Conversation of Intrinsic Disorder in Protein Domains and Families

Conversation of Intrinsic Disorder in Protein Domains and Families

Chen, Jessica Walton

Name: JChen_Thesis_FINAL.pdf
Size: 831.1Kb
Format: PDF

View/Open

Permanent Link: http://hdl.handle.net/1805/331
Date: 2005-08
Keywords: disorder ; protein domains

Abstract:

Protein regions which lack a fixed structure are called ‘disordered’. These intrinsically disordered regions are not only very common in many proteins, they are also crucial to the function of many proteins, especially proteins involved in signaling and regulation. The goal of this work was to identify the prevalence, characteristics, and functions of conserved disordered regions within protein domains and families. A database was created to store the amino acid sequences of nearly one million proteins and their domain matches from the InterPro database, a resource integrating eight different protein family and domain databases. Disorder prediction was performed on these protein sequences. Regions of sequence corresponding to domains were aligned using a multiple sequence alignment tool. From this initial information, regions of conserved predicted disorder were found within the domains. The methodology for this search consisted of finding regions of consecutive positions in the multiple sequence alignments in which a 90% or more of the sequences were predicted to be disordered. This procedure was constrained to find such regions of conserved disorder prediction that were at least 20 amino acids in length. The results of this work were 3,653 regions of conserved disorder prediction, found within 2,898 distinct InterPro entries. Most regions of conserved predicted disorder detected were short, with less than 10% of those found exceeding 30 residues in length. Regions of conserved disorder prediction were found in protein domains from all available InterPro member databases, although with varying frequency. Regions of conserved disorder prediction were found in proteins from all kingdoms of life, including viruses. However, domains found in eukaryotes and viruses contained a higher proportion of long regions of conserved disorder than did domains found in bacteria and archaea. In both this work and previous work, eukaryotes had on the order of ten times more proteins containing long disordered regions than

did archaea and bacteria. Sequence conservation in regions of conserved disorder varied, but was on average slightly lower than in regions of conserved order. Both this work and previous work indicate that in some cases, disordered regions evolve faster, in others they evolve slower, and in the rest they evolve at roughly the same rate. A variety of functions were found to be associated with domains containing conserved disorder. The most common were DNA/RNA binding, and protein binding. Many ribosomal protein families also were found to contain conserved disordered regions. Other functions identified included membrane translocation and amino acid storage for germination. Due to limitations of current knowledge as well as the methodology used for this work, it was not determined whether or not these functions were directly associated with the predicted disordered region. However, the functions associated with conserved disorder in this work are in agreement with the functions found in other studies to correlate to disordered regions. This work has shown that intrinsic disorder may be more common in bacterial and archaeal proteins than previously thought, but this disorder is likely to be used for different purposes than in eukaryotic proteins, as well as occurring in shorter stretches of protein. Regions of predicted disorder were found to be conserved within a large number of protein families and domains. Although many think of such conserved domains as being ordered, in fact a significant number of them contain regions of disorder that are likely to be crucial to their function.

Description:

Submitted to the faculty of the Bioinformatics Graduate Program in partial fulfillment of the requirements for the degree Master of Science in the School of Informatics, Indiana University August 2005

This item appears in the following Collection(s)

- Informatics Graduate Theses and PhD Dissertations (/handle/1805/303)
- Informatics School Theses and Dissertations (/handle/1805/954)

 Show Statistical Information (#)

My Account

- Login
- Register

Statistics

- Most Popular Items
- Statistics by Country
- Most Popular Authors