Chapter 3
Protein Structure Prediction

Hirak Jyoti Chakraborty
Central Inland Fisheries Research Institute, India

Sayak Ganguli
Amplicon Institute of Interdisciplinary Science and Technology, India

Aditi Gangopadhyay
Jhargram Raj College, India

Abhijit Datta
Jhargram Raj College, India

ABSTRACT

The great disagreement between the number of known protein sequences and the number of experimentally determined protein structures indicate an enormous necessity of rapid and accurate protein structure prediction methods. Computational techniques such as comparative modeling, threading and ab initio *modelling allow swift protein structure prediction with sufficient accuracy. The three phases of computational protein structure prediction comprise: the pre-modelling analysis phase, model construction and post-modelling refinement. Protein modelling is primarily comparative or* ab initio*. Comparative or template-based methods such as homology and threading-based modelling require structural templates for constructing the structure of a target sequence. The* ab initio *is a template-free modelling approach which proceeds by satisfying various physics-based and knowledge-based parameters. The chapter will elaborate on the three phases of modelling, the programs available for performing each, issues, possible solutions and future research areas.*

INTRODUCTION: THE PROTEIN FOLDING PROBLEM

The protein folding problem is one of the top 125 problems in science (Dill, Ozkan, Shell, & Weikl, 2008). It is both bewildering and beautiful how cells have been structuring amino acid strings into their precise folds through millions of years of evolution. It is one of the supreme mysteries of Nature that man is striving to understand. So how old is this folding problem? It began quite harmlessly in the 1960s when mankind first set eyes on the atomic structure of the protein. Back then, he expected to see more regular structures, instead of the irregularly packed globin. It was then that he set his foot on "How do

DOI: 10.4018/978-1-5225-2607-0.ch003

proteins fold?"and has been on the journey ever since. The folding problem has three facets: the folding code, the structure prediction and the folding process (Dill et al., 2008). The folding code problem deals with the system thermodynamics that determine the fold the protein is adopting. Structure prediction concerns itself with predicting the structures of proteins from amino acid sequences with computational power. The question of the protein folding process tries to find answers to the routes that proteins follow to achieve a particular structure. This chapter deals with the second facet of the folding problem: the computational approach to predicting protein structures.

BACKGROUND: MAKING PROTEINS WITH MACHINES

The wide disparity between the number of experimentally derived protein structures and the number of protein sequences in databases clearly indicates that experimental structure prediction methods are lagging behind the sequencers. Less than 1/500th of protein sequences have corresponding experimentally available structures (Moult, Fidelis, Kryshtafovych, Schwede, & Tramontano, 2016). However, the structure of a protein is a mandatory requirement for several applications in biology and medicine such as evolution, interactome study, drug design, protein function, enzymology or molecular biology. Experimental techniques such as NMR, crystallography or cryo electron microscopy are used to determine accurate structures of proteins and biomolecules. Some proteins pose a challenge during crystallisation, such as the membrane proteins, which owing to partial hydrophobicity and instability are difficult to purify (Carpenter, Beis, Cameron, & Iwata, 2008; White, 2004). Structural information of these proteins may be vital for drug design, as in the very case of membrane proteins, which make up more than 40% of drug targets (Overington, Al-Lazikani, & Hopkins, 2006). Also, experimental techniques for protein structure determination are hugely time consuming. In situations like these, protein structure prediction asks for computational techniques with the ability to provide rapid and reliable structures of proteins (Al-Lazikani, Jung, Xiang, & Honig, 2001; Dorn, E Silva, Buriol, & Lamb, 2014; Hardin, Pogorelov, & Luthey-Schulten, 2002). It is necessary to note that since comparative computational protein structure prediction is dependent on existing experimental structures, the accuracy and reliability in such cases heavily relies on the robustness of template selection and alignment.

PHASES OF PREDICTION

Protein modelling can be generally categorised into three phases: pre-modelling analyses, modelling and post-modelling refinement (Figure 1).

Figure 1. Phases of protein modelling: pre modelling analyses, modelling and post-modelling refinement