

**PepBinding: A workflow for predicting peptide binding structures by combining Peptide
Docking and Peptide Gaussian Accelerated Molecular Dynamics Simulations**

Jinan Wang^{1,a}, Kushal Koirala^{1,a,b}, Hung N. Do^{c,d}, Yinglong Miao,^{a,*}

a. Computational Medicine Program and Department of Pharmacology, University
of North Carolina – Chapel Hill, Chapel Hill, North Carolina, USA 27599

b. Curriculum in Bioinformatics and Computational Biology, University of North Carolina at
Chapel Hill, North Carolina 27599

c. Computational Biology Program, Department of Molecular Biosciences, University of Kansas,
Lawrence, KS 66047, USA

d. Current address: Theoretical Biology and Biophysics Group, Theoretical Division, Los
Alamos National Laboratory, Los Alamos, New Mexico, USA 87545

¹ These authors contributed equally to this work

*To whom correspondence should be addressed: Yinglong_Miao@med.unc.edu

Abstract

Predicting protein-peptide interactions is crucial for understanding peptide binding processes and designing peptide drugs. However, traditional computational modeling approaches face challenges in accurately predicting peptide-protein binding structures due to the slow dynamics and high flexibility of peptides. Here, we introduce a new workflow termed “PepBinding” for predicting peptide binding structures, which combines peptide docking, all-atom enhanced sampling simulations using the Peptide Gaussian accelerated Molecular Dynamics (Pep-GaMD) method and structural clustering. PepBinding has been demonstrated on seven distinct model peptides. In peptide docking using HPEPDOCK, the peptide backbone root-mean-square deviations (RMSDs) of their bound conformations relative to X-ray structures ranged from 3.8 Å to 16.0 Å, corresponding to the medium to inaccurate quality models according to the Critical Assessment of PRediction of Interactions (CAPRI) criteria. The Pep-GaMD simulations performed for only 200 ns significantly improved the docking models, resulting in five medium and two acceptable quality models. Therefore, PepBinding is an efficient workflow for predicting peptide binding structures and is publicly available at <https://github.com/MiaoLab20/PepBinding>.

Keywords: Peptide docking, Peptide Gaussian accelerated Molecular Dynamics, PepBinding, Structural clustering, Enhanced sampling

1. Introduction

Protein-peptide interactions play a critical role in various biological processes, including signal transduction, cellular regulation, immune response, protein trafficking pathway, and so on.^{1, 2} Peptides exhibit unique capability to bind target protein, particularly those with shallow pockets, which are difficult to modulate by small molecules. Peptides have thus emerged as promising candidates for new therapeutics. The number of marketed peptide-based drugs keeps increasing in recent years.³⁻⁶ Therefore, accurate characterization of peptide-protein interactions is important for both biological research and drug development.³⁻⁶

Experimental techniques including X-ray crystallography, cryo-electron microscopy (cryo-EM) and nuclear magnetic resonance (NMR) have been widely utilized to determine high-resolution structures of peptide-protein complexes.⁷ These structures are usually deposited into the Protein Data Bank (PDB).⁸ However, due to time-consuming, high cost and technical difficulties, experimental structures have been resolved for only a small fraction of protein-peptide complexes.⁹ In this context, computational modelling has significantly facilitated the study of protein-peptide interactions. By exploring and ranking possible peptide binding conformations, molecular docking has provided key insights into the protein-peptide interactions¹⁰ and is widely used in the peptide drug design.¹¹ The docking methods could be roughly grouped into three categories: template-based, local docking, and global docking. The template-based docking methods such as GalaxyPepDock¹² are highly efficient and accurate if high-quality templates are available¹². While local docking approaches such as MDockPep,¹³ HADDOCK,¹⁴ and Rosetta FlexPepDock¹⁵ need a priori knowledge of the peptide binding site, they are able to generate high quality models according to the Critical Assessment of PRediction of Interactions (CAPRI) criteria.¹⁶ In contrast, there is no need for any pre-defined binding sites for global docking

programs such as HPEPDOCK,¹⁷ CABS-dock,¹⁸ PIPER-FlexPepDock,¹⁹ PatchMAN²⁰ and PeptiDock.²¹ The global docking approaches provide sampling of peptide binding over the entire protein surface. However, it is still challenging to account for the system flexibility. In this regard, HPEPDOCK considers the peptide flexibility through an ensemble of peptide conformations generated by MODPEP program.¹⁷ *HPEPDOCK* is thus able to alleviate the peptide flexibility problem through ensemble docking of the peptides. Recently, AlphaFold2²² has shown great promise for providing structural insight into a wide range of peptide–protein complexes based on deep learning. Nevertheless, the docking calculations or deep learning could only provide snapshots of the protein-peptide interactions. Addressing the high flexibility of peptides and the dynamic interactions between protein and peptide remains a significant challenge. Overall, peptide docking often generates poor predictions that require further refinement to obtain quality models.

Molecular Dynamics (MD) is a powerful technique that enables all-atom simulations of biomolecules. It can fully account for the flexibility of both peptides and proteins during their interactions.^{5, 23-26} Therefore, MD has been used to refine binding poses of peptides to proteins obtained from docking.²⁷⁻³¹ For example, conventional MD (cMD) has been used to refine peptide binding poses in proteins in the pepATTRACT²⁸ and AnchorDock³⁰ docking protocols. Furthermore, cMD simulations have been widely applied to explore peptide binding mechanisms.^{5, 23-26, 32-35} cMD simulations performed for 200 μ s using the Anton specialized supercomputer have captured 70 binding and unbinding events between an intrinsically disordered protein fragment of the measles virus nucleoprotein and the X domain of the measles virus phosphoprotein complex, which enables detailed understanding of the peptide “folding-upon-binding” mechanism.³⁵ Microseconds of high temperature MD simulations with RSFF2C force field captured repetitive peptide binding events, which allowed to accurately predict protein-peptide binding complex.³⁶

Furthermore, the combination of MD and machine learning has significantly enhanced the efficacy and accuracy of predicting the structure of cyclic peptide-protein complexes.³⁷ Despite these remarkable advances, it is still challenging to sufficiently sample peptide-protein interactions through cMD simulations, due to the slow dynamics and limited simulation timescales.

On the other hand, enhanced sampling methods have provided improved sampling of peptide-protein interactions, which could efficiently capture both peptide binding and dissociation processes. The widely used enhanced sampling methods include steered MD,³⁸⁻⁴⁰ umbrella sampling,⁴¹⁻⁴³ metadynamics,⁴⁴⁻⁴⁹ adaptive biasing force,^{50, 51} supervised MD,²⁴ weighted ensemble,³⁴ Modeling by Employing Limited Data (MELD) using temperature and Hamiltonian replica exchange MD,^{52, 53} temperature-accelerated MD,⁵⁴ multi-ensemble Markov state models,⁵⁵ accelerated MD (aMD),^{56, 57} Gaussian accelerated MD (GaMD),⁵⁷⁻⁶⁰ and so on. In particular, the Peptide GaMD (Pep-GaMD)⁶¹ is a more recently developed peptide enhanced sampling method that works by selectively boosting essential interaction potential energy of the peptide. Pep-GaMD has successfully captured repetitive dissociation and binding of model peptides within microsecond simulations. Nevertheless, enhanced sampling simulations of peptide binding to proteins have been under explored. In this work, we present a new computational workflow termed “PepBinding” that combines global peptide docking using *HPEPDOCK* and all-atom enhanced sampling simulations using Pep-GaMD to model protein-peptide interactions. Seven model peptides have been selected from the PeptiDB database of non-redundant peptide-protein complex structures.⁹ Starting with the top model obtained from the *HPEPDOCK*, Pep-GaMD is applied to refine the peptide-protein complex structures.

2. Methods

PepBinding Workflow Combining *HPEPDOCK* and *Pep-GaMD*

The PepBinding workflow has been developed to predict protein-peptide binding structures. The workflow integrated peptide docking using HPEPDOCK¹⁷ with Pep-GaMD⁶¹ enhanced sampling simulations (**Fig. 1**). The HPEPDOCK, a hierarchical algorithm for blind and global docking, addresses peptide flexibility by generating an ensemble of peptide conformation with MODPEP program.⁶² Using MDock⁶³ with rigid docking, the sampled peptide structures are then globally docked against the entire protein. After completion of the peptide docking, HPEPDOCK provides interactive view of top-ranked 10 models. For our study, only the very top pose was selected for further Pep-GaMD refinement.

System Setup

Seven model peptides were selected from the PeptiDB database of non-redundant peptide-protein complex structures⁹. They included peptide “GPPPAMPARPT” (Peptide 1), “HTLKGRRLVFDN” (Peptide 2), “KSLTIYAQVQK” (Peptide 3), “ARTKQT” (Peptide 4), “KSTQATLERWF” (Peptide 5), “NMTPYRSPPPYVP” (Peptide 6), “GASDGSGWSENNPW” (Peptide 7), which binds to SH3 domain, WD-repeat 5, C-terminal regions in FEN-1, DNA polymerase sliding clamp, and dystrophin-glycoprotein complex (DGC) and TolB protein of *E. coli*, respectively. The free structures of target proteins are 1OOT, 1H1R, 1D1Z, 2H14, 1RWZ, 1EG4, and 1CRZ, respectively. The corresponding bound structures of protein-peptide complexes are 1SSH, 2CCH, 1D4T, 2H9M, 1RXZ, 1EG4 and 2IVZ, respectively. The free X-ray PDB structure of target protein and the peptide sequence were used as input for the HPEPDOCK docking. The top ranked of each peptide was further refined using Pep-GaMD simulations.

