Association Rule Mining of Protein Contact Maps from Protein 3D Structures

Dr K Suvarna Vani and K Praveen kumar

***Abstract*—Contact maps have been used in *ab initio* methods for the problem of protein structure prediction problem. Secondary structures and contacts made by the residues are clearly visible in the contact maps where helices are seen as thick bands and the beta sheets are seen as orthogonal to the diagonal. This paper explores the idea of extracting rules from contact maps to represent "protein fold" information. Association Rule Mining is carried out on the data set of All Alpha proteins in order to extract rules to predict EF-hand like fold and All Beta proteins to predict SH3 fold. The alpha helix regions are predicted with sensitivity of 94% and positive predictive value of over 85%. A novel method called Triangle Subdivision Method (TSM) is proposed to extract clustering patterns present in the off-diagonal interactions of the contact map to find surprising conserved patterns in the protein contact map. Rules based on the diagonal information and the off-diagonal patterns present in the contact map are extracted in order to predict EF-hand like fold with accuracy of 94% and SH3 fold with accuracy of 60%. The paper proposes a new method called TSM to extract conserved patterns from protein contact maps. The promising results of the current work shows that the methodology can be potentially applied for "fold prediction" in general.**

***Index Terms*—Proteins, Contact Maps, EF-Hand Fold, SH3-Fold.**