a valid email address is provided at the time of job submission.

Output

The docking output is provided to users through a result web page when the job is done, as shown in Figure 2. The docking results include three types of files for download:

- Receptor PDB file uploaded by users or constructed by the server from the user-provided FASTA sequence.
- The individual peptide binding models for the top 20 predictions.
- The compressed packages for the top 10 predictions, the top 100 predictions, and all the docking results.

Since the top 10 binding models are normally deemed the most important predictions in docking calculations, the result page provides an interactive view of the top 10 models using the Jmol software (36). Users can choose to view any of the top 10 models or all together by different representations and styles.

The result page also gives a summary of the docking scores and rankings for the top 10 binding models. However, it should be noted that the docking scores here do not reflect the real binding affinities, but a relative ranking among different binding models. It is recommended that users download their docking results as soon as possible after their job is done, as the job results will only be stored on our server for two weeks.

In addition, if only a sequence is provided as the input for a protein, the result page will also show the information of the protein model built by homology modeling, including the used template, and the sequence identity and sequence alignment between the template and the input sequence.

Performance of the HPEPDOCK server

Global docking with unbound structure. The HPEPDOCK server has been tested on the benchmark of 57 unbound/unbound protein–peptide complexes for blind global docking. Figure 3 shows the success rate and average number of hits per complex in binding mode prediction by HPEPDOCK when several numbers of top predictions were considered. The IRMSDs of individual cases are listed in Supplementary Table S3. For comparison, we have submitted the same benchmark of 57 unbound-unbound cases to the pepATTRACT server and downloaded the predicted peptide binding models from the server. The same criteria were used to assess the pepATTRACT models and the corresponding IRMSDs of individual cases are listed in Supplementary Table S4. It can be seen from Figure 3A that HPEPDOCK achieved a significantly better performance than pepATTRACT server and obtained a success rate of 15.8% and 33.3% when the top 1 and 10 predictions were considered, compared to 10.5% and 21.1% for pepATTRACT server. Similar trend can be observed in the average number of hits per complex (Figure 3B). For example, HPEPDOCK obtained an average of 0.912 hits per complex within top 10 predictions, compared to 0.614 hits for pepATTRACT sever. Although HPEPDOCK had a comparable success rate to pepATTRACT server when the top 50 predictions were considered (Figure 3A), HPEPDOCK performed significantly better than pepATTRACT server in the average number of hits per complex (Figure 3B). The overall better performance of HPEPDOCK than pepATTRACT server indicates the benefit of considering the peptide flexibility through an ensemble of peptide conformations in our hierarchical algorithm.

Global docking with homology model. The HPEPDOCK server has also been evaluated on the peptide-homology modeling benchmark of 57 cases. Figure 3 shows the success rate and average number of hits per complex of HPEPDOCK with homology models. The IRMSDs of individual cases are listed in Supplementary Table S5. It can be seen from Figure 3 that HPEPDOCK achieved comparable per-

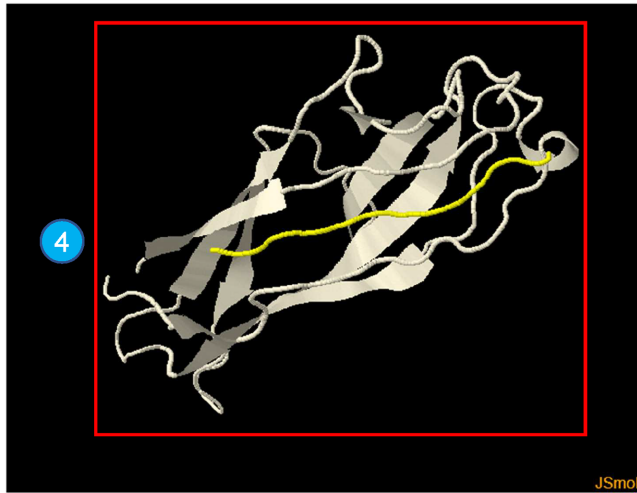
Your HPEPDOCK results for job 5a61ef7ce6e95

Download Files

[Receptor PDB file](#)

[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20]

[Top 10 Predictions](#) [Top 100 Predictions](#) [All the results in a package](#)



MODEL No.