Pep-GaMD Enhanced Sampling Simulations

For structural refinement of the peptide-protein docked complex, we used a more recently developed enhanced sampling method Pep-GaMD.⁶¹ The complexes were solvated in explicit water using the *tleap* program within AMBER22 package.⁶⁴ The charge in each system was neutralized using Na⁺ and Cl⁻ ions. The AMBER ff14SB⁶⁵ force field parameter was used for the protein/peptide and TIP3P as a water model⁶⁶ for water molecules. Each system was neutralized by adding counter ions and immersed in a cubic TIP3P water box,⁶⁶ which was extended 15 Å from the protein-peptide complex surface. For energy minimization of each system, the steepest descent for 50,000 steps and the conjugate gradient for additional 50,000 steps were performed with and without applying harmonic position restraints of 1 kcal/mol*Å² on heavy atoms of protein and peptides. Minimization was followed by heating the system from 0 to 300 K in constant number, volume, and temperature (NVT) ensemble for 1ns simulation by applying harmonic position restraints of 1 kcal/mol. Å² on heavy atoms of protein and peptides. System was then further equilibrated using constant number, pressure, and temperature (NPT) ensemble at 1 atm and 300K for 1ns with same restraints as in NVT run. Then, the system was fully equilibrated for a short 2 ns cMD without any constraints under the NPT ensemble. In the Pep-GaMD equilibration: a 2ns cMD run was firstly performed to calculate potential statistics (such as maximum, minimum, average and standard deviation) for adding the boost potential; then 18ns Pep-GaMD equilibration was performed with the boost potential added. Finally, three independent production runs of each system were performed for 200 ns with randomized initial atomic velocities. The simulation frames were saved every 0.2 ps for trajectory analysis. In the Pep-GaMD simulations, the threshold energy for applying the first and second boost potential was set to the lower bound. The upper limits of the standard deviation of the second potential boost (σ_0D) were set to 6.0 kcal/mol for all systems. By default, the upper limits of the standard deviation of the first potential boost (σ_0P)

were set to 6.0 kcal/mol for the initial tests. However, these acceleration parameters led to peptide dissociation during Pep-GaMD equilibration in the 1D4T, 2CCH, and 1EG2 systems. Since the PepBinding approach was focused on refining the local peptide binding pose rather than capturing the peptide dissociation and rebinding process, $\sigma_0 P$ was decreased to maintain the peptides in the bound state. Detailed description about each system setup, including the system size, the number of ions for neutralization, Pep-GaMD simulation parameters, and the boost potential values were summarized in **Tables S1 & S2**. All snapshots from the three production runs were used to analysis with CPPTRAJ.⁶⁷ Root-mean square deviations (RMSDs) were calculated for the peptide backbone relative to the X-ray structure, with the protein backbone aligned. RMSDs of the protein backbones relative to the X-ray structures with the protein aligned, distances between the center-of-mass of protein and center-of-mass of peptide, the number of native contact between peptide and protein were calculated using CPPTRAJ.⁶⁷ The PyReweighting toolkit⁶⁸ was utilized to reweight the structural clusters with three Pep-GaMD simulations combined. The peptide RMSD with protein aligned and protein RMSD were chosen as reaction coordinate for calculating the 2D PMF profile. The bin size was set to 1.0 Å. The cutoff for the number of simulation frames in one bin was set to 500 for reweighting 2D PMF profiles. To obtain the peptide binding pose, structural clustering was performed using the agglomerative algorithm implemented in CPPTRAJ.⁶⁷ Structural alignment was first performed on the protein backbone. Clustering was then performed using the peptide backbone RMSDs to obtain 10 representative structures. The structures from the PepBinding predictions shown in **Figs. 2 and S3** are those with the lowest peptide backbone RMSD among the 10 representative structures.

3. Results

Prediction of Peptide Binding Structures through Docking and Pep-GaMD Simulations

Peptide docking using *HPEPDOCK* exhibited different levels of accuracy, as evidenced by RMSDs of the peptide backbone when compared to the bound X-ray structures. The peptide RMSD values for the seven model peptides were 3.87 Å, 4.50 Å, 5.49 Å, 7.80 Å, 8.50 Å, 14.12 Å, and 16.40 Å, respectively (**Fig. 2** and **Table 1**). According to the CAPRI criteria,¹⁶ the first two peptides reached predictions of the medium quality, while the third to fifth peptides fell within the acceptable predictions. The last two peptides, predictions were classified as incorrect.

Pep-GaMD simulations were conducted with three replicas, each lasting 200 ns, to refine the docking models. For Peptides 1 and 2, RMSDs of the peptide backbone relative to the X-ray structures decreased to less than 3.0 Å during the Pep-GaMD simulations (**Figs. 2A-2B & 3A-3B**). Peptides 1-3 exhibited tight binding to the protein target site throughout the three Pep-GaMD simulations, with slight fluctuations observed in Sim3 of the Peptide 1 (**Figs. 3A-3C**). Two Pep-GaMD simulations of Peptide 4 resulted in a significant reduction in the peptide RMSD (**Figs. 3D**). In comparison, Peptides 5-6 displayed higher fluctuations and transiently sampled acceptable conformations with RMSD <10 Å during the Pep-GaMD simulations (**Figs. 3E & 3F**). Peptide 7 appeared trapped in a metastable state in the 200 ns Pep-GaMD simulations and RMSD of the peptide backbone dropped below 10 Å (**Fig. 3G**). RMSDs of the seven model peptides showed notable decreases compared with the initial docking models (**Figs. 2-3 and Table 1**).

Moreover, Pep-GaMD simulation snapshots of the peptide conformations were clustered based on backbone RMSDs relative to the X-ray structures (See methods). The best models in the resulting clustering was found to be 2.11 Å, 2.79 Å, 4.12 Å, 1.36 Å, 9.87 Å, 4.80 Å, and 9.64 Å, for the seven peptides (**Table 1**). According to the CAPRI criteria¹⁶, PepBinding predictions of Peptides 1-4 and 6 were determined of medium quality, while those of Peptides 5 and 7 were

considered acceptable quality. Therefore, Pep-GaMD simulations significantly improved the docking prediction quality of the seven peptide binding structures. The predicted bound conformations of the peptides were mostly similar to experimental X-ray structures, showcasing the peptide backbone RMSDs ranging from 2.1 Å to 9.8 Å. In contrast, docking poses derived from *HPEPDOCK* of the seven peptides exhibited RMSDs spanning from 3.87 Å to 16.40 Å (**Table 1**).

2D potential of mean forces (PMFs) were further calculated from the Pep-GaMD simulations, utilizing protein and peptide backbone RMSDs relative to the bound X-ray structures as reaction coordinates (**Fig. 4**). For peptides 1 to 4, only a single low-energy minimum was identified near the native bound structure (**Figs. 4A-4D**). Peptides 5 to 7 also exhibited a single low-energy minimum, but with moderate deviation and higher RMSDs from the native bound structure (**Fig. 4E-4G**). Specifically, Peptide 1 sampled a low-energy well, centered at ~2.1 Å and ~0.8 Å backbone RMSDs for the peptide and protein relative to the bound X-ray structure, being consistent with structural clustering findings of the peptide. For Peptide 2, Pep-GaMD explored a broad low-energy well, centered at ~2.8 Å and ~1.2 Å backbone RMSDs for the peptide and protein relative to the bound X-ray structure. Similarly, Peptide 3 exhibited a low-energy well, centered at around 4.1 Å and 0.8 Å backbone RMSDs for the peptide and protein relative to the bound X-ray structure. Peptide 4, featured a low-energy well, centered at approximately 1.4 Å and 0.8 Å backbone RMSDs for the peptide and protein relative to the bound X-ray structure. In comparison, Peptides 5 to 7 displayed low-energy wells with centers at higher ~9.8 Å and ~1.1 Å, ~4.9 Å and ~1.0 Å, and ~9.5 Å and ~1.0 Å backbone RMSDs for the peptide and protein relative to the bound X-ray structures, respectively. To further characterize the Pep-GaMD refinement, we calculated the time course of the number of native contacts between the protein and peptide, and

their center-of-mass distance (**Fig. S2**). The results indicated that the current Pep-GaMD simulations did not sample dissociation of the peptides. During the simulations, the number of peptide native contacts increased and RMSD of the peptide relative to the X-ray structure mostly decreased. Therefore, the Pep-GaMD simulations primarily refined local interactions between the peptide and protein rather than capturing the complete peptide dissociation and binding events. These findings provide important insights into the energetically favorable conformations sampled by Pep-GaMD simulations of each peptide.

Effects of Terminal Residue Charges on Peptide Binding

In addition to the zwitterion terminus model as described above, we conducted further Pep-GaMD simulations of the seven peptides with neutral terminal patched residues. Similar results were obtained for most of the peptides, as depicted in **Figs S3-S5**. For Peptides 1-7 in the neutral terminal model, the top-ranked PepBinding predictions exhibited peptide backbone RMSDs of 1.28 Å, 5.34 Å, 4.42 Å, 0.69 Å, 5.82 Å, 9.06 Å, and 8.98 Å relative to the X-ray bound structures (**Table 2** and **Fig. S3**). These prediction yielded one high-quality, two medium-quality, and four acceptable-quality models. Significant improvements were observed for peptides 1, 4, and 5 with the neutral termini, for which the backbone RMSDs decreased from 2.11 Å to 1.28 Å, 1.36 Å to 0.69 Å, and 9.87 Å to 5.82 Å, respectively. On the other hand, Peptides 2 and 6 exhibited larger RMSDs with the neutral termini. Their RMSDs increased from 2.79 Å to 5.34 Å, and 4.8 Å to 9.06 Å, respectively. Overall, the simulation-predicted peptide bound conformations with neutral termini exhibited slightly smaller peptide RMSDs from the native X-ray structures when compared to the zwitterion terminus models. It is worth noting that the experimental structures are obtained from the zwitterion terminus models. Therefore, it is better to adjust terminal residue charges in

predictions of peptide binding structures to match the experimental condition since no significant differences were observed between these two models in our simulations.

4. Discussion

We have developed and demonstrated the PepBinding workflow to predict peptide binding structures. Peptides typically bind to shallow protein surface with large pockets, while small-molecule ligands can access deeply buried sites. Peptide-protein interactions tend to be weaker than protein-protein interactions, primarily due to the smaller interface between peptides and their target proteins. Most peptides lack stable structures, making it challenging to incorporate their high flexibility and large conformational changes (folding and unfolding) into computational modeling, particularly in peptide docking scenarios. Utilizing seven peptides with diverse lengths and difficulty levels as model systems, we observed significant improvement of the peptide binding structure predictions. The Pep-GaMD refinement yielded models of medium quality for Peptides 1-5 and acceptable quality for Peptides 6 and 7 with the zwitterion terminus models. This promising outcome suggests wide applicability of PepBinding to many other peptide-protein binding systems. It is worth noting that our current PepBinding approach utilizes only the top binding pose from HPEPDock.

Next, we examined whether different HPEPDOCK-predicted structures of the peptides could be effectively refined using the PepBinding approach. Peptides 1 and 2 were selected as test systems using the second and third binding poses obtained from HPEPDock for additional Pep-GaMD simulations. As shown in **Figs. S6 & S7**, docking pose 3 of both peptides showed similar quality as pose 1 with correct peptide binding orientation. However, docking pose 2 of both peptides appeared in a reversed binding orientation of the peptide, leading to significant high peptide RMSD

values of 15.8 Å and 18.3 Å for peptide 1 and 2, respectively. The same parameters and approaches were applied in the additional Pep-GaMD simulations. For peptide 1, Pep-GaMD improved the model of binding pose 3, reducing the peptide RMSD from 5.5 Å to 2.3 Å (**Fig S6**). Even with the reversed peptide direction in pose 2, Pep-GaMD could improve the prediction quality from inaccurate to mediate (**Fig S6**). In the more challenging case of peptide 2, Pep-GaMD improved the model of docking pose 3, reducing the peptide RMSD from 4.8 Å to 3.2 Å (**Fig S7**). For pose 2 with the reversed peptide direction, although Pep-GaMD reduced the peptide RMSD from 18.3 Å to 15.2 Å, the refined model remained inaccurate. These tests suggested that including more binding poses for refinement in PepBinding may increase the success rate of obtaining improved quality models of peptide binding structures. In this regard, combining Pep-GaMD with replica exchange by incorporating different binding poses into different replicas might enhance the efficiency of PepBinding binding predictions. This will be subject to future studies.

