Arbitrarily Flexible Protein-Protein Docking

Investigation of hybrid solvent enhanced sampling molecular dynamics simulations for flexible protein-protein and protein-peptide interactions.

L. Schulig[1], N. Geist[2], M. Delcea[2], A. Link[1*],
[1] Institute of Pharmacy, University of Greifswald,
[2] Institute of Biochemistry, University of Greifswald

In Short

- Method development, investigation and optimization for computational host-guest recognition of protein complexes
- All-atom molecular systems with explicit solvation for flexible protein-protein docking using replica exchange molecular dynamics simulations
- TIGER2h with hybrid solvent potential energy calculations for optimal resource efficiency with a low number of replicas

Introduction The investigation of protein-protein interactions (PPI) is one of the central points in structural biology and is important for the understanding of protein's functions in biological processes and for the design of novel PPI inhibitors. An exhaustive search in the 6D rotational-conformational space of two interacting flexible molecules is computationally demanding and highly challenging. However, since the experimental elucidation of these complex structures is much more difficult as for the individual proteins, there is a high demand for accurate computational methods. Comparable to protein-ligand interactions with small molecules, it is possible to apply docking methods to study PPIs, as well.[1]

Protein-Protein Docking Methods The two main approaches for protein-protein docking are template- and FFT-based predictions. While the first needs high quality templates from similar proteins to provide accurate results, the FFT methods can be freely applied without prior knowledge of the complex.

A wide range of applications and servers like PIPER, ZDOCK, HADDOCK or pyDock are available to perform these calculations.[2] They are very fast and quite accurate for rigid body docking, but accounting for receptor and ligand flexibility is difficult as the number of possible complex conformations explodes.

To solve this problem at least partially, individual protein conformations can be pre-calculated using Monte Carlo simulations with coarse-grained protein representations or by molecular dynamics simulations (MDS). But efficiently sampling the backbone and sidechain positions of both interaction partners simultaneously during the docking process is still a major challenge.[3]

Molecular dynamics simulations MDS using explicit solvation have been proven to be a useful tool to accurately study the behavior of proteins and their corresponding complexes in solution. But due to sometimes large conformational changes and significant energy barriers between individual states, especially during protein complex formation, they often get stuck in local potential energy minima. Therefore, the successful application for the study of unknown protein-protein/peptide interactions strongly depends on the molecular system and is infeasible within typical simulation times.

However, a wide range of enhanced sampling methods have been developed to overcome this issue. While most of them are used to study the conformational space of a single molecule, e.g. protein folding, some can be generally applied to enhance the sampling of the overall molecular system like temperature replica exchange molecular dynamics (T-REMD) simulations. These are very accurate but are so computationally expensive that their use is impractical.

TIGER2h(s) We recently developed an updated algorithm of TIGER2 (temperature intervals with global exchange of replicas) by Li et al. which combines the resource efficient global replica exchange with a fast implicit solvent potential energy calculation while sampling the molecular system in explicit solvation (TIGER2h).[4] A significantly improved version of this algorithm, in terms of accuracy, reproducibility and efficiency, was developed as part of a previous HLRN project (mvb00012) and is currently submitted for publication.

Compared to T-REMD, only a few replicas are needed for each simulation, reducing the amount of computational resources to a fraction.[5]

The flexibility of both, host and guest molecules, can be freely adjusted using collective variables while limiting the conformational space and preventing significant unfolding beyond the binding site. This allows an induced fit like aggregation into the protein complex. All simulations are performed with explicit solvation to additionally include solvation effects and bridging water molecules.

