



HHS Public Access

Author manuscript

J Comput Chem. Author manuscript; available in PMC 2019 October 30.

Published in final edited form as:

J Comput Chem. 2018 October 30; 39(28): 2409–2413. doi:10.1002/jcc.25555.

MDockPeP: An *ab-initio* protein-peptide docking server

Xianjin Xu[#], Chengfei Yan[#], and Xiaoqin Zou^{*}

Dalton Cardiovascular Research Center, Department of Physics and Astronomy, Department of Biochemistry, Informatics Institute, University of Missouri, Columbia, MO 65211, USA

[#] These authors contributed equally to this work.

Abstract

Protein-peptide interactions play a crucial role in a variety of cellular processes. The protein-peptide complex structure is a key to understand the mechanisms underlying protein-peptide interactions and is critical for peptide therapeutic development. We present a user-friendly protein-peptide docking server, MDockPeP. Starting from a peptide sequence and a protein receptor structure, the MDockPeP Server globally docks the all-atom, flexible peptide to the protein receptor. The produced modes are then evaluated with a statistical potential-based scoring function, ITScorePeP. This method was systematically validated using the peptiDB benchmarking database. At least one near-native peptide binding mode was ranked among top 10 (or top 500) in 59% (85%) of the bound cases, and in 40.6% (71.9%) of the challenging unbound cases. The server can be used for both protein-peptide complex structure prediction and initial-stage sampling of the protein-peptide binding modes for other docking or simulation methods. MDockPeP Server is freely available at <http://zougrouptoolkit.missouri.edu/mdockpep>.

Graphical abstract

MDockPeP is a publicly accessible web server (<http://zougrouptoolkit.missouri.edu/mdockpep>) for predicting protein-peptide complex structures. The server requires only the peptide sequence and the protein structure. MDockPeP docks the all-atom, flexible peptide onto the whole protein without the knowledge of the binding site. MDockPeP is computationally efficient, and achieves excellent performance on mode sampling and good performance on mode prediction.

Keywords

Protein-peptide interactions; Complex structure prediction; Molecular docking; Molecular modeling; Web server

Introduction

Protein-peptide interactions are crucial to a variety of cellular processes including transcription regulation, signal transductions and immune response [1]. An increasing number of peptides have been designed and approved as drugs [2]. The structure of the

^{*}To whom correspondence should be addressed. zoux@missouri.edu.

Conflict of Interest: none declared.

protein-peptide complex is a key to understand the underlying mechanism of the protein-peptide interaction, and is therefore critical for peptide therapeutic development. Yet, the number of the resolved protein-peptide complex structures deposited in the Protein Data Bank (PDB) [3] is only a fraction of the whole protein-peptide interaction universe, due to the difficulty and cost for determining complex structures by experimental techniques such as X-ray crystallography and NMR.

Facing this challenge, several *in silico* methods have recently been developed for predicting protein-peptide complex structures and can be categorized into three classes: template-based modeling, molecular docking, and molecular dynamics (MD) simulation. The template-based methods are computationally efficient, but suffer from limited available protein-peptide templates [4–5]. On the other hand, regarding MD simulations, impractically expensive computational cost is the major stumbling block to their large-scale applications [6–7]. Molecular docking is a compromising strategy, which aims to account for both accuracy and computational efficiency. Among the recently developed docking methods, Rosetta FlexPepDock [8] and HADDOCK [9] focus on local docking with known binding sites. pepATTRACT [10] and AnchorDock [11] start with crudely sampling the whole protein surface, followed by extremely time-consuming MD refinement. The CABS-dock server [12] has the ability to dock a fully flexible peptide onto the whole protein surface within reasonable computational time. It uses a coarse-grained model for both the protein and the peptide; the peptide secondary structure either is provided by the user or is generated by PSI-PRED, a protein secondary structure prediction tool. PIPER-FlexPepDock [13] is another approach that performs the global blind docking. Briefly, a number of pre-generated peptide conformers are docked to a whole protein surface using a rigid sampling algorithm, and then the selected models are refined by considering the peptide flexibility and the protein sidechain flexibility. A thorough summary of state-of-the-art in the field can be found in a very recent review [14].