We compared PepBinding prediction accuracies with *AutoDock CrankPep (ADCP)*⁶⁹ on our dataset. ADCP is a recent developed local docking method that requires information about the binding site and has shown excellent prediction performance on the PepSet dataset among eight local docking methods. However, the ADCP only obtained one mediate quality, 5 acceptable quality models and 1 inaccurate predictions at the top-rank model (**Table S3**). This result underscores the challenging nature of our chosen peptides for different peptide docking software. Since peptide docking often yielded similar quality models, Pep-GaMD is promising to refine the peptide docking models, being consistent with our previous findings of GaMD combination with *PeptiDock*.³¹ However, further testing of more peptide systems is needed for Pep-GaMD in peptide refinement and PepBinding predictions, and its potential generalization to different docking programs require additional exploration. In addition to conformational sampling, the reliability of

MD simulations relies on the accuracy of the force field. In a study by Chen et al. ,³⁶ various force fields, including RSFF2C and AMBER ff14SB, were compared. While high-temperature MD simulations with both force fields produced near-native conformations, RSFF2C showed superior performance. The impact of different force fields on the overall performance remains a subject for future investigations. In comparison to GaMD and cMD approaches, our current Pep-GaMD method showcases more efficient sampling, requiring shorter simulation lengths for peptide structural refinement. For peptides 1 (1SSH), 3 (1D4T) and 6 (1EG4), we compared Pep-GaMD and cMD simulations of the same length (**Fig. S8 & Table S4**). For peptides 1 and 3, cMD improved the binding pose quality to the medium level. However, for the challenging peptide 6, no improvement was observed with cMD, while Pep-GaMD could improve the model quality from level of incorrect to medium with the enhanced sampling. While our previous GaMD refinement required four 300 ns simulations of rather short peptides³¹. Pep-GaMD achieves similar refinement with only three 200 ns simulations on even longer peptides. Notably, the three replicas of Pep-GaMD simulations were found to converge well, as PMFs calculated from the combined and individual simulation trajectories were similar (**Figs. S9 & S10**). In particular, the lowest-energy states identified from the combined PMFs were close to those from PMFs of the individual simulations, despite slight differences in the peptide configuration space sampled in the 1SSH and 2H9M systems. Nevertheless, these brief Pep-GaMD simulations did not sample any dissociation events, preventing convergence on peptide binding. This indicates that enhancing the current PepBinding approach should primarily focus on refining the local interactions between the peptide and protein. Longer simulations with higher acceleration of Pep-GaMD are needed for capturing dissociation and rebinding processes, which will be subject to future studies. High-performance Pep-GaMD simulations conducted using AMBER 22 on NVIDIA L40 GPU cards further affirm

computational viability, with each 200 ns GaMD simulations of Peptides 1 to 7 requiring less than two days. This enhanced efficiency positions PepBinding as a highly promising approach for wider applications in peptide binding structure predictions.

5. Conclusion

In summary, PepBinding has been demonstrated on predicting the peptide-protein complex structures, using seven distinct peptides as model systems. However, Pep-GaMD simulations of different lengths, the effects of different force fields (e.g. RSFF2C, CHARMM36m) and solvent models (e.g., TIP4P, implicit solvent, etc.)⁷⁰ and various structural clustering algorithms are to be further investigated in the future. For refinement of the docking poses with Pep-GaMD, because AMBER22 was applied for running the simulations, the widely used AMBER ff14SB force field was selected. Development of novel protocols to increase the accuracy of peptide-protein structural prediction will facilitate peptide drug design. Advances in computational methods and computing power are expected to help us to address these challenges.

6. Author Information

Corresponding Author

Yinglong Miao- Computational Medicine Program and Department of Pharmacology, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina, USA 27599

Authors

Jinan Wang- Computational Medicine Program and Department of Pharmacology, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina, USA 27599

Kushal, Koirala- Computational Medicine Program and Department of Pharmacology, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina, USA 27599

Hung N. Do- Computational Biology Program, Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66047, USA; Current address: Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos, National Laboratory, Los Alamos, New Mexico, USA 87545

Notes

The authors declare no competing financial interest.

7. Supporting Information

Tables S1-S4, Figures S1-S10 are provided in the Supporting Information. This information is available free of charge via the Internet at <http://pubs.acs.org>.

8. Acknowledgements

This work used supercomputing resources with allocation award TG-MCB180049 through the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296, and project M2874 through the National Energy Research Scientific Computing Center (NERSC), which is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231, and Research Computing at the University of North Carolina – Chapel Hill. This work was also supported by National Science Foundation (2121063) and the startup package at the University of North Carolina – Chapel Hill.

9. References

- (1) Petsalaki, E.; Russell, R. B., Peptide-mediated interactions in biological systems: new discoveries and applications. *Curr. Opin. Biotechnol.* **2008**, *19* (4), 344-350.
- (2) Das, A. A.; Sharma, O. P.; Kumar, M. S.; Krishna, R.; Mathur, P. P., PepBind: a comprehensive database and computational tool for analysis of protein-peptide interactions. *Genomics, Proteomics Bioinf.* **2013**, *11* (4), 241-246.
- (3) Ahrens, V. M.; Bellmann-Sickert, K.; Beck-Sickinger, A. G., Peptides and peptide conjugates: therapeutics on the upward path. *Future Med. Chem.* **2012**, *4* (12), 1567-1586.
- (4) Fosgerau, K.; Hoffmann, T., Peptide therapeutics: current status and future directions. *Drug Discov. Today* **2015**, *20* (1), 122-128.
- (5) Kahler, U.; Fuchs, J. E.; Goettig, P.; Liedl, K. R., An unexpected switch in peptide binding mode: from simulation to substrate specificity. *J. Biomol. Struct. Dyn.* **2018**, *36* (15), 4072-4084.
- (6) Lee, A. C.; Harris, J. L.; Khanna, K. K.; Hong, J. H., A Comprehensive Review on Current Advances in Peptide Drug Development and Design. *Int. J. Mol. Sci.* **2019**, *20* (10), 2383.
- (7) Schirò, A.; Carlon, A.; Parigi, G.; Murshudov, G.; Calderone, V.; Ravera, E.; Luchinat, C., On the complementarity of X-ray and NMR data. *J. Struct. Bio.* **2020**, *4*, 100019.
- (8) Burley, S. K.; Berman, H. M.; Kleywegt, G. J.; Markley, J. L.; Nakamura, H.; Velankar, S., Protein Data Bank (PDB): the single global macromolecular structure archive. *Methods. Mol. Biol.* **2017**, 627-641.
- (9) London, N.; Movshovitz-Attias, D.; Schueler-Furman, O., The structural basis of peptide-protein binding strategies. *Structure* **2010**, *18* (2), 188-199.
- (10) Lei, Y.; Li, S.; Liu, Z.; Wan, F.; Tian, T.; Li, S.; Zhao, D.; Zeng, J., A deep-learning framework for multi-level peptide-protein interaction prediction. *Nat. Commun.* **2021**, *12* (1), 5465.
- (11) Ciemny, M.; Kurcinski, M.; Kamel, K.; Kolinski, A.; Alam, N.; Schueler-Furman, O.; Kmiecik, S., Protein-peptide docking: opportunities and challenges. *Drug Discov. Today* **2018**, *23* (8), 1530-1537.
- (12) Lee, H.; Heo, L.; Lee, M. S.; Seok, C., GalaxyPepDock: a protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Res.* **2015**, *43* (W1), W431-W435.
- (13) Xu, X.; Yan, C.; Zou, X., MDockPeP: An ab-initio protein-peptide docking server. *J. Comput. Chem.* **2018**, *39* (28), 2409-2413.
- (14) Trellet, M.; Melquiond, A. S.; Bonvin, A. M., A unified conformational selection and induced fit approach to protein-peptide docking. *Plos One* **2013**, *8* (3), e58769.
- (15) Raveh, B.; London, N.; Zimmerman, L.; Schueler-Furman, O., Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors. *Plos One* **2011**, *6* (4), e18934.
- (16) Janin, J.; Henrick, K.; Moult, J.; Eyck, L. T.; Sternberg, M. J.; Vajda, S.; Vakser, I.; Wodak, S. J.; Critical Assessment of, P. I., CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* **2003**, *52* (1), 2-9.
- (17) Zhou, P.; Jin, B.; Li, H.; Huang, S.-Y., HPEPDOCK: a web server for blind peptide-protein docking based on a hierarchical algorithm. *Nucleic Acids Res.* **2018**, *46* (W1), W443-W450.
- (18) Kurcinski, M.; Jamroz, M.; Blaszczyk, M.; Kolinski, A.; Kmiecik, S., CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Res.* **2015**, *43* (W1), W419-W424.