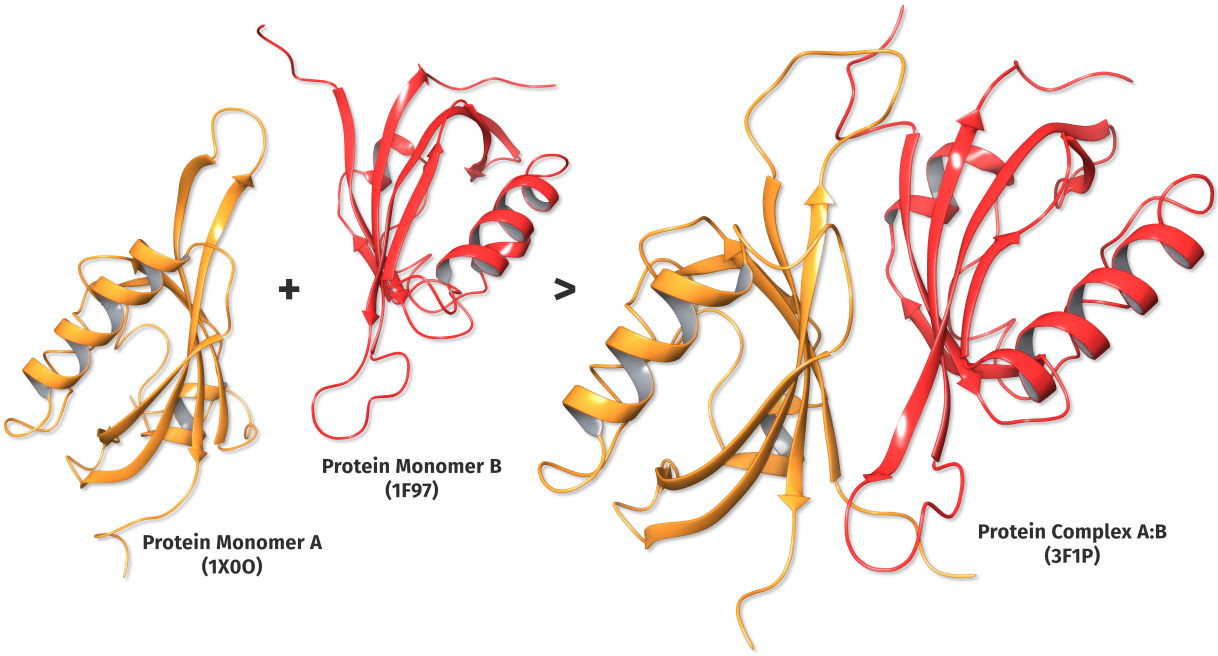

Figure 1: *Individual and heterodimer complex conformations of HIF2 alpha and ARNT C-terminal PAS domains. This protein complex is part of the Protein-Protein Docking Benchmark 5.5 with a high level of difficulty since significant conformational changes of loop regions are necessary for an accurate prediction.*

While these simulations are much more computationally demanding than the FFT-based methods and currently not suitable for high throughput, we expect a significant increase in accuracy, especially for larger conformational changes during the binding process.

In one of our previous studies, we were able to determine the homodimeric structure of the N-terminal zinc finger of the protein BCL11B for the wildtype and various mutants by using TIGER2h simulations with excellent agreement to experimental data.[6]

Objective The aim of this project is to investigate the suitability of TIGER2h for flexible protein-protein docking on a larger data set and to identify any problems in its practical application.

Benchmark To estimate the overall performance of TIGER2h for protein-protein docking, we are using the Protein-Protein Docking Benchmark 5.5 by Weng et al., consisting of 250 experimentally determined protein complexes, divided into eight categories and three difficulty levels.[7]

WWW

https://github.com/SLx64/TIGER2hs

More Information

[1] K.A. Porter et al., *Curr Opin Struct Biol* **55**, 1–7 (2019). doi:10.1016/j.sbi.2018.12.010

[2] I.A. Vakser et al., *Biophys J* **107**, 1785–1793 (2014). doi:10.1016/j.bpj.2014.08.033

[3] M. Kurcinski et al., *Int J Mol Sci* **14**, 7341 (2021). doi:10.3390/ijms22147341

[4] M. Kulke et al., *J Phys Chem B* **122**, 7295–7307 (2018). doi:10.1021/acs.jpcb.8b05178

[5] N. Geist et al., *J Phys Chem B* **28**, 5995–6006 (2018). doi:10.1021/acs.jpcb.9b03134

[6] L. Schulig et al., *Int J Mol Sci* **7**, 3650 (2021). doi:10.3390/ijms22073650

[7] T. Vreven et al., *J Mol Biol* **19**, 3031–3041 (2015). doi:10.1016/j.jmb.2015.07.016

[8] J.C. Phillips et al., *J Comput Chem* **26**, (2005). doi:10.1002/jcc.20289