We recently developed a novel, *ab initio* protein-peptide docking method, referred to as MDockPeP [15]. The method starts with a given peptide sequence and a protein structure, and globally docks the all-atom, flexible peptide to the protein (Fig. 1). MDockPeP was systematically validated and achieved good performance based on the peptiDB benchmarking database [9,16]. Here, we present the MDockPeP Server, which is free and open to all users without registration. The server can be used for both protein-peptide complex structure prediction and initial-stage sampling of the protein-peptide binding modes for other docking or simulation methods.

Materials and Methods

Overview of MDockPeP

Here, we briefly introduce the MDockPeP method; the details are available in our recently published paper [15]. MDockPeP includes three primary stages (Fig. 1):

- (1) Model the peptide conformers based on the given peptide sequence;
- (2) Sample putative peptide binding modes on the whole protein surface;

- (3) Rank the sampled binding modes according to their energy scores with our newly derived scoring function for protein-peptide docking.

For a given peptide sequence, first, MDockPeP models up to 3 non-redundant conformers based on the similar-sequence fragments from monomeric proteins with lengths longer than 50 amino acids. This strategy is based on the argument that binding of a peptide on a protein is similar to the protein folding process and that protein-peptide binding interfaces share remarkable similarities with the interior of proteins [17]. Our systematic assessment showed that the modeled peptide conformer was within 5.3 Å of the backbone RMSD (bRMSD) in comparison with the bound peptide structure when the best conformer among the top 3 peptide models was considered [15], for the 103 non-redundant peptides in the peptiDB benchmarking database [9,16].

Next, the modeled peptide conformers are independently docked to the whole protein using a method modified from AutoDockVina [18]. The grid box was defined by extending 20 Å to both the minimum and the maximum of the coordinates of the protein structure in three dimensions. First, the peptide conformer is rigidly docked to the whole protein by randomly generating 10^5 translational and rotational configurations within the grid box. The generated models are ranked by the built-in Vina scoring function. Then, flexible sampling is performed for the model that has the lowest score. All rotatable bonds in the peptide are treated as flexible during sampling, by using the iterated local search (ILS) global optimizer approach in AutoDock Vina. If the peptide conformation of a Vina-accepted mode strays too far from the initial peptide conformer (e.g., with bRMSD > 5.5 Å), the rigid global sampling process will be repeated for the initial peptide conformer, followed by flexible sampling. The procedure stops when the maximum step number for ILS, N , is reached. N is dependent on both the number of torsional angles and the number of the movable atoms. The exhaustiveness value in Vina is set to 100 for the MDockPeP server, which means 100 independent runs are performed for each docking. Finally, up to 2×10^4 binding modes are generated for each initial peptide conformer.

These binding modes generated from different initial peptide conformers are combined and ranked according to their energy scores calculated by our recently developed scoring function ITScorePeP [15]. ITScorePeP is a statistical potential-based scoring function that is developed for protein-peptide dockings. Contributions from both interactions between the protein and the peptide (inter-score) and interactions among non-neighbored residues within the peptide (intra-score) are considered in the scoring function. For any two modes with ligand RMSD (L_{rms}) less than a cutoff, only the one with the lower score is kept. L_{rms} is calculated based on the backbone atoms of the peptide between the predicted binding mode and the native binding mode after the optimal superimposition of the protein structures. The cutoff is set to 4.0 Å for the prediction of top 10 models. For the enrichment of high-quality models ($L_{rms} < 3.0$ Å) in top 500 models that are provided to the user as the sampling results, the cutoff is set to 2.0 Å.

The peptiDB benchmarking database

The non-redundant protein-peptide database peptiDB was employed to validate the MDockPeP Server. After the examination of the 103 bound protein-peptide complex

structures and 69 unbound protein receptor structures, 3 bound complexes and 5 unbound protein receptors were discarded from the database [15]. The remaining entries, 100 bound cases and 64 unbound cases (see Table S1) were used to evaluate the sampling performance of the MDockPeP Server. The results in this study are slightly different from our original paper (Fig. 5 in ref. 14), in which the binding modes were sampled more exhaustively at a cost of longer computational time.

