**Literature Survey**

The identification of domains was initially done manually by human experts by examining the the protein structures. The first reported study is by Wetlaufer in 1973 [1] by constructing three-dimensional peptide chain models which was facilitated by the then available x-ray structures and stereoscopic images. He observed distinct structural “regions” which were contiguous as well as non-contiguous and defined these “regions” as a section of peptide chain that can be enclosed in a compact volume. This laid down the foundation of structural domains and their identification, which has been a topic of scientific research for over 40 years now.

While manual curation is an accurate and reliable approach, it lacks is consistency and speed. With an observed increase in the number of solved structures, automated methods have been proposed for domain identification. These methods are fast, systematic and takes into account various properties possessed by structural domains. A myriad of algorithms have been proposed with each having it’s own pros and cons, a common observation is that almost all of them employ a hierarchical approach. An extensive review of domain identification algorithms is given by Stella et al (2009) [2]. Below we briefly discuss some of the important domain identification approaches proposed till date.

References

1: Wetlaufer, D.B., (1973), Nucleation: Rapid Folding, and Globular Intrachain Regions in

proteins. Proceedings of the National Academy of Sciences, U S A, 70 (3), 697-701.

2: Veretnik S, Gu J, Wodak S: Identifying Structural Domains in Proteins. In In Genny Gu and Philip Bourne Structural Bioinformatics. Second edition. Wiley-Blackwell; 2009:485–513.