Rational Design of Functional De Novo Proteins

Proteins are composed of 20 different amino acids, providing a vast range of functions and a high level of diversity. Because the possible space is too huge, it remains difficult to rationally design de novo proteins without a thorough understanding of the fundamental relationships between amino acid sequence, structure, and function. To overcome this challenge, we are combining large-scale measurements with machine learning, including deep learning, to gain a better understanding of the relationships between amino acid sequence, structure, and function. Then, based on this new knowledge, we design de novo proteins to verify and re-analyze the new findings. In this way, we repeat such analysis and design to achieve both an understanding of the fundamental laws of proteins and rational design of de novo proteins.