Server Description

Inputs

As shown in Fig. 2A, two inputs, a peptide sequence and a protein structure, are required for job submission on the MDockPeP Server. The email address is optional but recommended. If the email address is provided, the user will receive an email notification after the job is completed.

Advanced options

The MDockPeP Server provides several advanced options for the user to improve prediction results (as shown in Fig. 2B).

First, the server allows the user to upload one initial peptide 3D structure. The server will generate up to two other initial peptide conformers. As the peptide conformation during sampling is restricted to be relatively close to the initial peptide conformation, a reliable initial peptide structure would significantly reduce the search space and improve the prediction. Furthermore, the user is also allowed to control the degree of restriction of the peptide conformations in the sampling process by changing the cutoff value (default = 5.5 Å) of the backbone RMSD (bRMSD).

Another option is the exhaustiveness value. By increasing the exhaustiveness value, a larger conformational space can be reached during the sampling process at the cost of the increase in computational time. The default exhaustiveness value is set to 100, namely, each docking calculation (docking one initial peptide conformer onto the protein) contains 100 independent runs.

In addition, the user is allowed to define a binding location by providing the XYZ coordinates of the center of the grid box. The box (cubic) size will be automatically determined according to the peptide length. Specifically, the side of the cubic box equals $(3.8 \times \text{peptide_sequence_length} + 40)$ Å. The value 3.8 is the distance between two CA atoms in adjacent residues. This option is recommended for a large protein receptor with known binding location.

Outputs

Once a job is submitted successfully, the job status is monitored on the “Queue” page. If the email address is given, the user will receive an email notification with a link of the results after the job is completed. As shown in Fig. 2C, the top 10 predicted protein-peptide complex structures are displayed via 3Dmol.js [19] on the result page. In addition, top 500 predicted protein-peptide binding modes are provided as the initial sampling results.

Computational resources and run time

Jobs being submitted are performed on a computing node containing 24 Intel Xeon cores [Intel(R) Xeon(R) CPU E5-2650 v3 @ 2.30GHz]. For our test on the peptiDB database, the MDockPeP server normally takes less than 10 hours for a job depending on the length of the peptide and the size of the protein.

Performance

The MDockPeP Server was assessed with a non-redundant protein-peptide benchmarking database peptiDB (Table S1). As shown in Fig. 3A, the MDockPeP server successfully predicted at least one near-native ($L_{rms} \leq 5.5 \text{ \AA}$) mode among the top 10 models for 59% of the bound docking cases (high quality model with $L_{rms} \leq 3.0 \text{ \AA}$: 36%; medium quality model with $3.0 \text{ \AA} < L_{rms} \leq 5.5 \text{ \AA}$: 23%), and for 40.6% of the more challenging unbound docking cases (high quality: 3.1%; medium quality: 37.5%). Here, L_{rms} is the ligand RMSD, which is calculated based on the backbone atoms of the peptide between the predicted binding mode and the native binding mode after optimal superimposition of the protein structures.

Fig. 3B shows the rates for successfully ranking at least one near-native mode among the top N models. Impressively, bound docking achieved a high success rate of 77% when top 100 models were considered. The success rate decreased to 60.9% for the challenging unbound docking cases. For enrichment studies (see Fig. 3C), when considering the top 500 models that are provided for the user in the sampling results, the successful rate is 85% for bound docking cases (high quality: 65%; medium quality: 20%), and 71.9% for the unbound docking cases (high quality: 36%; medium quality: 36.9%).

Discussion

In our previous MDockPeP method paper [15], we analyzed the relationship between the best sampled binding mode (the mode with the lowest L_{rms}) and bRMSD of the best modeled peptide conformer. Because a smaller exhaustiveness value (100) was used for the web server than the exhaustive value for the method paper (500), we re-calculated the correlations. Fig. 4A and 4B show the results for the bound docking cases and the unbound docking cases, respectively. Similar to those observed in the method paper, L_{rms} and bRMSD show very weak correlations, with Pearson correlation coefficients of 0.19 (for bpro-upep) and 0.14 (for upro-upep), respectively. Encouragingly, our sampling method successfully generated medium-quality or even high-quality models for several cases in which no high-quality peptide conformers were modeled (using bRMSD = 4.0 Å as the threshold).

