

Chapter 65

Extraction of Protein Sequence Motif Information using Bio-Inspired Computing

Gowri Rajasekaran

Periyar University, India

Rathipriya R

Periyar University, India

ABSTRACT

Nowadays there are many people affected by the genetic disorder, hereditary diseases, etc. The protein complexes and their functions are detected, in order to find the irregularity in the gene expression. In a group of related proteins, there exist some conserved sequence patterns (motifs) either functionally or structurally similar. The main objective of this work is to find the motif information from the given protein sequence dataset. The functionalities of the proteins are ideally found from their motif information. Clustering approach is a main data mining technique. Besides the clustering approach, the biclustering is also used in many Bioinformatics related research works. The PSO K-Means clustering and biclustering approach is proposed in this work to extract the motif information. The Motif is extracted based on the structure homogeneity of the protein sequence. In this work, the clusters and biclusters are compared based on homogeneity and motif information extracted. This study shows that biclustering approach yields better result than the clustering approach.

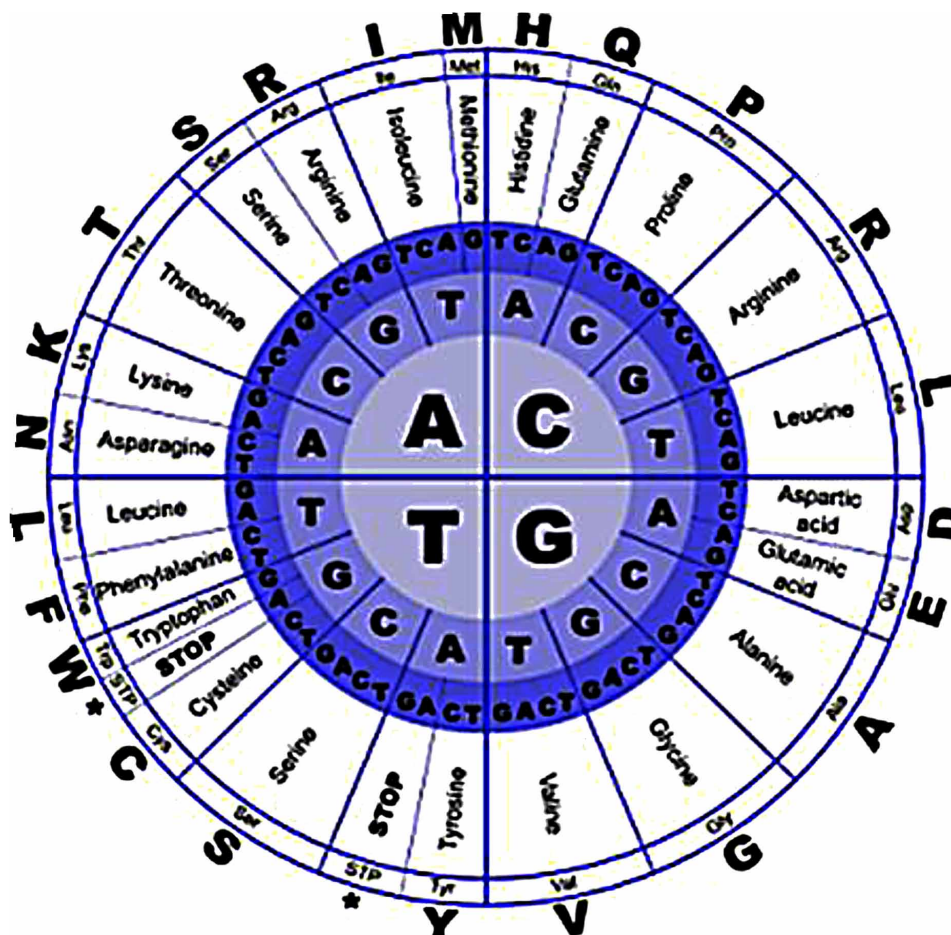
INTRODUCTION

Protein Sequence

Proteins (Vincent, Bernard, & SinanKockara, n.d.) are present in every cell of the organisms. They are involved virtually almost in all cellular activities. They are responsible for the various metabolic activities, nutrition transportation, regulations and etc. They exist as single chain molecule, as a three dimensional

DOI: 10.4018/978-1-7998-1204-3.ch065

Figure 1. The IUPAC code for PROTEIN from DNA code



structures or even in the bundle or complex forms. The protein plays a vital role in cellular processes. The protein consists of twenty amino acids. They possess different characteristics such as hydrophobic, hydrophilic, polar, non-polar and etc. It is the great challenge to the bioinformatics that to find which combination of proteins are responsible for what kind of activities. The structure and function discovery of proteins in living organisms is vital role in understanding the background of various cellular processes. It is helpful in treating various diseases, in detecting the drugs to peculiar diseases.

The irregularity in the proteins can be found if their actual functionality and their protein complex are known. The biologist will interpret the functionality of the protein complex based on their chemical properties.

These significant protein complexes of various organisms are discovered using the computational techniques. Protein complexes are an assembly of proteins that build up some cellular machinery; commonly spans a dense Sub-network of proteins in a protein interaction network. The gene expression would not help much in detecting the functionalities of the gene. Protein sequence can be generated from the DNA/mRNA sequence that codes for the protein are shown in the Figure 1. The four basic amino acid combines to form the protein that is shown. The 20 different amino acids that are present in the protein sequence along with their chemical name are listed in the Table 1.

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/extraction-of-protein-sequence-motif-information-using-bio-inspired-computing/243168

Related Content

Real Time Analysis Based on Intelligent Applications of Big Data and IoT in Smart Health Care Systems

Mamata Rath (2020). *Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications* (pp. 1839-1857).

www.irma-international.org/chapter/real-time-analysis-based-on-intelligent-applications-of-big-data-and-iot-in-smart-health-care-systems/243197

Proximate Breast Cancer Factors Using Data Mining Classification Techniques

Alice Constance Mensahand Isaac Ofori Asare (2019). *International Journal of Big Data and Analytics in Healthcare* (pp. 47-56).

www.irma-international.org/article/proximate-breast-cancer-factors-using-data-mining-classification-techniques/232335

EMG-Based Mobile Assessment System for Neck and Shoulder Fatigue

Pei Lun Lai, Hsiu-Sen Chiangand Qi-An Huang (2017). *International Journal of Big Data and Analytics in Healthcare* (pp. 39-50).

www.irma-international.org/article/emg-based-mobile-assessment-system-for-neck-and-shoulder-fatigue/204447

Regional Structure and Economic Development: Growth Empirics for U.S. Metropolitan Areas

David S. Bieri (2014). *Econometric Methods for Analyzing Economic Development* (pp. 153-176).

www.irma-international.org/chapter/regional-structure-and-economic-development/79698

Effects of Public Capital Investments on the Productivity of the United States, 1992-2022

Brian Slobodaand Yaya Sissoko (2024). *Data Envelopment Analysis (DEA) Methods for Maximizing Efficiency* (pp. 1-17).

www.irma-international.org/chapter/effects-of-public-capital-investments-on-the-productivity-of-the-united-states-1992-2022/336938