

# Non-linear Abstractions of Protein Folding Patterns

Parthan Kasarapu, Arun Konagurthu, Maria Garcia de la Banda

Simplified representations of protein folding patterns serve as a starting point for visualizing, rationalizing, comparing, classifying, searching and aligning large volumes of structural data. Traditionally, this simplification is performed at the level of secondary structures of proteins - helices and strands of sheet. Such representations fail to capture the entire essence of protein folding patterns resulting in significant loss of information. We present a segmentation method that generates biologically meaningful abstractions of protein folding patterns to support accurate large-scale analyses involving protein structures. This work investigates the description of folding pattern using a class of parametric curves, which are used to describe non-overlapping regions of protein structure. The approach to this problem is based on information theory and uses the statistical and inductive inference framework of minimum message length (MML).

