Abstract

A protein domain is a conserved part of a given protein sequence that can evolve, function, and exist independently of the rest of the protein chain. Each domain forms a compact three-dimensional structure and often can be independently stable and folded. Study the effects of changes in the genome on protein domain might also help to identify the factors cause of cancer. The number of human genes and proteins is high and relation between proteins and their domains is complex and recognizing the major causes of cancer by laboratory methods is also very difficult. All of these make it necessary to use computer science and statistics in the study of cancer. In this thesis the problem of finding protein domains involved in the formation and development of cancer has been studied. In this approach initially the probability of mutation on the genome has been calculated by a proposed method. In the next step by setting a threshold, candidate domains involved in each cancer has been extracted and based on these candidate domains, candidate stem cells and repair genes for each type of cancer has been discussed.

Keywords: *Protein Domain, Classification, Pfam, CATH, Cancer, Stem Cell, Repair Gene, Mitochondria, Pan Cancer, Computational models.*