Computational design of functional protein repertoires

Sarel Fleishman, Weizmann Institute of Science, Israel

Understanding how protein sequence and structure determine function would unlock vast opportunities in basic and applied research. Our lab is developing strategies that combine phylogenetic analysis and Rosetta atomistic calculations to design optimized variants of natural proteins. These strategies have been used by thousands of users worldwide to generate stable therapeutic enzymes, vaccine immunogens, highly active enzymes, and membrane proteins for a range of needs in basic and applied research. We now present a machine-learning strategy to design and economically synthesize millions of active-site variants that are likely to be stable, foldable, and active. We applied this approach to the chromophore-binding pocket of GFP to generate more than 16,000 active designs that comprise as many as eight mutations in the active site. The designs exhibit extensive and potentially useful changes in every experimentally measured parameter, including brightness, stability, and pH sensitivity. We also applied this strategy to design millions of glycoside hydrolases that exhibit significant backbone changes in the active site. Here too, we isolated more than 10,000 catalytically active and very diverse designs. Contrasting active and inactive designs illuminates areas for improving enzyme design methodology. This new approach to high-throughput design allows the systematic exploration of sequence and structure spaces of enzymes, binders, and other functional proteins.