

REVEALING THE CONFORMATION AND PROPERTIES OF HUMAN GENOME, PROTEIN MOLECULES AND PROTEIN DOMAIN CO-OCCURRENCE NETWORK

Zheng Wang

Dr. Jianlin Cheng, Dissertation Supervisor

ABSTRACT

Deoxyribonucleic acid, or DNA, encodes genetic instructions for the functionalities of organisms. For human beings, 23 pairs of chromosomes, containing DNA strands, form a globule structure in the nucleus. This chromosomal conformation influences the subsequent biological processes including transcription and translation by positioning sequentially remote genes spatially close. Chapter 2 reveals human chromosomal conformation and studies gene-gene interactions and “transcription factor binding site” interactions based on chromosomal spatial proximity.

Proteins are the biological units that conduct biological functions. The three-dimensional structure of a protein molecule determines its particular functions. Chapters 3 and 4 discuss research in predicting protein tertiary structures. Algorithms that can predict residue-specific qualities of predicted structures were constructed and benchmarked. A knowledge database of soybean transcription factors is presented in Chapter 5, which contains predicted protein tertiary structures. Chapter 6 shows a computer system predicting protein functions using profile-sequence alignment, profile-profile alignment, and protein domain co-occurrence network.

A biological process is usually performed by multiple proteins. Biological network provides a global perspective of studying lives, which usually considers the entire set of the same type of biological molecules of the target organism. Chapter 7 introduces a novel biological network, protein Domain Co-occurrence Network (DCN), and demonstrates that DCN has great potentials in inferring species phylogenies and predicting protein functions.