Fig. 4C and 4D show the dependence of the sampling performance on the peptide size for the bound docking cases and the unbound docking cases, respectively. The peptide lengths in the peptiDB benchmark range from 5 to 15. MDockPeP was able to generate high-quality models ($L_{rms} \leq 3 \text{ \AA}$) for most cases with short- or medium-size peptides (less than 12 residues). For a number of cases with peptide length ≥ 12 , our method failed to generate high-quality models or even medium-quality models. This is reasonable, because long

peptides typically contain more rotatable bonds than short peptides and therefore require larger conformational spaces for sampling. Another concern is the use of the same L_{rms} threshold for different peptide lengths. It is well known that the RMSD value is dependent of the size of a ligand [20]. How to normalize the RMSD value based on the ligand size remains an open question.

It is further noted that to optimize the performance of the MDockPeP server for users, the whole peptiDB database were used for the training of the scoring function. Overfitting is not expected to be an issue, because in our method paper [15] 3-fold cross-validation was used to assess the scoring function to avoid overlap between the training set and the test set; no significant difference was found between the two scoring performances. In both scoring studies, the success rate of unbound docking is significantly lower than the success rate of bound docking. A possible reason is that the decoys used in the training process were generated using bound protein structures and the protein structures were treated as rigid bodies in the sampling process. The scoring function needs to be improved in future studies.

Conclusion

The MDockPeP Server provides a useful and efficient means to produce models of protein-peptide complexes via a user-friendly web interface. The server can be used for both protein-peptide complex structure prediction and initial-stage sampling of the protein-peptide binding modes for other docking or simulation methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This work is supported by NSF CAREER Award DBI-0953839 and the NIH R01GM109980 (X.Z.). The computations are performed on the high-performance computing infrastructure supported by NSF CNS-1429294 (principal investigator, Chi-Ren Shyu) and the HPC resources supported by the University of Missouri Bioinformatics Consortium (UMBC).

References

1. Petsalaki E, Russell RB RB, Curr. Opin. Biotechnol 2008, 19, 344–350. [PubMed: 18602004]
2. Fosgerau K, Hoffmann T, Drug Discov Today, 2015, 20,122–128. [PubMed: 25450771]
3. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE, Nucleic Acids Res. 2000, 28, 235–242. [PubMed: 10592235]
4. Verschueren E, Vanhee P, Rousseau F, Schymkowitz J, Serrano L, Structure, 2013, 21, 789–797. [PubMed: 23583037]
5. Lee H, Heo L, Lee MS, Seok C, Nucl. Acids Res 2015, 43(W1), W431–W435. [PubMed: 25969449]
6. Niv MY, Weinstein H, J. Am. Chem. Soc 2005, 127, 14072–14079. [PubMed: 16201829]
7. Antes I, Protein 2010, 78, 1084–1104.
8. Raveh B, London N, Zimmerman L, Schueler-Furman O, PLOS ONE 2011, 6, e18934. [PubMed: 21572516]