Model 1

Model 2

Model 3

Model 4

Model 5

Model 6

Model 7

Model 8

Model 9

Model 10

Color By:

Model

Chain

Group

Action:

Spin

Stereo

Style:

Cartoon

Spacefill

Wireframe

Summary of the top 10 models

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-436.035	-286.762	-282.055	-279.462	-277.385	-265.734	-263.059	-253.898	-253.375	-252.432
(a)	Row 1: The ranks of the models.									
(b)	Row 2: The docking energy scores.									

huanglab@hust.edu.cn, Lab of Biophysics and Molecular Modeling

Figure 2. The HPEPDOCK server result page. At the top of the page is the job name or a unique job ID (1), and the files for download (2). Optional buttons on the right can control Jmol to visualize the binding model (3) on the left (4). The docking scores of the top 10 peptide models are shown on the bottom (5).

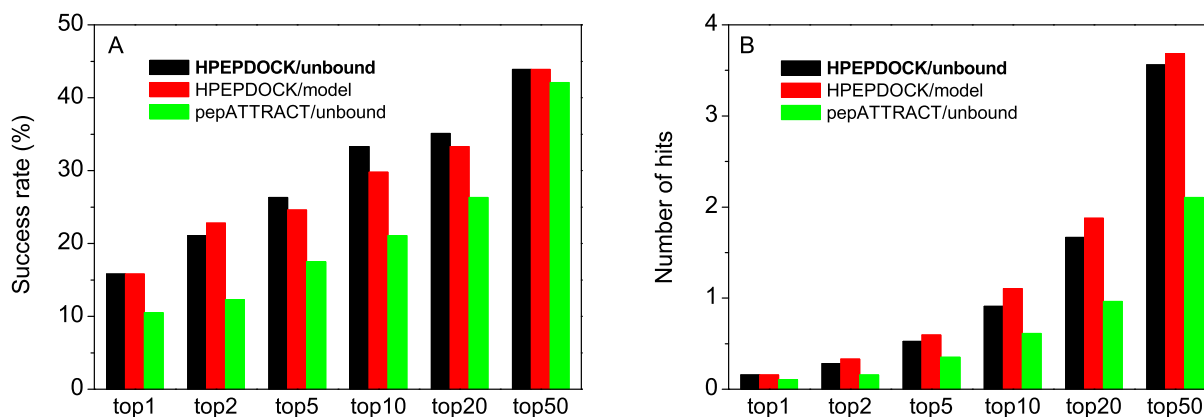


Figure 3. The success rate (A) and average number of hits per complex (B) in binding mode predictions by HPEPDOCK and pepATTRACT server for global peptide docking on the benchmark of 57 test cases when several numbers of top predictions were considered. For each number of top predictions, from left to right are HPEPDOCK with unbound structure, HPEPDOCK with homology model, and pepATTRACT server with unbound structure of the protein, respectively. The corresponding IRMSDs of individual cases are listed in Supplementary Tables S3-S5.

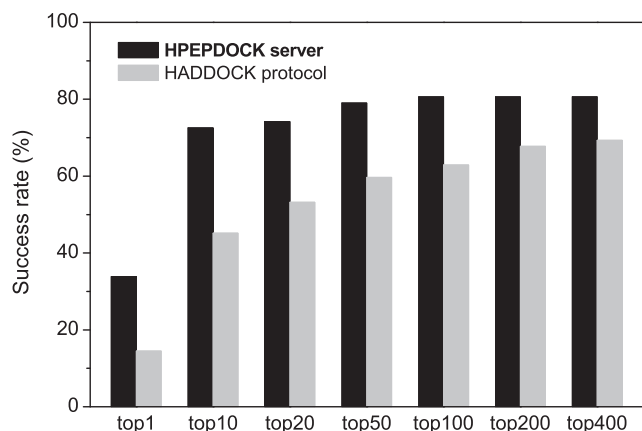


Figure 4. The success rates in binding mode predictions by HPEPDOCK server and HADDOCK peptide protocol for local peptide docking on the benchmark of 62 unbound cases when several numbers of top predictions were considered. The corresponding IRMSDs of individual cases are listed in Supplementary Table S6.

performances when docking against homology models and unbound structures. For example, HPEPDOCK with homology models obtained a success rate of 15.8%, 29.8% and 43.9% when the top 1, 10 and 50 predictions were considered, compared to 15.8%, 33.3% and 43.9% for HPEPDOCK server with unbound structures (Figure 3A). Similar trend can be observed in the average number of hits per complex for docking with homology models and unbound structures (Figure 3B). The comparable performances of HPEPDOCK with homology model and unbound structure suggests the robustness of our algorithm in docking with sequence input.

