Chapter 8
Supercomputers in Modeling of Biological Systems

Randall Maples
Oklahoma State University, USA

Sindhura Ramasahayam
Oklahoma State University, USA

Gerard G. Dumancas
Oklahoma Medical Research Foundation, USA

ABSTRACT

Modeling of biological systems has become an important facet in today's scientific community because it has aided in the simulation of the minute biological entities comprising a living individual. With the advent in the advances of supercomputers, most challenges in understanding the complexities of biological networks and processes occurring in the human body can now be understood. Proteins, which are large biomolecules comprised of amino acids, play a critical role in the proper functioning of a living organism, and, thus, the prediction of its structure is essential in medicine for drug design or in biotechnology, such as in the designing of novel enzymes. This chapter focuses on how supercomputers facilitate in the prediction of protein structures in its different forms, modeling of protein-ligand binding site identification, as well as in the protein-surface interactions modeling.

INTRODUCTION

Over the years, the biological sciences have evolved to a field where it is intertwined with computational sciences. This has become necessary since computational (mathematical) models may provide the basis of activity patterns exhibited by different biological phenomena and, therefore, play an important role in understanding the processes of life from a holistic point of view.

Computational modeling is a powerful approach for understanding biological systems complexity. The development of novel mathematical representations and simulation algorithms are vital for the success of modeling efforts in biological systems. For example, several successful attempts have been made for simulating complex biological

DOI: 10.4018/978-1-4666-7461-5.ch008

processes like metabolic pathways, gene regulatory networks and cell signaling pathways (Meng, Somani, & Dhar, 2004). Throughout the years, a number of diverse methods have been developed to model and visualize the biological systems. Supercomputers play a very critical role in the success of biological systems modeling. For example, in December 1999, IBM announced a five-year $100 million initiative to build a petaflop-scale supercomputer to attack problems such as protein folding. As such, the IBM Blue Gene project that utilized a massively parallel computer was initiated to use large-scale biomolecular simulation to advance the understanding of biologically important processes, specifically protein folding (Allen et al., 2001). With this project, modeling the protein folding trajectories addressed questions such as why do proteins consistently fold into specific structures and are there one or several folding pathways per protein?

Supercomputers also play a critical role in the drug design and development aspect. For example, the average cost of developing and bringing one drug to the market can range from a few hundred million dollars to more than a billion and taking from 10-15 years before patients can avail the medications they need. As such, scientists for example are using the Oak Ridge Leadership Computing Facility to speed up the screening process while increasing the chance for developing a successful drug for a fraction of the cost. In that regard they were able to create 3D biological simulations of compounds docking with receptors in the body and run it using the world's fastest computers to screen millions of drug candidates in a few days (Baudry, 2012).

In 2012, the U.S. Department of Energy administered a program named Advanced Scientific Computing Research Leadership Computing Challenge that would award up to two million hours on the Titan supercomputer at the Oak Ridge National Laboratory (ORNL) for research. One of the funded proposals in this program in 2012 was awarded to scientists at the U.S. Department of Energy (DOE) Brookhaven National Laboratory to study how protein folds into their three-dimensional shapes. The Titan supercomputer was ranked as number 2 in the Top500 Supercomputers in the world in November 2013 listing as provided in the Appendix of this book (Rutkin, 2012).

While it is possible to cover many different topics on the modeling of biological systems, this chapter will focus on the applications of supercomputers in the modeling of biological systems specifically in protein structure prediction, protein-ligand binding site identification, and protein-surface interactions.

Computational Methods for Protein Structure Prediction

The knowledge of the native protein structure could provide insights into its functions. Protein structures comprise of polymers of amino acids joined together by peptide bonds. Protein structures are classified into primary, secondary, and tertiary structures. The linear sequence of the polypeptide chain refers to its primary structure. The secondary structure is the polypeptide chain which comprises of α-helix and β-sheets or β-strands. The α-helix and β-strands are connected through coil or loop. Tertiary structure is the 3-D structure of the protein molecules in which α-helix and β-strands are folded into a compact globule.

Even though the large scale sequencing projects have generated an abundance of protein sequence data, the experimental determination of a protein structure and/or function is tedious, expensive and requires intense labor. Consequently, there is still a growing gap between proteins with experimentally derived structures and proteins with unknown structures. In order to address this, a collection of automated methods or bioinformatics' tools that utilize sophisticated supercomputers have been used in determining structures of the novel proteins from its amino acid. This chapter will present a set of bioinformatics tools that cover