9. Trellet M, Melquiond AS, Bonvin AM. PLOS ONE 2013, 8, e58769. [PubMed: 23516555]
10. Schindler CE, de Vries SJ, Zacharias M, Structure 2015, 23, 1507–1515. [PubMed: 26146186]
11. Ben-Shimon A, Niv MY, Structure, 2015, 23, 929–940. [PubMed: 25914054]
12. Kurcinski M, Jamroz M, Blaszczyk M, Kolinski A, Kmiecik S, Nucl. Acids Res 2015, 43(W1), W419–W424. [PubMed: 25943545]
13. lam N, Goldstein O, Xia B, Porter KA, Kozakov D, Schueler-Furman O, PLOS Comput. Biol 2017, 13: e1005905. [PubMed: 29281622]
14. Ciemny M, Kurcinski M, Kamel K, Kolinski A, Alam N, Schueler-Furman O, Kmiecik S, Drug Discov. Doday 2018, doi: 10.1016/j.drudis.2018.05.006
15. Yan C, Xu X, Zou X, Structure, 2016, 24, 1842–1853. [PubMed: 27642160]
16. London NN, Movshovitz-Attias D, Schueler-Furman O, Structure, 2010, 18, 188–199. [PubMed: 20159464]
17. Vanhee P, Stricher F, Baeten L, Verschueren E, Lenaerts T, Serrano L, Rousseau F, Schymkowitz J, Structure, 2009, 17, 1128–1136. [PubMed: 19679090]
18. Trott O, Olson AJ, J. Comput. Chem 2010, 31, 455–461. [PubMed: 19499576]
19. Rego N, Koes D, Bioinformatics 2014, 31, 1322–1324. [PubMed: 25505090]
20. Irving JA, Whisstock JC, Lesk AM, Proteins 2001, 42, 378–382. [PubMed: 11151008]

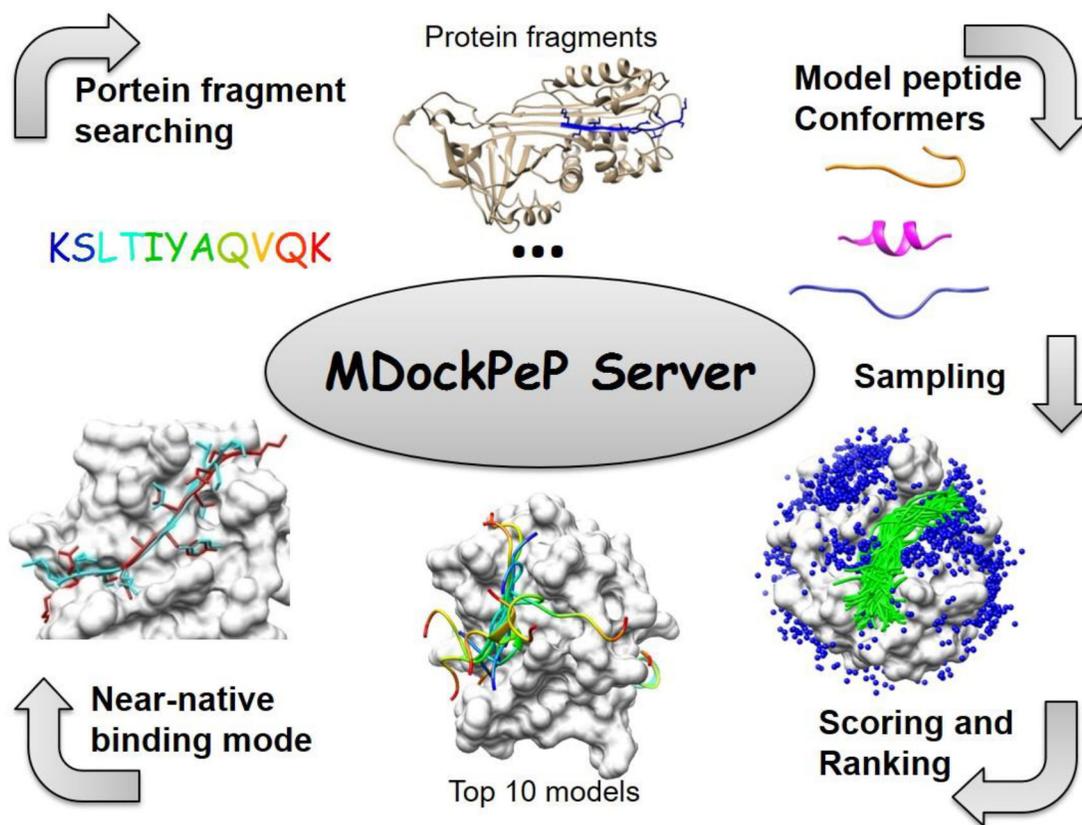


Figure 1.
The flowchart of the MDockPeP Server.

