StAIR: Stepwise All-Atom Iterative Reconstruction of Large Coarse-Grained Protein Structures

Wong, Erin (School: William A. Shine Great Neck South High School)

Molecular dynamics (MD) simulations model how proteins change conformation during key processes like folding and substrate binding. However, simulating every atom, as is done in all-atom simulations, is computationally expensive for larger proteins. To address this, coarse-graining (CG) groups atoms into fewer “beads”, reducing the number of degrees of freedom in the CG simulation. Still, backmapping, which restores atomic details to these simplified, low resolution CG structures after simulation, remains challenging. Current methods can only accurately backmap from CG structures where one bead represents an amino acid, limiting the speedup of CG simulations. This study introduces StAIR, a novel backmapping framework that successfully backmap from much lower-resolution CG structures using as few as five beads. Unlike the current one-step method, CGVAE, StAIR backmaps twice, employing two machine learning models: the first backmaps to a higher CG resolution, and the second restores full atomic detail. StAIR was trained on real all-atom MD simulation data from the eIF4E and SARS-CoV-2 PLpro proteins and evaluated against CGVAE using root mean square deviation (RMSD) for accuracy and Ramachandran plots for reconstruction quality, comparing backmapped and real structures for each. Overall, StAIR achieved 3x lower RMSD while backmapping from five beads, 10x fewer than what CGVAE can accurately handle, enabling simulations to be run up to 176x faster. Adding a third step further improved reconstruction quality, suggesting how StAIR can be scaled in a similar manner to backmap more complex proteins. StAIR enables highly efficient, biologically accurate simulations, making large-scale protein studies and structure-based drug design significantly faster and more accessible.

Awards Won: