ERPD: A Novel Protein Design Approach to Engineering and Optimizing Smaller Fluorescent Proteins

Liu, Xinyi (School: Pine View School)

Green fluorescent proteins (GFP) are used as markers to image β-amyloid plaques in Alzheimer's diagnosis, visualize cancer progression, and measure drug toxicity. However, the large size of GFP can impair the target protein's function and prevent accurate visualization. Some GFPs have shown to be prone to aggregate, potentially due to the dominating beta sheets, which leads to toxicity and cell death. Additionally, rapid viral gene expression processes require proteins with high fluorescence, stability, and maturation rate, characteristics that GFPs don't simultaneously have. I developed ERPD (Editable Refinement Protein Design) - a pipeline approach combining diffusion models, homology modeling, MD simulations, and physical-based energy analysis to design stable and experimentally feasible proteins with desired functions. Experimental and computational testing verified the increase in stability, fluorescence, and maturation rate of the proteins designed using ERPD, compared to traditional GFPs, with some outperforming EGFP by 50% in quantum yield. The designs can potentially replace GFP as more fluorescent and stable fluorescent markers, and their smaller sizes allow them to be less disruptive when expressed in vivo with target proteins. ERPD serves as a novel protein engineering approach that reduces time and cost from current practices of testing uncertain proteins such as directed evolution. Generating promising proteins with over 50% increase in folding feasibility in just two designs, ERPD prevents running extensive algorithms and manually sorting through thousands of generated designs. Applying ERPD to fluorescent protein engineering demonstrates the effectiveness of ERPD, which can be used to design virtually any protein for any function.