- (19) Alam, N.; Goldstein, O.; Xia, B.; Porter, K. A.; Kozakov, D.; Schueler-Furman, O., High-resolution global peptide-protein docking using fragments-based PIPER-FlexPepDock. *PLoS Comput. Biol.* **2017**, *13* (12), e1005905.
- (20) Khramushin, A.; Ben-Aharon, Z.; Tsaban, T.; Varga, J. K.; Avraham, O.; Schueler-Furman, O., Matching protein surface structural patches for high-resolution blind peptide docking. *Proc. Natl. Acad. Sci. U.S.A* **2022**, *119* (18), e2121153119.
- (21) Porter, K. A.; Xia, B.; Beglov, D.; Bohnuud, T.; Alam, N.; Schueler-Furman, O.; Kozakov, D., ClusPro PeptiDock: efficient global docking of peptide recognition motifs using FFT. *Bioinformatics* **2017**, *33* (20), 3299-3301.
- (22) Tsaban, T.; Varga, J. K.; Avraham, O.; Ben-Aharon, Z.; Khramushin, A.; Schueler-Furman, O., Harnessing protein folding neural networks for peptide-protein docking. *Nat. Commun.* **2022**, *13* (1), 176.
- (23) Knapp, B.; Demharter, S.; Esmailbeiki, R.; Deane, C. M., Current status and future challenges in T-cell receptor/peptide/MHC molecular dynamics simulations. *Brief Bioinform.* **2015**, *16* (6), 1035-1044.
- (24) Salmaso, V.; Sturlese, M.; Cuzzolin, A.; Moro, S., Exploring Protein-Peptide Recognition Pathways Using a Supervised Molecular Dynamics Approach. *Structure* **2017**, *25* (4), 655-662 e2.
- (25) Yadahalli, S.; Li, J.; Lane, D. P.; Gosavi, S.; Verma, C. S., Characterizing the conformational landscape of MDM2-binding p53 peptides using Molecular Dynamics simulations. *Sci Rep-Uk* **2017**, *7* (1), 15600.
- (26) Wan, S.; Knapp, B.; Wright, D. W.; Deane, C. M.; Coveney, P. V., Rapid, Precise, and Reproducible Prediction of Peptide-MHC Binding Affinities from Molecular Dynamics That Correlate Well with Experiment. *J. Chem. Theory Comput.* **2015**, *11* (7), 3346-3356.
- (27) Antes, I., DynaDock: A new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility. *Proteins* **2010**, *78* (5), 1084-1104.
- (28) de Vries, S. J.; Rey, J.; Schindler, C. E. M.; Zacharias, M.; Tuffery, P., The pepATTRACT web server for blind, large-scale peptide-protein docking. *Nucleic Acids Res.* **2017**, *45* (W1), W361-W364.
- (29) Schindler, C. E.; de Vries, S. J.; Zacharias, M., Fully blind peptide-protein docking with pepATTRACT. *Structure* **2015**, *23* (8), 1507-1515.
- (30) Ben-Shimon, A.; Niv, M. Y., AnchorDock: Blind and Flexible Anchor-Driven Peptide Docking. *Structure* **2015**, *23* (5), 929-940.
- (31) Wang, J.; Alekseenko, A.; Kozakov, D.; Miao, Y., Improved Modeling of Peptide-Protein Binding Through Global Docking and Accelerated Molecular Dynamics Simulations. *Front. Mol. Biosci.* **2019**, *6* (112), 112.
- (32) Ahmad, M.; Gu, W.; Helms, V., Mechanism of fast peptide recognition by SH3 domains. *Angew. Chem.-Int. Edit.* **2008**, *47* (40), 7626-7630.
- (33) Chen, J.; Wang, J.; Zhang, Q.; Chen, K.; Zhu, W., Probing origin of binding difference of inhibitors to MDM2 and MDMX by polarizable molecular dynamics simulation and QM/MM-GBSA calculation. *Sci Rep-Uk* **2015**, *5* (1), 17421.
- (34) Zwier, M. C.; Pratt, A. J.; Adelman, J. L.; Kaus, J. W.; Zuckerman, D. M.; Chong, L. T., Efficient Atomistic Simulation of Pathways and Calculation of Rate Constants for a Protein-Peptide Binding Process: Application to the MDM2 Protein and an Intrinsically Disordered p53 Peptide. *J. Phys. Chem. Lett.* **2016**, *7* (17), 3440-3345.
- (35) Robustelli, P.; Piana, S.; Shaw, D. E., Mechanism of Coupled Folding-upon-Binding of an Intrinsically Disordered Protein. *J. Am. Chem. Soc.* **2020**, *142* (25), 11092-11101.

- (36) Chen, J.-N.; Jiang, F.; Wu, Y.-D., Accurate Prediction for Protein–Peptide Binding Based on High-Temperature Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2022**, *18* (10), 6386-6395.
- (37) Miao, J.; Descoteaux, M. L.; Lin, Y.-S., Structure prediction of cyclic peptides by molecular dynamics + machine learning. *Chem. Sci.* **2021**, *12* (44), 14927-14936.
- (38) Cuendet, M. A.; Michielin, O., Protein-protein interaction investigated by steered molecular dynamics: the TCR-pMHC complex. *Biophys. J.* **2008**, *95* (8), 3575-3590.
- (39) Gonzalez, A.; Perez-Acle, T.; Pardo, L.; Deupi, X., Molecular basis of ligand dissociation in beta-adrenergic receptors. *Plos One* **2011**, *6* (9), e23815.
- (40) Cuendet, M. A.; Zoete, V.; Michielin, O., How T cell receptors interact with peptide-MHCs: a multiple steered molecular dynamics study. *Proteins* **2011**, *79* (11), 3007-3024.
- (41) Kastner, J., Umbrella sampling. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1* (6), 932-942.
- (42) Torrie, G. M.; Valleau, J. P., Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23* (2), 187-199.
- (43) Rose, A. S.; Elgeti, M.; Zachariae, U.; Grubmuller, H.; Hofmann, K. P.; Scheerer, P.; Hildebrand, P. W., Position of transmembrane helix 6 determines receptor G protein coupling specificity. *J. Am. Chem. Soc.* **2014**, *136* (32), 11244-11247.
- (44) Laio, A.; Parrinello, M., Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A* **2002**, *99* (20), 12562-12566.
- (45) Alessandro, L.; Francesco, L. G., Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **2008**, *71* (12), 126601.
- (46) Saleh, N.; Saladino, G.; Gervasio, F. L.; Clark, T., Investigating allosteric effects on the functional dynamics of beta2-adrenergic ternary complexes with enhanced-sampling simulations. *Chem. Sci.* **2017**, *8* (5), 4019-4026.
- (47) Saleh, N.; Ibrahim, P.; Saladino, G.; Gervasio, F. L.; Clark, T., An Efficient Metadynamics-Based Protocol To Model the Binding Affinity and the Transition State Ensemble of G-Protein-Coupled Receptor Ligands. *J. Chem. Inf. Model.* **2017**, *57* (5), 1210-1217.
- (48) Saleh, N.; Ibrahim, P.; Clark, T., Differences between G-Protein-Stabilized Agonist-GPCR Complexes and their Nanobody-Stabilized Equivalents. *Angew. Chem.-Int. Edit.* **2017**, *56* (31), 9008-9012.
- (49) Zou, R.; Zhou, Y.; Wang, Y.; Kuang, G.; Agren, H.; Wu, J.; Tu, Y., Free Energy Profile and Kinetics of Coupled Folding and Binding of the Intrinsically Disordered Protein p53 with MDM2. *J. Chem. Inf. Model* **2020**, *60* (3), 1551-1558.
- (50) Darve, E.; Rodriguez-Gomez, D.; Pohorille, A., Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **2008**, *128* (14), 144120.
- (51) Darve, E.; Pohorille, A., Calculating free energies using average force. *J. Chem. Phys.* **2001**, *115* (20), 9169-9183.
- (52) Morrone, J. A.; Perez, A.; MacCallum, J.; Dill, K. A., Computed Binding of Peptides to Proteins with MELD-Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2017**, *13* (2), 870-876.
- (53) Morrone, J. A.; Perez, A.; Deng, Q.; Ha, S. N.; Holloway, M. K.; Sawyer, T. K.; Sherborne, B. S.; Brown, F. K.; Dill, K. A., Molecular Simulations Identify Binding Poses and Approximate Affinities of Stapled alpha-Helical Peptides to MDM2 and MDMX. *J. Chem. Theory Comput.* **2017**, *13* (2), 863-869.

- (54) Lamothe, G.; Malliavin, T. E., re-TAMD: exploring interactions between H3 peptide and YEATS domain using enhanced sampling. *Bmc Struct Biol* **2018**, *18* (1), 4.
- (55) Paul, F.; Wehmeyer, C.; Abualrous, E. T.; Wu, H.; Crabtree, M. D.; Schöneberg, J.; Clarke, J.; Freund, C.; Weikl, T. R.; Noé, F., Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations. *Nat. Commun.* **2017**, *8* (1), 1095.
- (56) Hamelberg, D.; Mongan, J.; McCammon, J. A., Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919-11929.
- (57) Miao, Y.; Feher, V. A.; McCammon, J. A., Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11* (8), 3584-3595.
- (58) Pang, Y. T.; Miao, Y.; Wang, Y.; McCammon, J. A., Gaussian Accelerated Molecular Dynamics in NAMD. *J. Chem. Theory Comput.* **2017**, *13* (1), 9-19.
- (59) Miao, Y.; McCammon, J. A., Chapter Six - Gaussian Accelerated Molecular Dynamics: Theory, Implementation, and Applications. In *Annu. Rep. Comput. Chem.*, Dixon, D. A., Ed. Elsevier: 2017; Vol. 13, pp 231-278.
- (60) Miao, Y.; McCammon, J. A., Mechanism of the G-protein mimetic nanobody binding to a muscarinic G-protein-coupled receptor. *Proc. Natl. Acad. Sci. U.S.A* **2018**, *115* (12), 3036-3041.
- (61) Wang, J.; Miao, Y., Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. *J. Comput. Phys.* **2020**, *153* (15).
- (62) Yan, Y.; Zhang, D.; Huang, S.-Y., Efficient conformational ensemble generation of protein-bound peptides. *J. Cheminf.* **2017**, *9*, 1-13.
- (63) Huang, S. Y.; Zou, X., Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* **2007**, *66* (2), 399-421.
- (64) Case, D. A.; D.S. Cerutti; T.E. Cheatham, I.; T.A. Darden, R. E. D., T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, AMBER 2022, University of California, San Francisco. **2022**.
- (65) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696-3713.
- (66) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926-935.
- (67) Roe, D. R.; Cheatham, T. E., 3rd, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084-3095.
- (68) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A., Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **2014**, *10* (7), 2677-2689.
- (69) Zhang, Y.; Sanner, M. F., Docking Flexible Cyclic Peptides with AutoDock CrankPep. *J. Chem. Theory Comput.* **2019**, *15* (10), 5161-5168.

(70) Kuzmanic, A.; Pritchard, R. B.; Hansen, D. F.; Gervasio, F. L., Importance of the Force Field Choice in Capturing Functionally Relevant Dynamics in the von Willebrand Factor. *J. Phys. Chem. Lett.* **2019**, *10* (8), 1928-1934.

Table 1. Summary of the seven model protein-peptide complexes and performance of the *HPEPDOCK* and PepBinding predictions. *The model quality was assessed according to the CAPRI criteria and denoted in parentheses.

