Abstract

There is no known life without proteins. Proteins are indispensable for virtually all biologically meaningful processes, functioning as structural components, transporters, containers, receptors, signaling molecules, regulatory factors, and, above all, enzymes. They are the basic building blocks of life, and understanding their properties is essential for advancing the frontiers of medicine. However, proteins themselves cannot exist without RNA—another fundamental biopolymer. Notably, recent advances have made the study of RNA increasingly parallel to that of proteins. Both proteins and RNAs occupy the boundary between the micro- and macro-worlds, attracting the attention of researchers from biology, chemistry, biophysics, bioinformatics, and pharmacy. While most existing drugs act by modulating protein activities, the importance of RNA as a target in medical research is rapidly growing.

A central theme in protein science is the relationship between sequence, structure, and function. Proteins are biopolymers composed of amino acids, assembled by ribosomes (large RNA-protein complexes) within our cells. The resulting chain of amino acids can (ideally) fold into a functional protein in a physics-driven process. While the amino acid sequence determines the structure and dynamics of the folded protein, the resulting structure and dynamics ultimately govern its biological functions.

The significance of structural protein science was underscored by last year's Nobel Prize in Chemistry, awarded "for computational protein design" (David Baker) and "for protein structure prediction" (Demis Hassabis and John Jumper, representing the AlphaFold project). The AI revolution has fundamentally transformed the field: from roughly 100,000 experimentally determined protein structures, we now have accurate structural predictions for all known amino acid sequences—over 200 million in total—allowing unprecedented opportunities for large-scale protein analysis.

Classically, protein structures are described in terms of four levels of organization, but this overlooks a crucial property influencing the entropic freedom of protein chains: topology. For the past 25 years, it has been recognized that some proteins are intrinsically knotted—a property not merely incidental, but one that dramatically affects their stability and dynamics. More recently, a wider variety of topologically complex motifs have been discovered, including knots, slipknots, lassos, θ-curves, and handcuffs—discoveries largely inspired by the application of knot theory to protein science.

In this thesis, I present my contributions to knot theory and the structural biology of topologically complex proteins and other biopolymers. I first describe a solution to a 50-year-old problem posed by John Conway: the classification of algebraic tangles. Algebraic tangles provide the foundation for classifying spatial graphs, themselves a basis for recognizing newly discovered, topologically complex protein folds. Secondly, I present the detailed characterization of lasso proteins and slipknotted transmembrane proteins. Furthermore, I describe the discovery of new knotted folds: a 6_3 protein knot, a Brunnian handcuff in protein, and—most notably—the first naturally occurring RNA molecule with a trefoil (3_1) knot. Additionally, I introduce practical tools for the broader scientific community, including AlphaKnot, AlphaLasso, and the Topoly Python package, which facilitate topological analysis of biopolymers, even for those without advanced training in topology or programming.