International Journal of Recent Development in Computer Technology & Software Applications
Vol. 2, Issue 2 - 2018

A NOVEL AND RELIABLE PROMOTER, PROTEIN CODING AND PROTEIN STRUCTURE PREDICTION USING ARTIFICIAL IMMUNE SYSTEM BASED 116-NEIGHBORHOOD HYBRID CELLULAR AUTOMATA

P KIRAN SREE*

The development of a potential classifier towards addressing Promoter, Protein Coding and Protein Structure Prediction was identified and the plan to achieve it was formulated. This research creates an intuition, how interrelated and logical problems in Bioinformatics like protein coding regions, promoter regions, protein structure prediction etc. can be addressed with a common framework.

The proposed approach aims at developing four independent classifiers which can perform their operations individually and also in a combined form. These classifiers AIS-MACA series uses the basic framework of Cellular Automata (CA) and features of the modified CLONAL classifier like self monitoring and non uniformity which is potential, versatile and robust. We use 16-Neighborhood Hybrid Cellular Automata to design and develop our classifiers.

SIGNIFICANCE OF RESEARCH

- Identifying the protein coding region plays a vital role in understanding these genes.
- If we identify the promoter region, we can extract information regarding gene expression patterns, cell specificity and development.
- Identifying the quaternary protein structure helps in drug design.

SPECIFIC OBJECTIVES

1. **Accuracy:** To develop a system, this can predict all the three (Protein Coding, Promoter and Protein Structure) with an average accuracy more than 94.6%.
2. **Sequence Lengths**: The DNA sequences to be processed by the classifier will be in the lengths of 54, 108, 162, 252, 354, 567, 1026 with a combination of four nucleotides (A, G, C, and T). So the proposed classifier has to handle different lengths of inputs. In the case of predicting the quaternary structure of protein, the classifier has to take the amino acid sequence as input which is a variation of twenty alphabets.
3. **Boundary Reporting:** Due to the limited information available, it is difficult to predict the coding regions from random and report the exact boundaries.
4. **Large Data Sets:** The proposed classifier for predicting protein and promoter regions has to process large data sets for training and testing to attain better accuracy. The system needs to handle more than one crore data sets to achieve that high accuracy and acceptability.

*Professor, Dept of CSE, UCEK, JNTUK. ***Correspondence E-mail Id:*** editor@eurekajournals.com

 ISSN: 2581-6276

5. Identifying the structure of the genes can extract lots of information regarding human body than can be obtained from a sequence. It is very easy to find the similarities at a higher level by predicting structure. Identifying the quaternary protein structure helps in drug design.
6. The identification of protein coding regions plays a vital role in understanding the genes. If we analyze the protein coding regions, we can extract lots of information like what is the disease causing gene, whether it is inherited from father or mother, how can we regulate the disease growth, how one cell is going to control another cell.
7. If we identify the promoter region, we can extract information regarding gene expression patterns, cell specificity and development. Promoters will regulate a gene expression. Some of the genetic diseases which are associated with the absence of promoters are asthma, beta thalassemia, rubinstein-taybi syndrome, cancer and endocrine diseases.

 ISSN: 2581-6276