Chapter 21
Prediction of Structural and Functional Aspects of Protein: In-Silico Approach

Arun G. Ingale
North Maharashtra University, India

ABSTRACT

To predict the structure of protein from a primary amino acid sequence is computationally difficult. An investigation of the methods and algorithms used to predict protein structure and a thorough knowledge of the function and structure of proteins are critical for the advancement of biology and the life sciences as well as the development of better drugs, higher-yield crops, and even synthetic bio-fuels. To that end, this chapter sheds light on the methods used for protein structure prediction. This chapter covers the applications of modeled protein structures and unravels the relationship between pure sequence information and three-dimensional structure, which continues to be one of the greatest challenges in molecular biology. With this resource, it presents an all-encompassing examination of the problems, methods, tools, servers, databases, and applications of protein structure prediction, giving unique insight into the future applications of the modeled protein structures. In this chapter, current protein structure prediction methods are reviewed for a milieu on structure prediction, the prediction of structural fundamentals, tertiary structure prediction, and functional imminent. The basic ideas and advances of these directions are discussed in detail.

INTRODUCTION

A protein is a sequence produced by combination of 20 amino acids. Amino acids are characterized by polar (hydrophilic) and non polar (hydrophobic) based on its residue. These amino acids are associated to each other with peptide bond to form a protein sequence (Laskowski*et al.,* 2003). The fabrication of a protein sequence is much effortless than the determination of a protein structure. However, the structure of a protein gives much more impending in the function of the protein than its sequence. Consequently, a number of methods for the computational prediction of protein structure from its sequence have been

DOI: 10.4018/978-1-5225-1762-7.ch021

developed. Protein structure prediction is the prediction of the three-dimensional structure of a protein from its amino acid sequence that is, the prediction of its secondary, tertiary, and quaternary structure from its primary structure. The native structure is the 3D structure of the protein sequence. It necessitates knowing the native protein structure so that we can find out the function of that protein structure. There are numerous algorithms used to predict the native confirmation of a protein. Protein sequences are folded on lattice with non overlapping amino acid chain. These self circumventing conformations produced native structures which have smallest energy configuration. The protein sequence-structure gap is broadening speedily. The number of identified protein sequences (*Bairoch and Apweiler, 2000)* is blasting as a outcome of genome and other sequencing projects. The rising number of protein sequences is much bigger than the escalating number of known protein structures (Berman *et al*., 2000). Hence, computational predictive tools for protein structures are badly needed to narrow the widening gap. Prediction of 3D protein structure from the amino acid sequence is one of the most exigent tribulations in theoretical structural biology (Montelione *et al*., 1999; Skolnick *et al*., 2000).Three-dimensional protein structures still cannot be correctly predicted straight from sequences. A transitional but helpful step is to predict the protein secondary structure, which is a way to simplify the prediction problem by projecting the very complicated 3D structure onto one dimension, i.e. onto a string of secondary structural coursework for each residue.

Fortitude of the exclusive tertiary (three-dimensional) structure of a protein from its amino acid sequence alone is one of the most significant and exigent problems in modern biology. The information on the tertiary structure of a protein is pretty decisive in accepting the function and biological responsibility of the protein. At present, genome-sequencing projects are producing an extraordinary amount of linear amino-acid sequences. An exponential growth of protein sequence database in current years by far outpaces the experimental determination of protein tertiary structures. Structure prediction is vitally dissimilar from the inverse problem of protein design. Protein structure prediction is one of the most key targets accomplished by bioinformatics and theoretical chemistry; it is extremely important in medicine and biotechnology for designing of the drug and novel enzymes.

This chapter begins with a thorough introduction to the protein structure prediction problem and is divided into four themes: a background on structure prediction, the prediction of structural elements, tertiary structure prediction, and functional insights.

The Protein Structure

Proteins are building blocks of life. It reveals more sequence and chemical complexity than DNA or RNA. It is a polymeric macromolecule fictitious of amino acid building blocks set in a linear chain and coupled jointly by peptide bonds. A protein sequence is a linear hetero polymer made up of one of the 20 different amino acids. Protein structures range in size from tens to several thousand residues (Brocchieri and Karlin, 1987). Protein structure is the bimolecular structure of a protein molecule that is a sequence formed from various residues. To be able to perform their biological function, proteins fold into one or more specific spatial conformations, driven by a number of non-covalent interactions such as hydrogen bonding, ionic interaction, Van der waals forces, and hydrophobic packing. The primary structure is usually symbolized by a sequence of letters above a 20-letter alphabet linked with the 20 naturally arising amino acids. Proteins are the foremost building blocks and useful molecules of the cell, pleasing up roughly twenty percent of a eukaryotic cells weight, the biggest involvement after water. They execute ample range of functions in the living organism, playing diverse catalytic, structural, regulatory and sig-