

Combinatorial Biophysics: Library and Statistical Protein Design

Thomas J. Magliery

Departments of Chemistry and Biochemistry, The Ohio State University, 100 W. 18th Ave.,
Columbus, OH 43210-1185, USA

Both the prediction and design of protein structure, using computational and rational approaches, remain significant challenges in protein chemistry. A major limitation to developing a comprehensive physicochemical model of the protein structure-sequence relationship is the vastness of sequence space and the low-throughput nature of biophysical studies. We are pursuing two avenues to understand better the sequence structure-relationship: sorting large libraries of protein variants for structured proteins, and statistical analysis of ubiquitous protein families for protein redesign. In the combinatorial approach, we have developed a high-throughput cell-based screen for activity of the well-studied four-helix bundle protein Rop. Rop controls the copy number of ColE1 plasmids in *E. coli*, and therefore the expression level of ColE1-borne genes (here, GFP). Due to our detailed understanding of Rop structure from systematic studies, we can generate rational libraries of Rop variants (e.g., variants with randomized hydrophobic cores) such that, to a reasonable approximation, active variants are folded variants. A number of surprising “families” of soluble, active variants have already been isolated from small libraries, including overpacked variants with molten-globule-like thermodynamic properties. We are also interested in the role of correlated occurrences of amino acids in natural protein families. To that end, we have generated a consensus version of triose phosphate isomerase (TIM), which can be thought of as a “correlation-free” variant, as a host to interrogate the roles of correlated positions by mutagenesis and library methods. Surprisingly, the consensus TIM, which is only 65% homologous to any natural TIM, is active; more surprisingly, it has some molten-globule like properties that are seemingly inconsistent with the accepted mechanism of catalysis. The implications and utility of these combinatorial approaches to biophysical problems will be discussed.