P87

PRIGSA: Protein Repeat Identification by Graph Spectral Analysis.

Broto Chakrabarty and Nita Parekh

Centre for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, India

Abstract

Repetition of a structural motif within protein is associated with a wide range of structural and functional roles. In most cases the repeating units are well conserved at the structural level while at the sequence level, they are mostly undetectable suggesting the need for structure-based methods. Since most known methods require a training dataset, *de novo* approach is desirable. Here we propose an efficient graph based approach for detecting structural repeats in proteins. In a protein structure represented as a graph, interactions between inter- and intra-repeat units are well captured by the eigen spectra of *adjacency matrix* of the graph. These conserved interactions give rise to similar connections and a unique profile of the principal eigen spectra for each repeating unit. The efficacy of the approach is shown on eight repeat families annotated in UniProt, comprising of both solenoid and non-solenoid repeats with varied secondary structure architecture and repeat lengths. The performance of the approach is also tested on other known benchmark datasets and the performance compared with two repeat identification methods. For a known repeat type, the algorithm also identifies the type of repeat present in the protein. A web tool implementing the algorithm is available at the URL: http://bioinf.iiit.ac.in/PRIGSA/