Computational Protein Engineering for Systematic Enhancement of Crystallization Propensity

Nooriel E Banayan[1], Blaine Loughlin[1], Dr. Shikha Singh[1], Dr. Farhad Forouhar[1], Dr. John F Hunt[1]

[1]*Columbia University*

neb2154@columbia.edu

Protein crystallization is the gold-standard method for elucidating protein structures, specifically drug-target structures, in industry and academia, yet many proteins do not crystallize and there are currently no reproducible methods to crystallize recalcitrant proteins. Crystallization propensity, therefore, is an intrinsic property that depends on the structural and thermodynamic properties of proteins. To elucidate the structural features that give rise to high crystallization propensity, we data-mined the Protein Data Bank (PDB), analyzed 87,684 crystal structures and found specific, generally high-entropy epitopes to be over-represented in crystal-packing interfaces. To explain this observation, we hypothesized that although there is an unfavorable entropy cost of stabilizing these epitope, the entropy cost can be overcome if these epitopes create enough favorable crystal-packing contacts in a crystal lattice--which may explain their over-representation in crystal-packing interfaces. To this end, we developed a computational pipeline called "BulkM" that uses protein homology analysis to engineer proteins at select sites to improve the crystallization propensity of recalcitrant proteins. We applied this method to three proteins--two of which crystallize readily and have structures available and one of which is recalcitrant with no publicly available structure—and in all cases, our computational pipeline significantly increased crystallization propensity leading to structure determination. Our observations indicate that "BulkM" successfully engineers proteins at favorable sites which make recalcitrant proteins crystallize while conserving the protein's wildtype structure. This method expediates the process of solving protein structures and allows for a two week turn- around time from protein expression to a crystal structure allowing scientists to quickly solve high-impact and high-resolution (<3.0 Angstroms) protein structures. This method can significantly streamline the drug-target discovery process and have major impacts in medicine and science.