Local docking with unbound structure. When information about the binding site is given, HPEPDOCK server will run local peptide docking. The benchmark of 62 unbound/unbound cases used by HADDOCK peptide protocol was used to test the performance of the local docking protocol of HPEPDOCK (Supplementary Table S2). Specifically, the molecular surface for the protein atoms within 5 Å from the native peptide was first generated using the DMS program (33). Next, the sphere points that represent the negative images of the molecular shape were generated using the SPHGEN algorithm (34), and selected to represent the binding site on the protein. Figure 4 shows the success rates of the local peptide docking protocol of HPEPDOCK server on the benchmark of 62 unbound cases when several numbers of top predictions were considered. The IRMSDs of individual cases are listed in Supplementary Table S6. For comparison, the figure also shows the corresponding success rates by the HADDOCK peptide docking protocol on the same benchmark, which were taken from the literature (11). It can be seen from Figure 4 that HPEPDOCK achieved a significantly better performance than HADDOCK and obtained a success rate of 33.9%, 72.6%, and 80.6% in binding mode prediction when the top 1, 10 and 100 binding modes were considered, compared to 14.5%, 45.2% and 62.9% for HADDOCK. When the top 400 models were considered, HPEPDOCK maintained the

same success rate of 80.6%, compared to 69.4% for HADDOCK (Figure 4). As a comparison, the local docking protocol of pepATTRACT obtained an overall success rate of 79% within top 1000 models (8). The significantly better performance of HPEPDOCK than HADDOCK demonstrates the strong predictive power of HPEPDOCK server in local peptide docking.

Computational efficiency

Figure 5 shows the running times of HPEPDOCK server for the protein–peptide docking jobs on two benchmarks. The detailed values of running times are listed in Supplementary Tables S3 and S6. It can be seen from Figure 5 that HPEPDOCK tends to consume more time for longer peptides, as expected, and can finish a global docking job as short as 5.5 mins for target 1JWG with a peptide of 5 aa. The longest running time happens to target 1JBU with a 15-aa peptide (Supplementary Table S3). On average, HPEPDOCK server consumed 29.8 mins for a global protein–peptide docking job (Figure 5A). Similar trend can be observed in the relationship between running time and peptide length for local peptide docking (Figure 5B). HPEPDOCK server consumed an average of 14.2 mins for a local peptide docking job with the longest time for target 1ER8 (34.8 mins) and the shortest time for 1NVR (3.2 mins) (Supplementary Table S6).

Examples of the docking models

Figure 6A shows an example of the docking model by HPEPDOCK web server. The protein–peptide target (PDB code: 3BFW) is a complex between the truncated FimG (FimGt) and the donor strand peptide of FimF (DSF). Sequences were provided as input for both the protein and the peptide when the docking job was submitted. Among the top 10 constructed models, there are six near-native predictions with an interface RMSD of <math><2.0\text{ \AA}</math>, of which the first model has a high accuracy of IRMSD = 0.84 Å and the pose 10 gives the best accuracy with IRMSD = 0.51 Å.

In addition to performing protein–peptide docking, HPEPDOCK can also be used to dock a peptide against nucleic acids (DNAs/RNAs). Figure 6B gives an example of the predicted peptide binding model on a DNA by HPEPDOCK. The target 2EZf is a complex between a truncated form of HMG-I(Y) and a DNA dodecamer. For this target, the pdb structure was submitted for the DNA and the sequence was provided for the peptide. Among top 10 constructed models, there are three acceptable predictions according to the CAPRI criteria (38), of which the pose 10 has the best accuracy with an IRMSD of 3.11 Å.

CONCLUSION

We have presented the user-friendly HPEPDOCK web server for blind prediction of protein–peptide complex structures. The server implements a hierarchical docking protocol with fast conformational sampling of peptide conformations and ensemble docking of generated peptide conformations against the protein. The docking server accepts both sequence and structure as input for protein/peptide.

