**Literature Survey**

The task of domain identification can be done manually by human experts who have prior knowledge on protein structures and domains. The first manual inspection was done 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 a sophisticated and an accurate approach, what it lacks is consistency and speed. With an observed increase in the number of solved structures, automated methods have emerged to tackle the problem of domain identification. These methods are fast, systematic and takes into account various properties possessed by structural domains. A myriad of algorithms have been devised with each having it’s own pros and cons, a common observation is that almost all of them employ a hierarchical approach. Stella et al.[2] did an extensive study on the all the major techniques which came up until the first decade of 21st century. We’ll discuss some of those while extending her work to incorporate the recent advances in the following sections.

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