

ABSTRACT

Impressive advances in genomic sequencing technologies are flooding us with the complete genetic blueprints of human, rat, mouse, chimpanzee and various microorganisms at an extremely rapid pace. As these DNA sequences continue to accumulate, the challenge to determine the function of each gene is paramount. These functions are determined by their unique three-dimensional (3D) structures as a result of protein folding. Protein folding can be defined as the process in which proteins spontaneously arrange their linear sequence of amino acids into native 3D structures that will allow them to function properly. Thus, elucidation of the 3D structure of a protein is vital in understanding its function. However, it is not known how the newly synthesized polypeptide chains fold into a protein with specific function. It was not until 1973 that Anfinsen [1] postulated that all the information needed for a protein to correctly fold into its native structure is encoded solely in its amino acid sequence and that the native state of the protein is the conformation with the lowest energy. Consequently, this important principle has brought immense interests among the scientists to investigate how proteins fold into their native structures and eventually determine the correct functional fold for the proteins. The two most mature and conventional experimental techniques to solve the structure of a protein are the X-ray crystallography and Nuclear Magnetic Resonance (NMR). John Kendrew and Max Perutz shared the Nobel Prize in 1962 for their pioneering achievement in solving the atomic level structure of the protein myoglobin and hemoglobin, respectively, using X-ray diffraction. Since then, many protein structures were solved and various roles of proteins in living cells were known. Despite the accuracy and the advances of these experimental techniques, such methods are very costly and it may take months to years for solving one structure. The current number of 3D protein structures is very small compared to the number of protein sequences, creating a huge gap between them. It has become more pressing with the growth of genome sequencing projects providing protein sequences for which structural information is not available. As this gap is expected to keep on growing with the ongoing genome projects, the experimental techniques certainly cannot be expected to keep pace with the rapid flow of the sequences. This has caused an urgent need to accurately predict the 3D structure of proteins from the linear chain of amino acid sequence using other methods especially computational work that relies heavily on theoretical studies. However, despite decades of active research and the impressive advances [2,3,4,5], computational protein structure prediction and protein folding remain one of the most important unsolved problems in structural biology today [6,7,8].