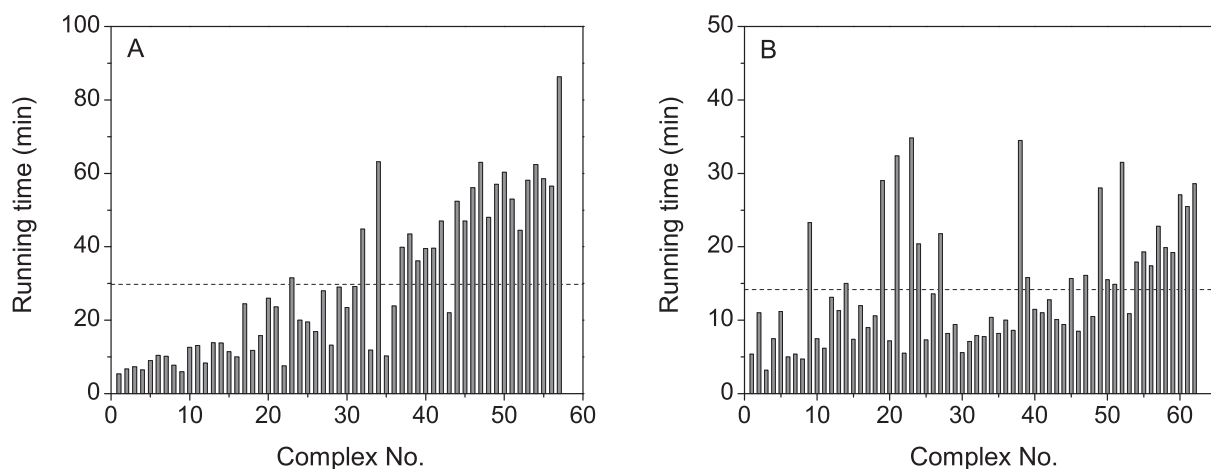


Figure 5. The running time of HPEPDOCK server for a protein–peptide docking job through global peptide docking (A) and local peptide docking (B), where the complex No. is consistent with that in Supplementary Tables S1 and S2, respectively. The dashed lines indicate their average running times over all the cases of the benchmark. The corresponding running times of individual cases are listed in Supplementary Tables S3 and S6.

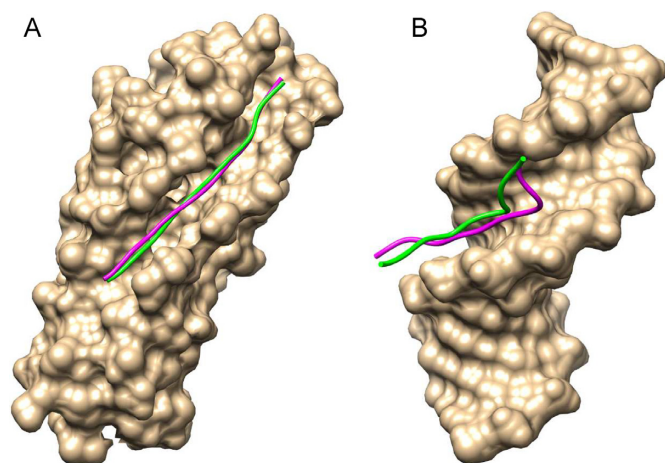


Figure 6. Comparison between the crystal structure (green) and HPEPDOCK server prediction (magenta) for two peptide docking examples where the receptor is represented in molecular surface: (A) protein–peptide target (PDB code: 3BFW; IRMSD = 0.51 Å); (B) DNA-peptide target (PDB code: 2ZKF; IRMSD = 3.11 Å),

The HPEPDOCK docking protocol was extensively tested on two benchmarks of unbound/unbound protein–peptide complexes from peptiDB and compared with state-of-the-art HADDOCK peptide docking protocol and pepATTRACT server. It was shown that overall HPEPDOCK achieved a significantly better performance in predicting near-native models for global and local peptide docking than pepATTRACT server and HADDOCK peptide protocol. The HPEPDOCK server can also be used to dock peptides against nucleic acids. HPEPDOCK is computationally efficient and the average running time is 29.8 mins for a global peptide docking job and 14.2 mins for a local peptide docking job. The accuracy of the HPEPDOCK server may be further improved by an additional refinement of output docking models in the future development. The two benchmarks of protein–peptide complexes and predicted peptide binding models by HPEPDOCK

and pepATTRACT server are available for download from our web site at http://huanglab.phys.hust.edu.cn/hpepdock/hpepdock_test.tgz

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Key Research and Development Program of China [2016YFC1305800, 2016YFC1305805]; National Natural Science Foundation of China [31670724]; startup grant of Huazhong University of Science and Technology. Funding for open access charge: NKRDP [2016YFC1305800, 2016YFC1305805].

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.
- London, N., Raveh, B. and Schueler-Furman, O. (2013) Peptide docking and structure-based characterization of peptide binding: from knowledge to know-how. *Curr. Opin. Struct. Biol.*, **23**, 894–902.
- Fosgerau, K. and Hoffmann, T. (2015) Peptide therapeutics: current status and future directions. *Drug Discov. Today*, **20**, 122–128.
- Craik, D. J., Fairlie, D. P., Liras, S. and Price, D. (2013) The future of peptide-based drugs. *Chem. Biol. Drug Des.*, **81**, 136–147.
- 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., Thevenet, 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.
- 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.
- Schindler, C. E., de Vries, S. J. and Zacharias, M. (2015) Fully blind peptide–protein docking with pepATTRACT. *Structure*, **23**, 1507–1515.
- Pallara, C., Jimenez-Garcia, B., Romero, M., Moal, I. H. and Fernandez-Recio, J. (2017) pyDock scoring for the new modeling

