A Novel Orientation-Based Statistical Potential for Efficient Prediction of Protein Structure

Sivaraman, Venkatesh

The ability to predict protein structure is a major challenge in computational biology, with wide-ranging applications from understanding disease mechanisms to designing targeted therapies for disorders such as Alzheimer's disease and cancer. Approximating protein stability is crucial to solving this problem because proteins generally fold into the structure that minimizes their Gibbs free energy. A novel approach called Segmented Positional Analysis of Residue Contacts (SPARC) was developed to address this need, using inter-residue orientations derived from a database of known proteins to estimate a structure's stability. SPARC was then refined by separating short- and long-range interactions and adding a hydrophobicity reference state based on protein volume. The resulting potential showed an accuracy of 80.5% on two databases of decoy structures, a significant improvement over RWplus and GOAP, two other well-established statistical potentials ($p < 0.0001$). Leveraging the accuracy of SPARC, a dynamic Monte Carlo simulation algorithm was developed to predict protein structure given only the amino acid sequence. The new algorithm simulates small segments individually, then recursively combines stable segments to efficiently generate full structures. Applied to a short simulation of bovine pancreatic trypsin inhibitor, SPARC achieves a conformation with root-mean-square deviation 10.94 Å. Together, SPARC and the new simulation algorithm will enable biologists to investigate systems that are difficult to study experimentally, such as transmembrane proteins, to study the causes of protein-related disorders, and to engineer new proteins.

Awards Won:

Fourth Award of $500