Automated Annotation of Protein Complexes with Domain Correlation

Ya Zhang, Xue-wen Chen, Information and Telecommunication Technology Center, University of Kansas, Lawrence, KS 66045. 2006 Annual Kansas City Area Life Sciences Research Day.

High throughput experiments such as HMS-PCI (High throughput Mass Spectrometric Protein Complexes Identification) and TAP-MS (Tandem Affinity Purification and Mass Spectrometry) have produced a large amount of protein complex data. However, the high throughput data are inherently noisy. The lack of overlap among independent data sets suggests that some experiments may only capture a partial protein complex. Moreover, the biological functions of many high-throughput protein complexes are unknown. Two important tasks are to computationally identify potential member of the complexes and predict the functions of the protein complexes. Previous efforts in analyzing protein complex data were largely based on the protein constitution of the complexes, such as discovering protein functional modules. However, studying complexes at protein level does not always reveal the function linkage among complexes. In some cases, protein complexes of the same function may not share any protein at all. Examples of such protein complexes include the cytoplasmic ribosomal large subunit and the mitochondrial ribosomal large subunit, both of which function for protein synthesis. On the other hand, when the two protein complexes are compared at domains level, 14 shared domains are identified. This fact clear suggests the importance of the domain composition in understanding the function of complexes. In this research, we study the domain composition of protein complexes to predict the functions of protein complexes. The underlying rationale behind this study is that the biological function of a complex is determined by its domain composition. If a set of domains are extensively shared among protein complexes, then the association among the set of domains are likely to reflect the functions of the complexes. We define such set of domains as domain function modules. The domain modules are treated as elementary functional units and indicators of the functions of the corresponding protein complexes.