- challenges in docking: protein-peptide, homo-multimers, and domain-domain interactions. *Proteins*, **85**, 487–496.
10. 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.
 11. Trellet, M., Melquiond, A.S. and Bonvin, A.M. (2013) A unified conformational selection and induced fit approach to protein-peptide docking. *PLoS One*, **8**, e58769.
 12. Ben-Shimon, A. and Niv, M.Y. (2015) AnchorDock: blind and flexible anchor-driven peptide docking. *Structure*, **23**, 929–940.
 13. 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.
 14. Donsky, E. and Wolfson, H.J. (2011) PepCrawler: a fast RRT-based algorithm for high-resolution refinement and binding affinity estimation of peptide inhibitors. *Bioinformatics*, **27**, 2836–2842.
 15. Antes, I. (2010) DynaDock: A new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility. *Proteins*, **78**, 1084–1104.
 16. Pierce, B.G. and Weng, Z. (2013) A flexible docking approach for prediction of T cell receptor-peptide-MHC complexes. *Protein Sci.*, **22**, 35–46.
 17. 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.
 18. Saladin, A., Rey, J., Thevenet, 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.
 19. Yan, C. and Zou, X. (2015) Predicting peptide binding sites on protein surfaces by clustering chemical interactions. *J. Comput. Chem.*, **36**, 49–61.
 20. de Vries, S.J., Rey, J., Schindler, C.E.M., Zacharias, M. and Tuffery, P. (2017) The pepATTRACT web server for blind, large-scale peptide-protein docking. *Nucleic Acids Res.*, **45**, W361–W364.
 21. Yan, Y., Zhang, D. and Huang, S.Y. (2017) Efficient conformational ensemble generation of protein-bound peptides. *J. Cheminform.*, **9**, 59.
 22. Huang, S.Y. and Zou, X. (2007) Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins*, **66**, 399–421.
 23. Huang, S.Y. and Zou, X. (2007) Efficient molecular docking of NMR structures: application to HIV-1 protease. *Protein Sci.*, **16**, 43–51.
 24. Lensink, M.F., Velankar, S. and Wodak, S.J. (2017) Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins*, **85**, 359–377.
 25. Yan, Y., Wen, Z., Wang, X. and Huang, S.Y. (2017) Addressing recent docking challenges: A hybrid strategy to integrate template-based and free protein-protein docking. *Proteins*, **85**, 497–512.
 26. Lensink, M.F., Velankar, S. and Wodak, S.J. (2017) Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins*, **85**, 359–377.
 27. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
 28. Remmert, M., Biegert, A., Hauser, A. and Soing, J. (2011) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods*, **9**, 173–175.
 29. Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 2444–2448.
 30. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.
 31. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
 32. Marti-Renom, M.A., Stuart, A., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 291–325.
 33. Richards, F.M. (1977) Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.*, **6**, 151–176.
 34. Ewing, T.J., Makino, S., Skillman, A.G. and Kuntz, I.D. (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.*, **15**, 411–428.
 35. Huang, S.Y. and Zou, X. (2008) An iterative knowledge-based scoring function for protein-protein recognition. *Proteins*, **72**, 557–579.
 36. Hanson, R.M., Prilusky, J., Renjian, Z., Nakane, T. and Sussman, J.L. (2013) JSmol and the next-generation web-based representation of 3D molecular structure as applied to Proteopedia. *Israel J. Chem.*, **53**, 207–216.
 37. London, N., Movshovitz-Attias, D. and Schueler-Furman, O. (2010) The structural basis of peptide-protein binding strategies. *Structure*, **18**, 188–199.
 38. Janin, J., Henrick, K., Moulton, J., Eyck, L.T., Sternberg, M.J.E., Vajda, S., Vasker, I. and Wodak, S.J. (2003) CAPRI: a critical assessment of predicted interactions. *Proteins*, **52**, 2–9.
 39. Jette, M.A., Yoo, A.B. and Grondona, M. (2002) SLURM: simple linux utility for resource management. In: *Lecture Notes in Computer Science: Proceedings of Job Scheduling Strategies for Parallel Processing (JSSPP)*, Springer-Verlag, Berlin, Vol. **2003**, pp. 44–60.
 40. Tange, O. (2011) GNU parallel – the command-line power tool. *login: The USENIX Mag.*, **36**, 42–47.