Chapter 5
Development of Novel Multi-Objective Based Model for Protein Structural Class Prediction

Bishnupriya Panda
Siksha O Anusandhan University, India

Babita Majhi
G.G Viswavidyalaya, India

ABSTRACT

Protein folding has played a vital role in rational drug design, pharmacology and many other applications. The knowledge of protein structural class provides useful information towards the determination of protein structure. The exponential growth of newly discovered protein sequences by different scientific communities has made a large gap between the number of sequence-known and the number of structure-known proteins. Accurate determination of protein structural class using a suitable computational method has been a challenging problem in protein science. This chapter is based on the concept of Chou's pseudo amino acid composition feature representation method. Thus the sample of a protein is represented by a set of discrete components which incorporate both the sequence order and the length effect. On the basis of such a statistical framework a low complexity functional link artificial neural network and a complex novel hybrid model using radial basis function neural network and multi-objective algorithm based classifier are introduced to predict protein structural class.

INTRODUCTION

Proteins play a significant role in every biological process. Proteins are large molecules consisting of amino acids which the cells need to function properly. The human body is made up of approximately 100 trillion cells - each one has a specific function. Each cell has thousands of different proteins, which together make the cell do its job - the proteins are tiny machines within the cell.

DOI: 10.4018/978-1-5225-0427-6.ch005

Amino acids are the basic building block of all proteins. All amino acids have the same basic structure an amino group, a carboxyl group and a hydrogen atom but differ due to the presence of a side-chain. This side-chain varies dramatically between amino acids, from a simple hydrogen atom in the amino acid glycine to a complex structure found in tryptophan. Hydrophilicity or Hydrophobicity is determined by the side chain.

The function of a protein is based on its structure. Knowledge of protein structure plays a significant role in molecular biology, cell biology, pharmacology and medical science. However, despite years of both experimental and theoretical study, protein structure determination remains one of the most difficult tasks in proteomics. With exponential growth of protein database, experimental determination of structure is not cost effective. Therefore, development of a robust computational model is highly significant.

Accurate determination of protein structural class is a two-step process: Effective representation of protein sequence and then developing a prediction model. Many in-sillico structural class prediction algorithm and methods have been proposed earlier. Amino Acid Composition (AAC) is highly related to protein structural class (Chou, 1995). Several classification methods such as distance classifier, principal component analysis (Du & Jiang, 2006), Bayesian classifier, fuzzy clustering (Ding & Zhang, 2007), support vector machine (Cai, Liu, Xu &Zhou, 2001) and multilayer artificial neural network (Cai & Zhou, 2000) have been proposed in the literature. Though many promising results have been achieved, AAC of protein lacks sequence order and sequence length information. Sequence order and sequence length information also play a significant role in predicting protein structural class because amino acid composition do not differentiate between protein molecules of different sequence order and sequence length. This chapter along with pseudo amino acid composition, amphiphillic correlation factors of protein molecule (Panda, Mishra, Majhi &Rout, 2013) and the spectral characteristics of the protein (Sahu & Panda, 2010) have been used to capture the sequence order information.

Classification is a supervised method where the class label is known and we train the model by using that class label. Classification has two steps. The first one is model construction and the second process is model usage. Model construction describes a set of predetermined classes. Model usage is for classifying future or unknown objects.

Many authors have proposed neural network as a good candidate for classification of protein structural class. But how to choose the number of layers and number of neurons in each layer to enhance the classification accuracy is highly complex problem.

SIGNIFICANCE OF PROTEIN SECONDARY STRUCTURE

The function of a protein is based on its structure. Knowledge of protein structure plays a significant role in molecular biology, cell biology, pharmacology and medical science. However, despite years of both experimental and theoretical study, the determination of protein structure remains one of the most difficult problems.

Protein secondary structure prediction plays a vital role due to following reasons:

1. Experimental determination of structure is not cost effective.
2. It helps in predicting the 3D structure.