FLEX-LZERD: PROTEIN DOCKING WITH EXTREMELY LARGE CONFORMATIONAL CHANGES

Charles Christoffer[1], and Daisuke Kihara[1,2,*]

[1] Department of Computer Science, Purdue University, West Lafayette, Indiana, 47907, USA
[2] Department of Biological Sciences, Purdue University, West Lafayette, Indiana, 47907, USA

Large flexibility of proteins as they interact is key to their function in many systems, such as with calmodulin, importin, or interleukin receptors. To understand these mechanisms at the molecular level, it is necessary to consider the 3D atomic structures involved. When these have not been determined, protein docking can generate useful structure models. However, traditional docking typically considers flexibility limited to small movements on the order of ~2 Å root mean square deviation (RMSD), or otherwise to restrictively small structures. These methods account for such minor backbone flexibility via conventional normal mode or other sampling, but modeling ordered proteins of arbitrary size which undergo large-scale conformational changes, on the order of 10 Å RMSD and above, requires different treatment to be generally feasible.

Here, we present Flex-LZerD, a computational method for assembling such complexes with large-scale conformational changes on the order of 10 Å RMSD and above. Via an initial partial-assembly multidomain docking and an iterative anisotropic normal mode analysis admitting curvilinear motions, we demonstrate the ability to model the assembly of a variety of protein-protein and protein-nucleic acid complexes. Also yielded are full-atom trajectories taking the input flexible protein from its initial conformation to its docked conformation, which could shed light on potential intermediate states. AlphaFold and other current deep learning-based methods predict static structure models well, but do not explicitly address flexibility. AlphaFold is further limited to proteins, while here protein-nucleic acid complexes are also considered. Flex-LZerD successfully modeled protein-protein complexes for 5 of 9 unbound test targets. On a further unbound-bound study of 15 protein-protein complexes and 8 protein-nucleic acid complexes, Flex-LZerD likewise successfully modeled 9 and 8, respectively. Flex-LZerD is thus a step toward the explicit treatment of flexibility and multiple states that is the future of structure modeling.