A

Job Submission

Inputs

→ Peptide sequence ? : Or upload a FASTA file: No file selected.

→ Protein structure ? : No file selected.

→ Job name (optional) :

→ Email address (optional) ? :

Example: Peptide sequence : 1AWQ_pep.fasta ; Protein receptor: 1AWQ_rec.pdb

B

Advanced options (optional)▼

→ Peptide structure in PDB format: No file selected.

→ The number of initial peptide models to used for docking:

→ The cutoff of bRMSD (Å) for the restriction of the peptide conformation: ?

→ The exhaustiveness value for the sampling: ?

→ The center of the binding site ? : / /

→ Do not show my job on the queue page

C

Top 10 predicted binding modes

Predicted models are ranked by a knowledge-based scoring function, ITScorePeP.
Click the "View" button on the right panel to display the corresponding binding mode.



Models	Active	View	Download
Model 1		<input type="button" value="View"/>	Download
Model 2		<input type="button" value="View"/>	Download
Model 3		<input type="button" value="View"/>	Download
Model 4		<input type="button" value="View"/>	Download
Model 5		<input type="button" value="View"/>	Download
Model 6		<input type="button" value="View"/>	Download
Model 7		<input type="button" value="View"/>	Download
Model 8		<input type="button" value="View"/>	Download
Model 9		<input type="button" value="View"/>	Download
Model 10		<input type="button" value="View"/>	Download

Figure 2.

Snapshots of the MDockPeP server. (A) Two inputs, a peptide sequence and a protein structure, are required for job submission. The email address is optional but recommended. The user will receive a notification through email once the job is completed. (B) Prediction results can be improved by adjusting the parameters in "Advanced options". For example, the sampling performance can be significantly improved by either increasing the exhaustiveness value (default = 100) or providing the binding site information. (C) The top 10 protein-peptide complex structures are reported and can be viewed online.

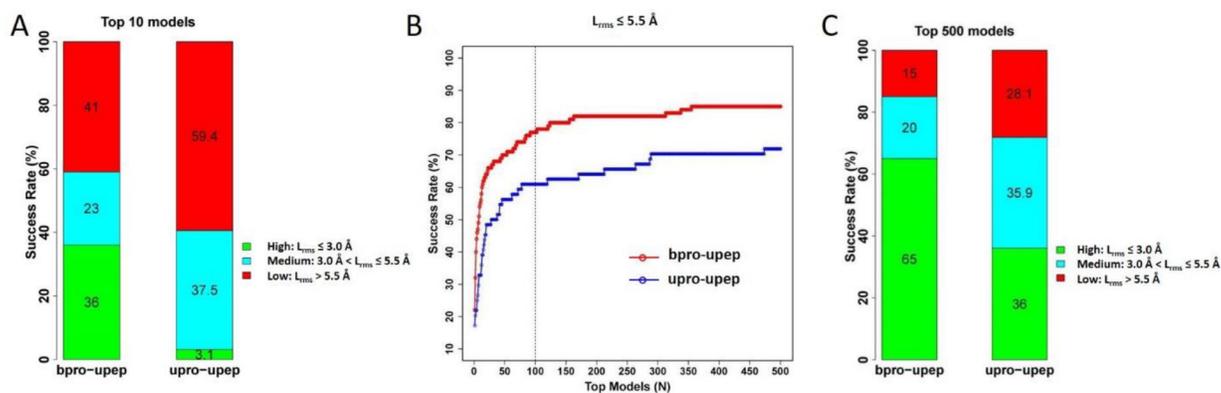


Figure 3.

The performance of MDockPeP server on the peptiDB database. (A) The performance according to the top 10 binding modes for the bound docking cases (bpro-upep) and unbound docking cases (upro-upep), respectively. (B) The success rates of ranking at least one near native ($L_{rms} \leq 5.5 \text{ \AA}$) mode among the top N models for the bound docking cases (bpro-upep) and the unbound docking cases (upro-upep), respectively. The dashed black line denotes the case in which N equals 100, for which the total success rates are 77% and 60.9% for the bound docking cases and the unbound docking cases, respectively. (C) The performance regarding the top 500 modes based on the criterion of L_{rms} for the bound docking cases (bpro-upep) and unbound docking cases (upro-upep), respectively.