Peptide	Bound	Free	Peptide	RMSD HPEPDOCK (Å)*	RMSD PepBinding (Å)*
1	1SSH(A)	1OOT(A)	GPPPAMPARPT	3.87(Medium)	2.11(Medium)
2	2CCH(B)	1H1R(B)	HTLKGRRLVFDN	4.50(Medium)	2.79(Medium)
3	1D4T(A)	1D1Z(A)	KSLTIYAQVQK	5.49(Acceptable)	4.12(Medium)
4	2H9M(A)	2H14(A)	ARTKQT	7.80(Acceptable)	1.36(Medium)
5	1RXZ(A)	1RWZ(A)	KSTQATLERWF	8.50(Acceptable)	9.87(Acceptable)
6	1EG4(A)	1EG3(A)	NMTPYRSPPPYVP	14.12(Incorrect)	4.80(Medium)
7	2IVZ(A)	1CRZ	GASDGSGWSENPNW	16.40(Incorrect)	9.64(Acceptable)

Table 2. Comparison of prediction models with different peptide terminus settings using the PepBinding workflow. *The model's quality was assessed according to CAPRI peptide docking criteria, and its classification was put inside the brackets.

Peptide		1	2	3	4	5	6	7
RMSD (Å)	PepBinding (Neutral terminus)	1.28 (Medium)	5.34 (Acceptable)	4.42 (Medium)	0.69 (High)	5.82 (Acceptable)	9.06 (Acceptable)	8.98 (Acceptable)
	PepBinding (Zwitterion terminus)	2.11 (Medium)	2.79 (Medium)	4.12 (Medium)	1.36 (Medium)	9.87 (Acceptable)	4.80 (Medium)	9.64 (Acceptable)

Figure captions

Figure 1. Workflow of the PepBinding that combines peptide docking with HPEDOCK, all-atom enhanced sampling simulations with Pep-GaMD and structural clustering.

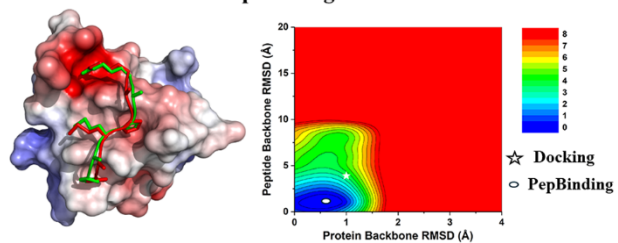
Figure 2. Binding poses of seven peptides predicted using HPEPDOCK (purple) and “PepBinding” (green) with the zwitterion terminus are compared with X-ray structures (blue): (A) peptide “GPPPAMPARPT” (Peptide 1), (B) “HTLKGRRLVFDN” (Peptide 2), (C) “KSLTIYAQVQK” (Peptide 3), (D) “ARTKQT” (Peptide 4), (E) “KSTQATLERWF” (Peptide 5), (F) “NMTPYRSPPPYVP” (Peptide 6), and (G) “GASDGSGWSENNPW” (Peptide 7). The binding poses from the PepBinding predictions were the representative structures with lowest peptide backbone RMSD obtained through clustering analysis of Pep-GaMD simulations. The top 1 binding poses from the HPEPDock, which were also used as initial Pep-GaMD simulations, were shown.

Figure 3. Time courses of peptide backbone RMSD obtained from three 200ns Pep-GaMD simulations on peptides with the zwitterion terminus for (A) peptide “GPPPAMPARPT” (Peptide 1), (B) “HTLKGRRLVFDN” (Peptide 2), (C) “KSLTIYAQVQK” (Peptide 3), (D) “ARTKQT” (Peptide 4), (E) “KSTQATLERWF” (Peptide 5), (F) “NMTPYRSPPPYVP” (Peptide 6), and (G) “GASDGSGWSENNPW” (Peptide 7).

Figure 4. 2D potential of mean force (PMF) regarding the peptide backbone RMSD and protein backbone RMSD for peptides with the zwitterion terminus: (A) Peptide 1, (B) Peptide 2, (C) Peptide 3, (D) Peptide 3, (E) Peptide 3, (F) Peptide 3 and (G) Peptide 7. The white asterisks indicate the initial docking poses obtained using HPEPDOCK.

Table of Contents

PepBinding



Supporting Information: PepBinding: A workflow for predicting peptide binding structures by combining Peptide Docking and Peptide Gaussian Accelerated Molecular Dynamics Simulations

Jinan Wang^{1,a}, Kushal Koirala^{1,a,b}, Hung N. Do^{c,d}, Yinglong Miao,^{a,*}

a. Computational Medicine Program and Department of Pharmacology, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina, USA 27599

b. Curriculum in Bioinformatics and Computational Biology, University of North Carolina at chapel Hill, North Carolina 27599

c. Computational Biology Program, Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66047, USA

d. Current address: Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA 87545

¹ These authors contributed equally to this work

*To whom correspondence should be addressed: Yinglong_Miao@med.unc.edu

Table S1. Summary of Pep-GaMD Simulations Performed on Refining HPEPDock models with zwitterion peptide terminus^a

System	Natoms	System Size (axbxc in Å)	Nions (Na+/Cl-)	σ_{OP}	σ_{OD}	$\Delta V(\text{kcal/mol})$
1SSH	23,240	67.1x66.0x66.2	0/0	6.0	6.0	12.89±3.66
2CCH	55,799	96.4x83.5x83.2	2/0	1.0	6.0	9.73±3.067
1D4T	28,470	77.0x70.8x65.6	0/4	1.5	6.0	11.24±3.14
2H9M	48,937	90.3x88.4x74.1	0/7	6.0	6.0	11.81±3.540
1RXZ	47,063	88.4x96.7x98.8	12/0	6.0	6.0	58.19±6.86
1EG4	59,071	102.9x85.4x80.7	0/1	0.5	6.0	7.99±2.81
2IVZ	71,746	112.2x91.5x83.8	3/0	6.0	6.0	13.37±3.71

^a Natoms is the number of atoms in the simulation; N_{ions} (Na+/Cl-) is the number of ions (Na+ or Cl-) used to neutralize system; σ_{OP} and σ_{OD} are the peptide essential and the second boost potential standard deviations in Pep-GaMD simulations; ΔV is the total boost potential.

Table S2. Summary of Pep-GaMD Simulations Performed on Refining HPEPDock models with neutral peptide terminus^a

System	Natoms	System Size (axbxc in Å)	Nions (Na+/Cl-)	σ_{OP}	σ_{OD}	$\Delta V(\text{kcal/mol})$
1SSH	23,397	70.4x69.7x63.4	0/0	6.0	6.0	12.56±3.55
2CCH	55,799	96.5x83.5x83.2	2/0	6.0	6.0	12.51±3.60
1D4T	28,470	77.0x70.8x65.6	0/4	1.5	6.0	10.96±3.40
2H9M	49,558	90.3x88.4x75.0	0/7	6.0	6.0	13.80±3.76
1RXZ	47,222	96.4x79.6x74.1	12/0	6.0	6.0	12.97±3.67
1EG4	59,077	102.9x85.4x80.7	0/1	6.0	6.0	12.420±3.64
2IVZ	73,690	113.0x92.7x84.1	2/0	6.0	6.0	12.05±3.59

^a Natoms is the number of atoms in the simulation; N_{ions} (Na+/Cl-) is the number of ions (Na+ or Cl-) used to neutralize system; σ_{OP} and σ_{OD} are the peptide essential and the second boost potential standard deviations in Pep-GaMD simulations; ΔV is the total boost potential.

Table S3. Comparison of docking models obtained using the *HPEPDOCK* and *AutoDock CrankPep (ADCP)*. *The model quality was assessed according to the CAPRI criteria and denoted in parentheses.

Peptide		1	2	3	4	5	6	7
RMSD (Å)	HPEPDOCK	3.87 (Medium)	4.50 (Medium)	5.49 (Acceptable)	7.80 (Acceptable)	8.50 (Acceptable)	14.00 (Inaccurate)	16.40 (Inaccurate)
	ADCP	9.12 (Acceptable)	5.78 (Acceptable)	5.22 (Acceptable)	4.55 (Medium)	7.13 (Acceptable)	9.70 (Acceptable)	17.24 (Inaccurate)

Table S4. Comparison of prediction models with different MD techniques (Pep-GaMD and cMD) for the PepBinding workflow. *The model's quality was assessed according to CAPRI peptide docking criteria, and its classification was put inside the brackets.

Peptide		1	3	6
RMSD (Å)	HPEPDock	3.87(Medium)	5.49(Acceptable)	14.12(Incorrect)
	Pep-GaMD	2.11 (Medium)	4.12 (Medium)	4.80 (Medium)
	cMD	1.57 (Medium)	4.65 (Medium)	13.07 (Incorrect)

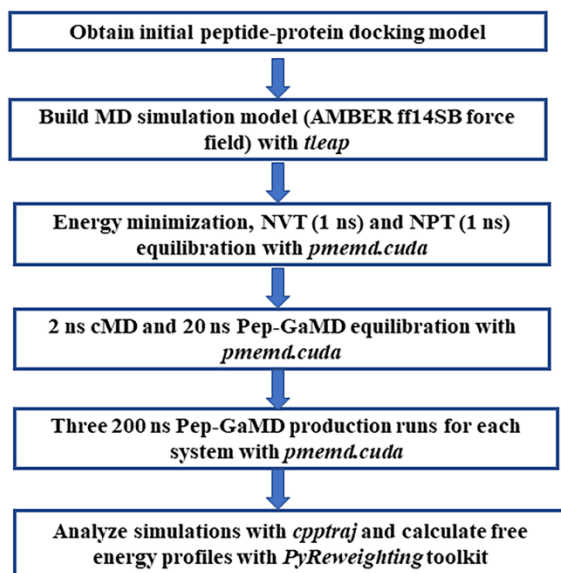


Figure S1. Protocol of Pep-GaMD simulations for refining peptide docking structures

