Abstract

Bioinformatics houses a rich collection of research related to the protein interactome and its applications in Evolution and Disease Biology. However, exploitation of the protein interactome for determination of the protein spatial proximity information and reduction of the huge computational overhead faced by whole-cell modeling and simulation are not yet studied up to their merit. In this thesis, we address some of these issues and present several new theoretical findings, efficient algorithms, and computational aspects concerning the aforementioned research problems. We have first shown how network and graph theory-based approaches can lead to computation of the protein spatial locality, which, in turn, aids in identifying the spatially localized protein clusters in the cell of an organism. These clusters are functionally consistent as indicated by a thorough Gene Ontology-based analysis, and hence, can be termed as the functional modules of the cell. Based on the characterization of the protein clusters, we computationally partition the simulation space (whole-cell of an organism) into simulation sub-cells, each encompassing its corresponding assigned functional module. Next, we present a novel parallel whole-cell simulation framework that can optimally simulate these sub-cells on separate computing units of the High-Performance Computing systems. Such simulations are efficiently guided by our designed hashing-based data structure called the Cellular Dictionary that caches the cell state during each simulation time instant. We also explore the role of protein localization in cell towards deciphering biological evolution. To address the hypothesis of the presence of evolution traces at the level of the inter-module interactions, we present detailed topological analyses followed by some statistical and biological results supporting the hypothesis. Further, we exploit the locality information and develop a novel Deep Learning framework for identifying the putative disease-associated proteins in human. Our model has an excellent predictive capability as reflected by the evaluation metric measures and it presents potential disease-associated human proteins eligible for experimental validations. Experimental results and detailed analyses have been furnished to demonstrate the usefulness and efficiency of the proposed techniques.