Chapter 40
Multi-Agent Systems in Three-Dimensional Protein Structure Prediction

Leonardo de Lima Corrêa
Federal University of Rio Grande do Sul, Brazil

Márcio Dorn
Federal University of Rio Grande do Sul, Brazil

ABSTRACT

Tertiary protein structure prediction in silico is currently a challenging problem in Structural Bioinformatics and can be classified according to the computational complexity theory as an NP-hard problem. Determining the 3-D structure of a protein is both experimentally expensive, and time-consuming. The agent-based paradigm has been shown a useful technique for the applications that have repetitive and time-consuming activities, knowledge share and management, such as integration of different knowledge sources and modeling of complex systems, supporting a great variety of domains. This chapter provides an integrated view and insights about the protein structure prediction area concerned to the usage, application and implementation of multi-agent systems to predict the protein structures or to support and coordinate the existing predictors, as well as it is advantages, issues, needs, and demands. It is noteworthy that there is a great need for works related to multi-agent and agent-based paradigms applied to the problem due to their excellent suitability to the problem.

INTRODUCTION

Bioinformatics consists in the study of biological problems and their intrinsic properties through the development of theoretical models and computational techniques (Chou, 2004; Gibas & Jambeck, 2001). Examples of bioinformatics studies include analysis and integration of -omics data, prediction of protein structure and function, and development of computational strategies to identify the affinity binding of drugs to a receptor and its effects. Depending on the biological question to be answered, the research in bioinformatics can be classified into two main lines: sequence analysis and structural bioinformatics.

DOI: 10.4018/978-1-5225-8903-7.ch040

The first is focused on the study and analysis of biological sequences, e.g., nucleotides and amino acid residues, over data mining and computational methods, such as sequence alignment, the inference of pathways from metabolic networks, morphometrics, and evolution (Chou, 2004). The second addresses biological questions from a three-dimensional point of view, covering most of the techniques included in computational chemistry or molecular modeling. It is related to researchers based on the three-dimensional (3-D) structures of molecules, including the computational prediction of protein/polypeptide structures, protein docking with different molecules (RNA, DNA, and proteins), simulation of dynamic behaviors of proteins, protein structure characterization and classification, and study of structure-function relationships (Chou, 2004; Xu, Xu, & Liang, 2007).

In this chapter, we provide an introduction and overview of the current state of multi-agent systems developed for study and prediction of three-dimensional structures of proteins. The study of protein structure and the prediction of their three-dimensional structures is one of the key research problems in structural bioinformatics. Over the last years, many computational methods, systems, and algorithms have been developed for the purpose of solving this complex issue. However, the problem still challenges biologists bioinformaticians, chemists, computer scientists, and mathematicians because of the complexity and high dimensionality of the protein conformational search space. Experimentally, the generation of a protein sequence is considerably easier than the determination of its 3-D structure. The 1990's GENOME projects resulted in a significant increase in the number of protein sequences. Unfortunately, the number of identified 3-D protein structures did not follow the same trend. Currently, the number of known sequences is far higher than the number of known 3-D structures; there is a large gap between the number of protein sequences we can generate and the number of new protein folds we can determine by experimental methods such as X-ray diffraction and Nuclear Magnetic Resonance (NMR). We intend to give an integrated view and insights about the protein structure prediction area concerned with the development and application of multi-agent systems to predict 3-D protein structures, as well as its advantages, issues, current needs, and demands. This chapter is useful for bioinformaticians, computer scientists, mathematicians, and biologists interested in beginning research in this field or to improve their current research.

This chapter is organized as follows. Section 2 provides some background knowledge relevant to proteins, its representation models, energy functions and molecular forces, and an overview of the protein structure prediction problem. In Section 3, prediction methods based on agents and multi-agent are introduced. The chapter concludes, and further research is outlined in Section 4.

BACKGROUND

Proteins

From a structural perspective, a protein or polypeptide is an ordered linear chain of building blocks known as amino acids. An amino acid residue is a small molecule containing an amino group (H2N+), a carboxyl group (COOH−), and a hydrogen atom attached to a central alpha carbon (Cα) (Figure. 1). Besides, each amino acid has also an R organic group (side-chain) connected to the Cα. The group R distinguishes one amino acid from another and confers the chemical properties of each amino acid residue. In nature, there are 20 distinct amino acid residues, each one with its chemical properties (Lodish et al., 1990). The side-chains of amino acids can differ in size; electric charge; and polarity. Also, depending on