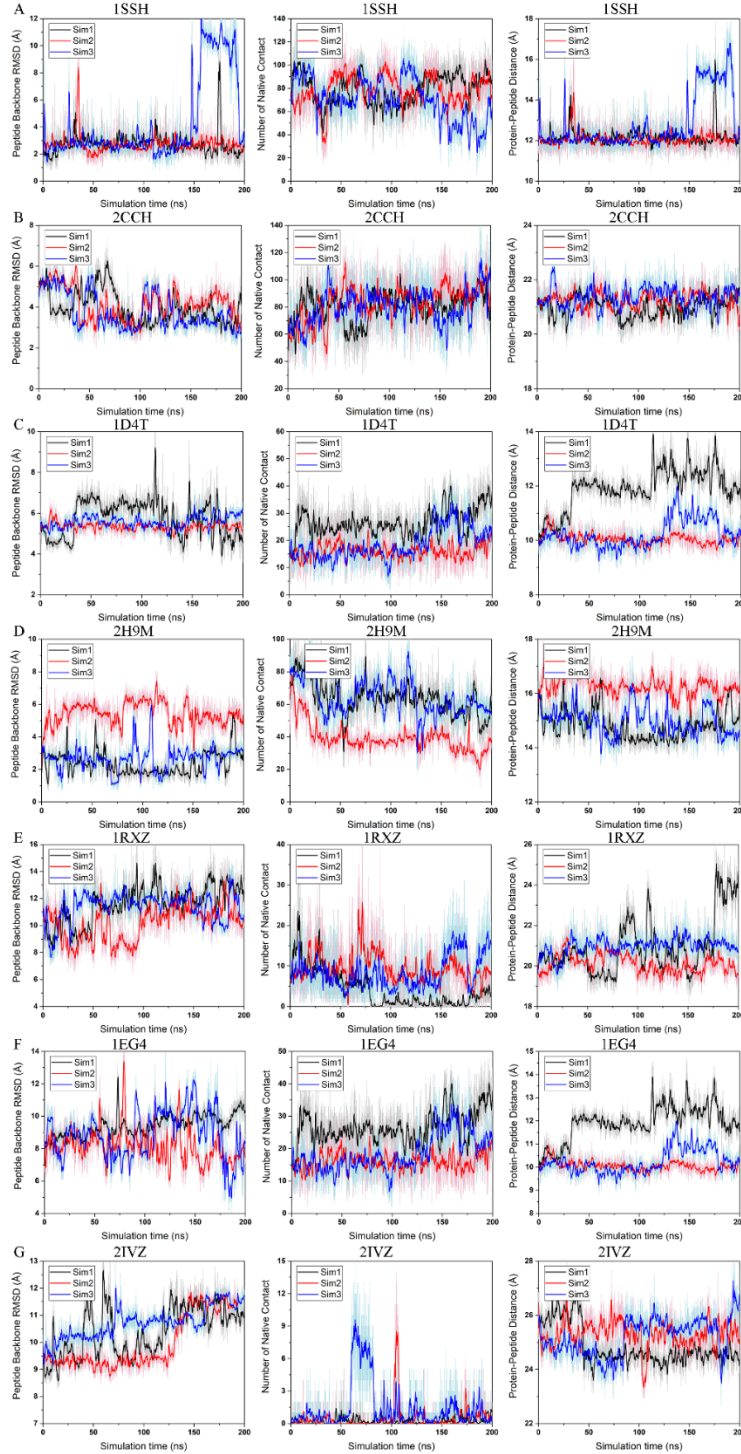


Figure S2. Time courses of peptide backbone RMSD, Number of native contacts and center-of-mass distance between peptide and protein obtained from three 200ns Pep-GaMD simulations with the zwitterion terminus on (A) peptide “GPPPAMPARPT” (Peptide 1, 1SSH), (B) “HTLKGRRLVFDN” (Peptide 2, 2CCH), (C) “KSLTIYAQVQK” (Peptide 3, 1D4T), (D) “ARTKQT” (Peptide 4, 2H9M), (E) “KSTQATLERWF” (Peptide 5, 1RXZ), (F) “NMTPYRSPPYVP” (Peptide 6, 1EG4), and (G) “GASDGSGWSENPNW” (Peptide 7, 2IVZ).

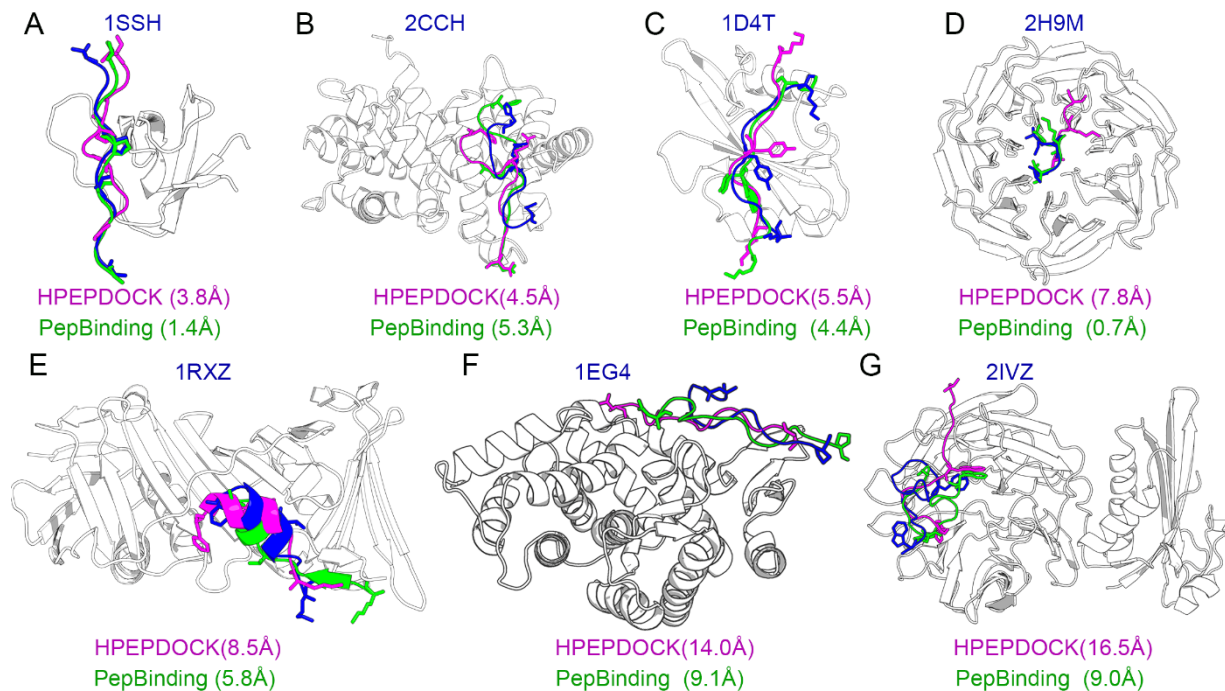


Figure S3. Docking poses of seven peptides obtained using *HPEPDOCK* (purple) and *PepBinding* (green) are compared with X-ray structures (blue) using the neutral peptide terminus: (A) peptide “GPPPAMPARPT” (Peptide 1), (B) “HTLKGRRLVFDN” (Peptide 2), (C) “KSLTIYAQVQK” (Peptide 3), (D) “ARTKQT” (Peptide 4), (E) “KSTQATLERWF” (Peptide 5), (F) “NMTPYRSPPPYVP” (Peptide 6), and (G) “GASDGSGWSSNNPW” (Peptide 7).

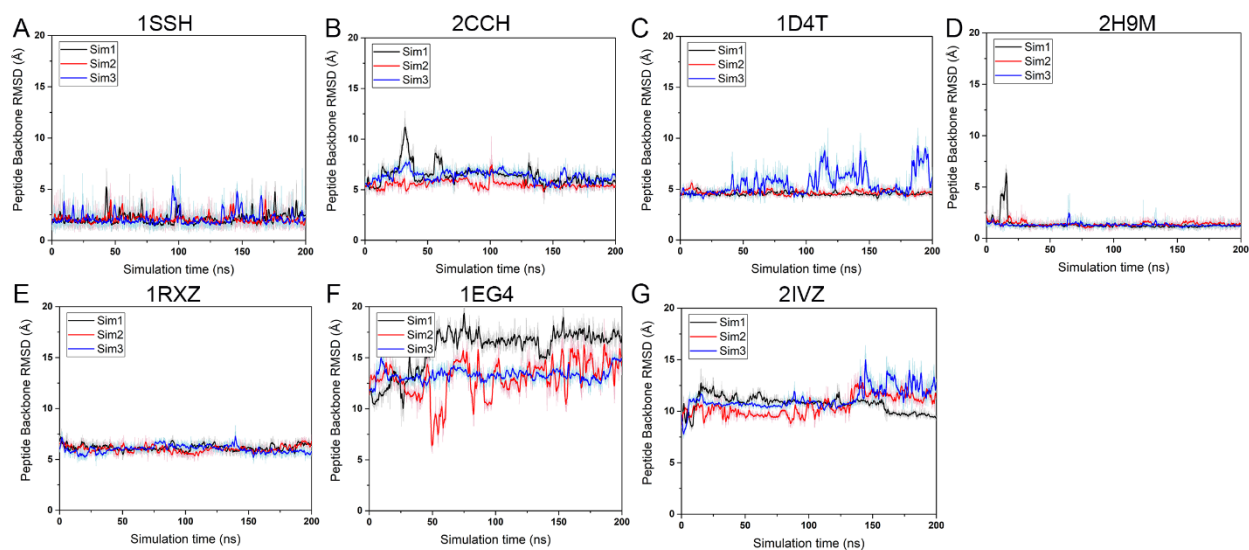


Figure S4. Time courses of peptide backbone RMSD obtained from three 200ns Pep-GaMD simulations with the *neutral terminus* on (A) peptide “GPPPAMPARPT” (Peptide 1), (B) “HTLKGRRLVFDN” (Peptide 2), (C) “KSLTIYAQVQK” (Peptide 3), (D) “ARTKQT” (Peptide 4), (E) “KSTQATLERWF” (Peptide 5)”, (F) “NMTPYRSPPPYVP” (Peptide 6), and (G) “GASDGSGWSENNPW” (Peptide 7).

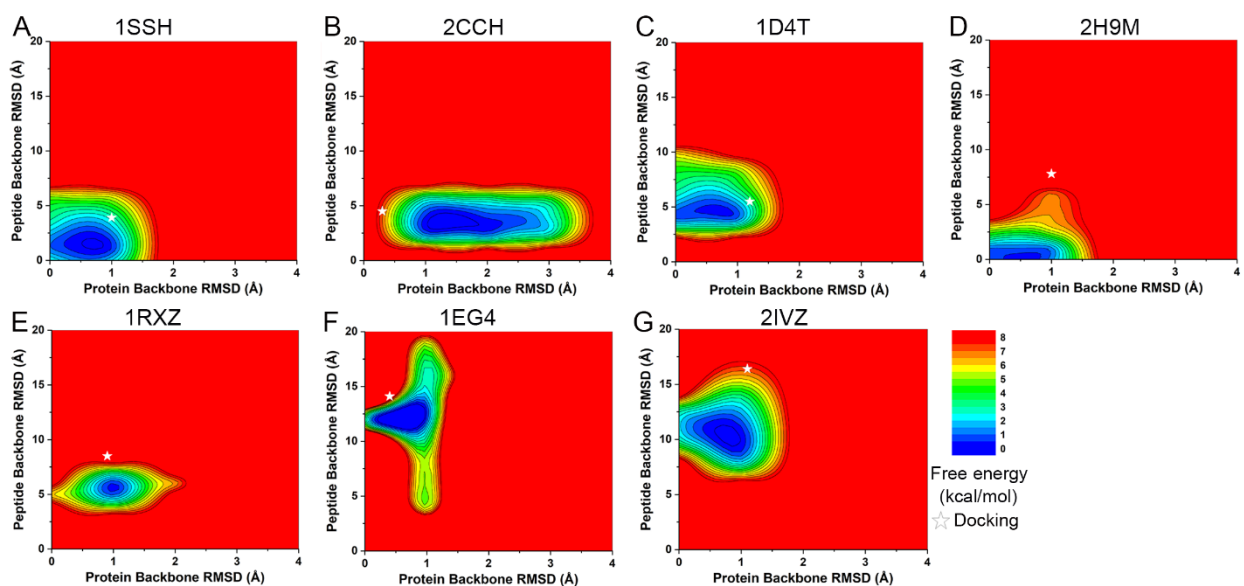


Figure S5. 2D potential of mean force (PMF) regarding the peptide backbone RMSD and protein backbone RMSD for peptides with the neutral terminus (A) “GPPPAMPARPT” (Peptide 1), (B) “HTLKGRRLVFDN” (Peptide 2), (C) “KSLTIYAQVQK” (Peptide 3), (D) “ARTKQT” (Peptide 4), (E) “KSTQATLERWF” (Peptide 5), (F) “NMTPYRSPPPYVP” (Peptide 6), and (G) “GASDGSGWSENNPW” (Peptide 7), respectively.