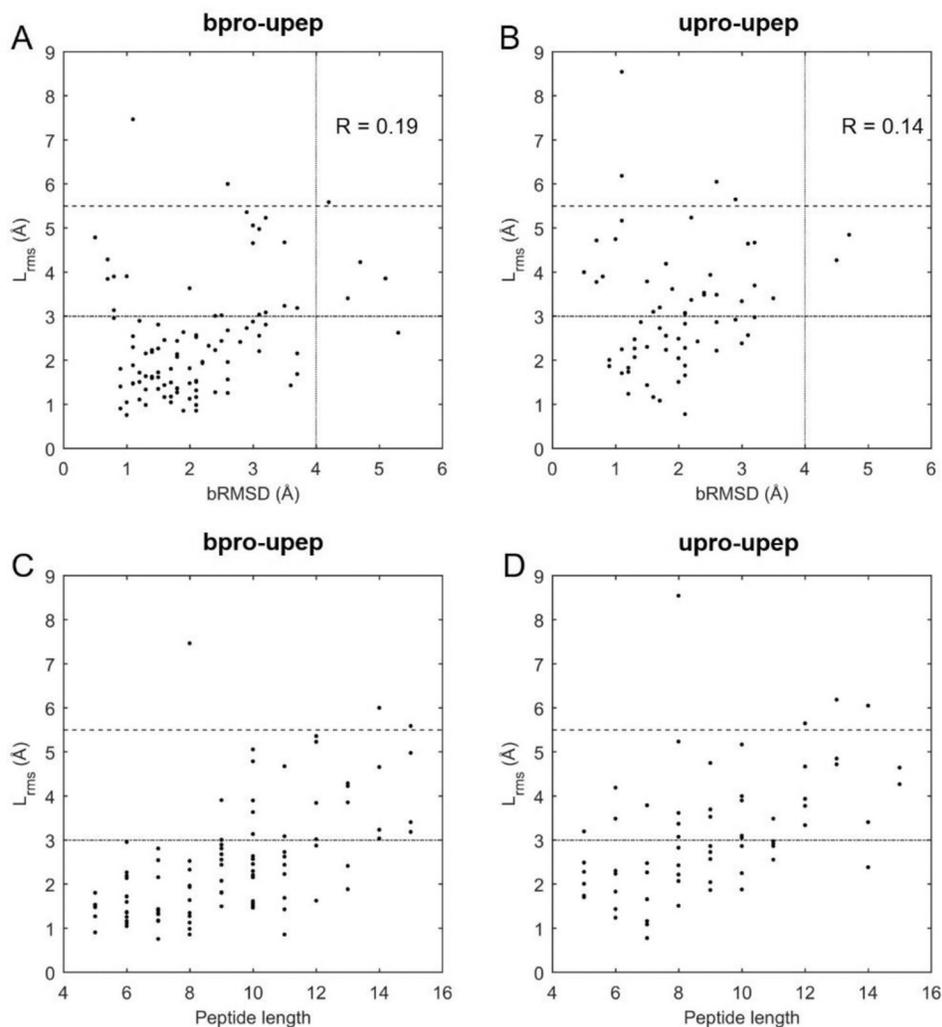


Figure 4.

The dependence of MDockPeP sampling performance on the quality of initial peptide conformers and on the peptide length. The thresholds for the high-quality sampled binding mode ($L_{rms} = 3.0$ Å) and the medium-quality sampled binding mode ($L_{rms} = 5.5$ Å) are shown as the horizontal dashed lines and broken lines, respectively. (A-B) The relationship between L_{rms} of the best sampled binding mode and bRMSD of the best peptide conformer for the bound docking cases (A) and the unbound docking cases (B), respectively. The threshold for effective peptide modeling ($bRMSD = 4.0$ Å) is shown as the vertical dashed line. (C-D) The distribution of L_{rms} of the best sampled binding mode as a function of the peptide length for the bound docking cases (C) and the unbound docking cases (D), respectively.