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

Figure 1. The IUPAC code for PROTEIN from DNA code
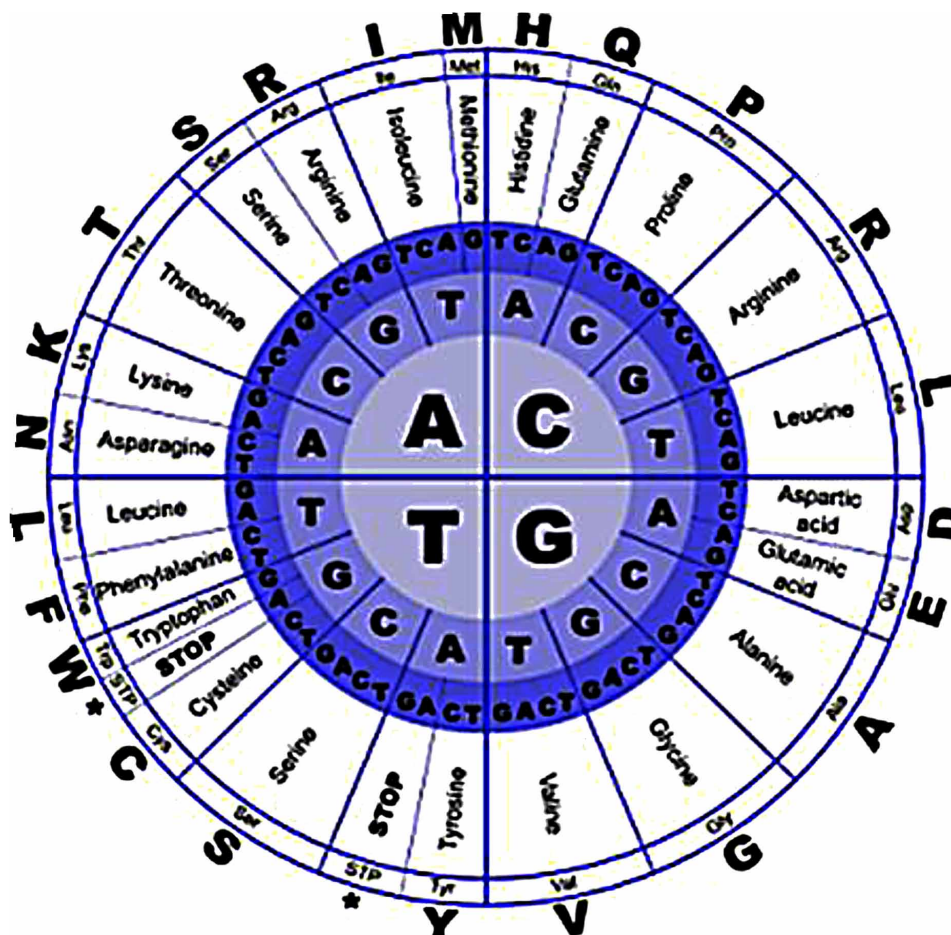
structures or even in the bundle or complex forms. The protein plays a vital role in cellular processes. The protein consists of twenty amino acids. They possess different characteristics such as hydrophobic, hydrophilic, polar, non-polar and etc. It is the great challenge to the bioinformatics that 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.

Related Content

Prediction Length of Stay with Neural Network Trained by Particle Swarm Optimization
Azadeh Oliyaeiand Zahra Aghababaee (2017). *International Journal of Big Data and Analytics in Healthcare (pp. 21-38).*
www.irma-international.org/article/prediction-length-of-stay-with-neural-network-trained-by-particle-swarm-optimization/204446

A Hybrid AHP-ELECTRE I Multicriteria Model for Performance Assessment and Team Selection
Ikram Khatrouch, Lyes Kermad, Abderrahman el Mhamediand Younes Boujelbene (2017). *Organizational Productivity and Performance Measurements Using Predictive Modeling and Analytics (pp. 115-127).*
www.irma-international.org/chapter/a-hybrid-ahp-electre-i-multicriteria-model-for-performance-assessment-and-team-selection/166518

Towards Clinical and Operational Efficiency through Healthcare Process Analytics
Vassiliki Koufi, Flora Malamateniouand George Vassilacopoulos (2016). *International Journal of Big Data and Analytics in Healthcare (pp. 1-17).*
www.irma-international.org/article/towards-clinical-and-operational-efficiency-through-healthcare-process-analytics/171401

Adaptive Edge Detection Method towards Features Extraction from Diverse Medical Imaging Technologies
Indra Kanta Maitraand Samir Kumar Bandhyopadhyaay (2017). *Intelligent Multidimensional Data Clustering and Analysis (pp. 159-192).*
www.irma-international.org/chapter/adaptive-edge-detection-method-towards-features-extraction-from-diverse-medical-imaging-technologies/172555

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