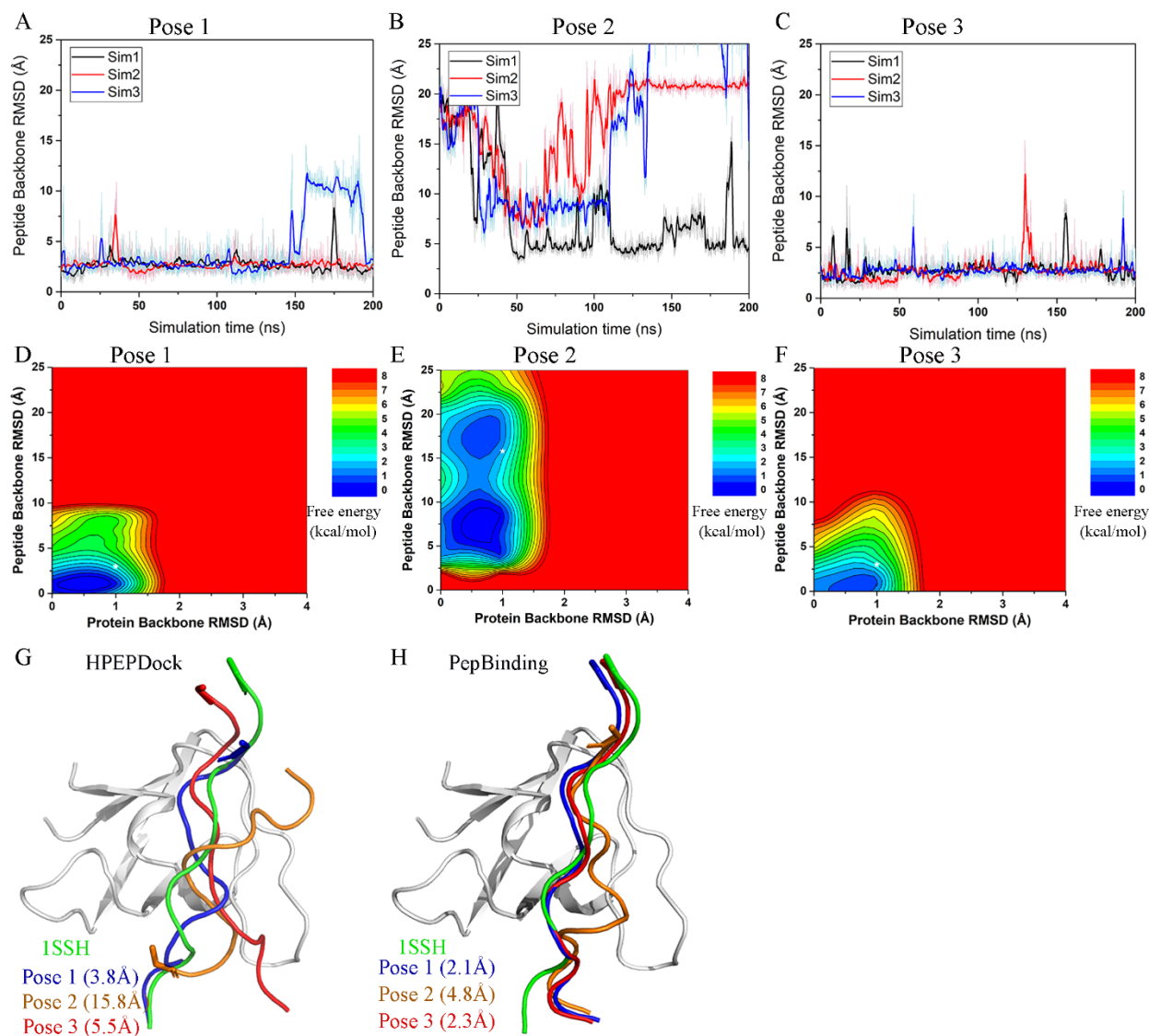


Figure S6. (A-C) Time courses of peptide backbone RMSD were obtained from Pep-GaMD simulations with the zwitterion terminus, starting from the top 3 poses predicted by HPEPDock for peptide 1 (1SSH). (D-F) The 2D PMFs were calculated based on the protein backbone RMSD and peptide backbone RMSD for peptide 1 starting from the top 3 binding poses. The white asterisks indicate the initial docking poses obtained using HPEPDock. (G-H) Binding poses of the three poses predicted using HPEPDock (G) and PepBinding (H) with the zwitterion terminus were compared with X-ray structures.

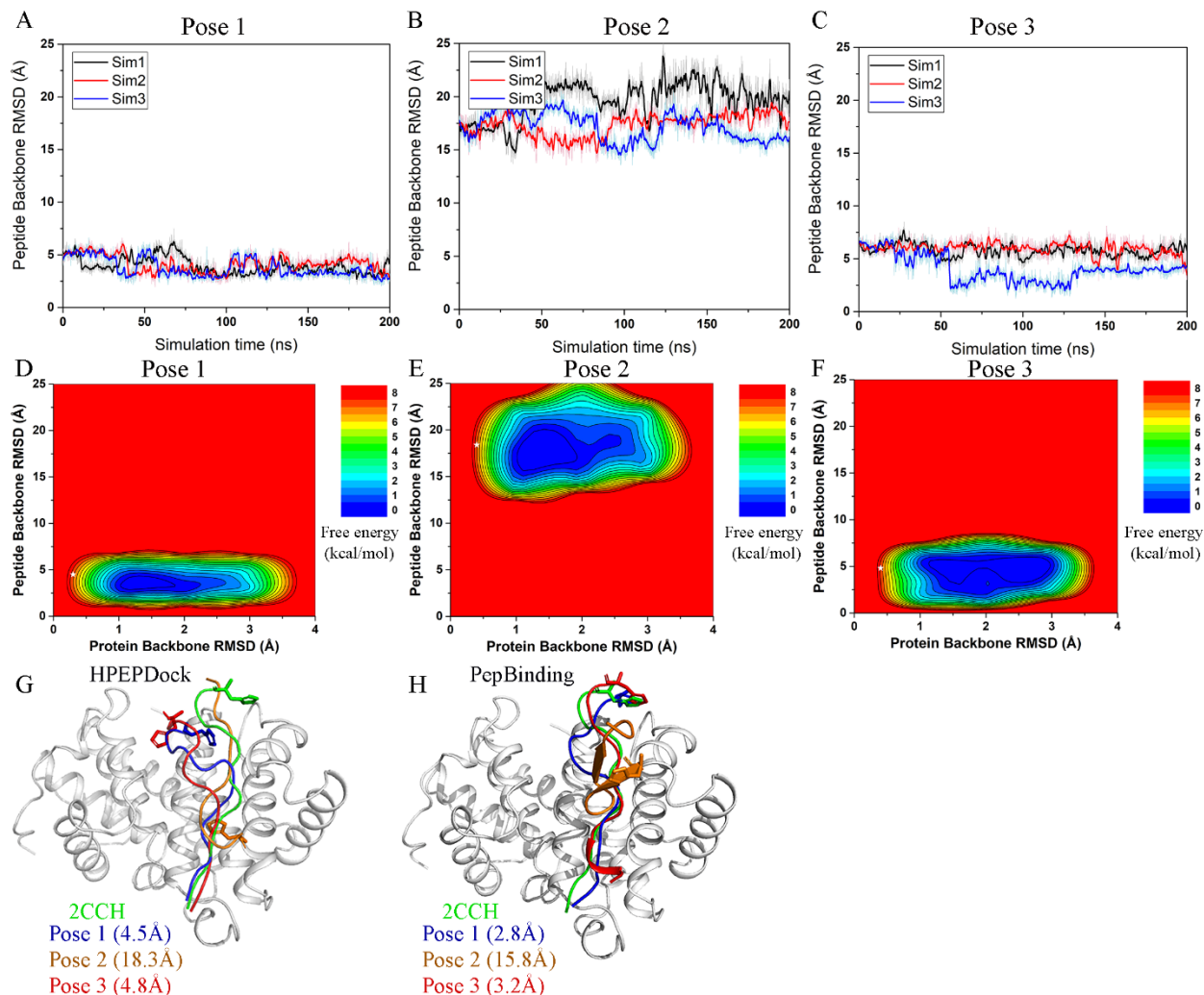


Figure S7. (A-C) Time courses of peptide backbone RMSD were obtained from Pep-GaMD simulations with the zwitterion terminus, starting from the top 3 poses predicted by HPEPDock for peptide 2 (2CCH). (D-F) The 2D PMFs were calculated based on the protein backbone RMSD and peptide backbone RMSD for peptide 3 starting from the top 3 binding poses. The white asterisks indicate the initial docking poses obtained using HPEPDock. (G-H) Binding poses of the three poses predicted using HPEPDock (G) and PepBinding (H) with the zwitterion terminus were compared with X-ray structures.

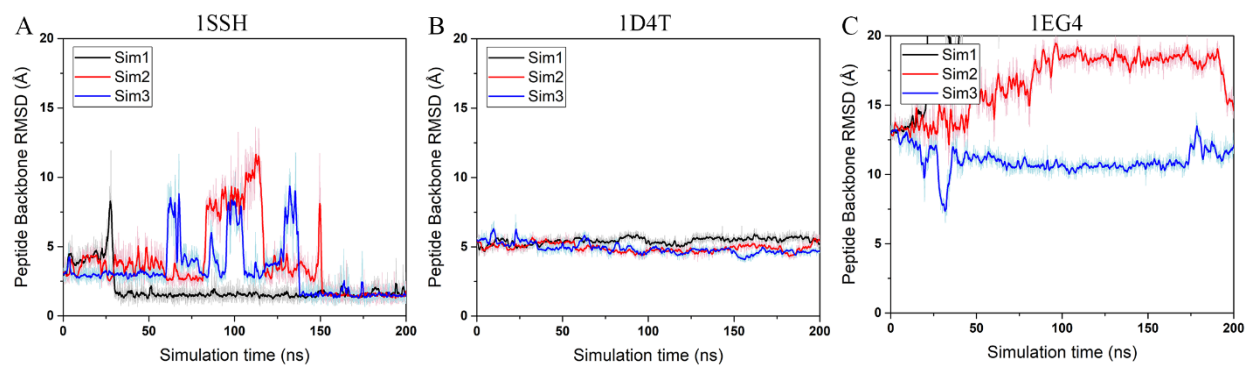


Figure S8. Time courses of peptide backbone RMSD obtained from three 200ns cMD simulations with the zwitterion *terminus* on (A) peptide “GPPPAMPARPT” (Peptide 1, 1SSH), (B) “KSLTIYAQVQK” (Peptide 3, 1D4T) and (C) “NMTPYRSPPPYVP” (Peptide 6, 1EG4).

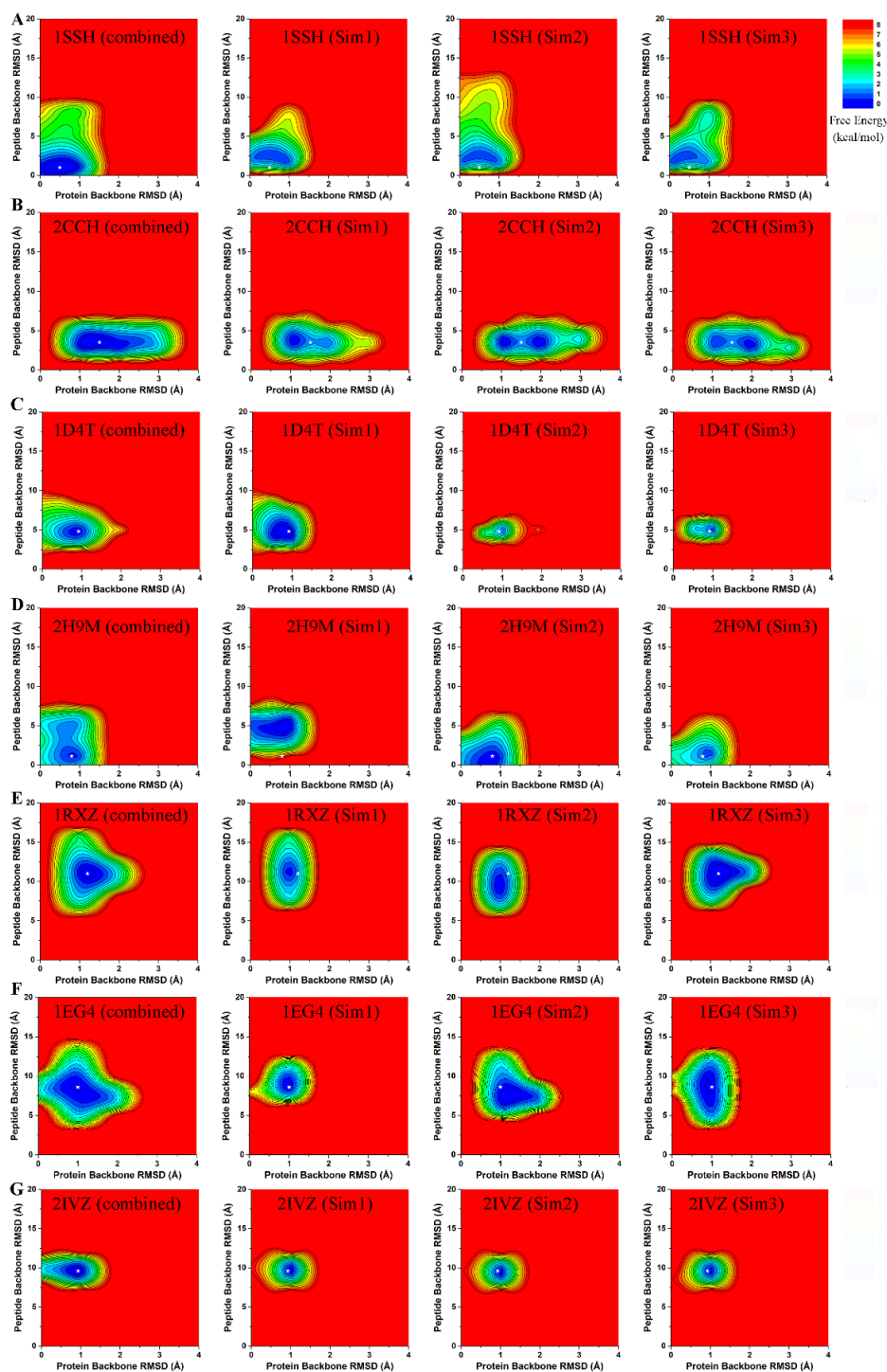


Figure S9. 2D PMFs calculated from three combined or individual Pep-GaMD trajectories regarding the peptide backbone RMSD and protein backbone RMSD for peptides with the zwitterion peptide terminus (A) “GPPPAMPARPT” (Peptide 1, 1SSH), (B) “HTLKGRRLVFDN” (Peptide 2, 2CCH), (C) “KSLTIYAQVQK” (Peptide 3, 1D4T), (D) “ARTKQT” (Peptide 4, 2H9M), (E) “KSTQATLERWF” (Peptide 5, 1RXZ), (F) “NMTPYRSPPPYVP” (Peptide 6, 1EG4), and (G) “GASDGSGWSENPNPW” (Peptide 7, 2IVZ), respectively. The lowest state from the combined PMFs was labeled as white star in all PMFs.

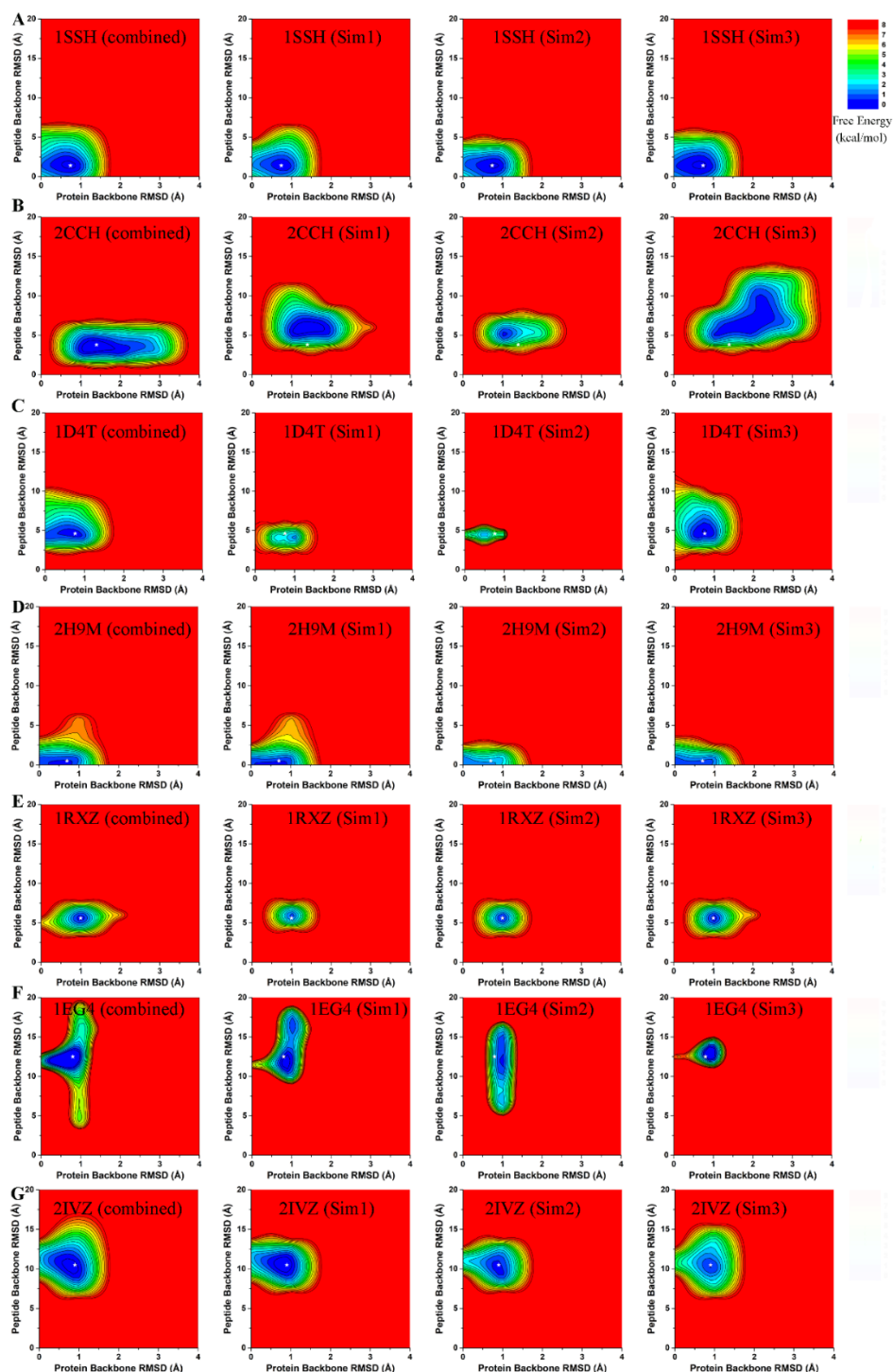


Figure S10. 2D PMFs calculated from three combined or individual Pep-GaMD trajectories regarding the peptide backbone RMSD and protein backbone RMSD for peptides with neutral peptide terminus (A) “GPPAMPARPT” (Peptide 1, 1SSH), (B) “HTLKGRRLVFDN” (Peptide 2, 2CCH), (C) “KSLTIYAQVQK” (Peptide 3, 1D4T), (D) “ARTKQT” (Peptide 4, 2H9M), (E) “KSTQATLERWF” (Peptide 5, 1RXZ), (F) “NMTPYRSPPPYVP” (Peptide 6, 1EG4), and (G) “GASDGSGWSSENNPW” (Peptide 7, 2IVZ), respectively. The lowest state from the combined PMFs was labeled as white star in all PMFs.