RESEARCH PAPER

Asian Journal of Bio Science, Vol. 4 Issue 2 : 274-277 (October, 2009 to March, 2010)

Functional annotation of histone proteins in human

MANOJ KUMAR SHARMA[1], ASTHA DUHAN[2] AND M.L. SANGWAN[3]

[1]Department of Bio-informactics, J.V. College, Baraut, BAGHPAT (U.P.) INDIA

[2]Department of Bio-informatics, Mata GujriCollege, FATEHGARHSAHEB (PUNJAB) INDIA

[3]Department of Animal Biotechnology, C.C.S. Hissar Agricultural University, HISSAR (HARYANA) INDIA

(Accepted : September, 2009)

Proteins are very much useful components and their functional annotation is also very useful as well. The function of protein is the measure of the expression of that particular protein. By knowing the function of one protein it can be found out the function of that protein also which has conserved region of the above protein whose function is known. PANDORA is a web based tool to aid biologist in interpretation of protein sets without the need of examining each individual protein. The general approach that PANDORA uses is based on annotation. In PANDORA, annotations are treated as binary properties that can be assigned to proteins. In relation with histone protein family we find out not only functional annotation but the evolutionary relationship with the family members of histone protein. PANDORA gives results for histone protein family. It provides more white the nodes which mean the sensitivity is higher, that are close to 1, reflect the result that fraction of the proteins with annotation. Specificity provides the data that is always more than 0, that gives the result that fraction of protein set has annotation.

Key words : PANDORA, Histone

INTRODUCTION

Proteins are complex nitrogenous organic biopolymers of amino acids showing great diversity in their organization and they are of prime biological importance. Proteins are the most complex chemicals synthesized in nature and must fold into complicated three-dimensional structures to become active. Family and super family classification also serves as the basis for rule-based procedures that provide rich automatic functional annotation among homologous sequences and perform integrity checks. Patterns or profiles, numerous rules have been defined to predict position-specific sequence features such as active sites, binding sites, modification sites, and sequence motifs. Linking protein data to more bibliographic data that describes or characterizes the proteins is crucial for increasing the amount of experimental information and improving the quality of protein annotation. The annotation of function by transference from proteins of related sequences is not the only possibility for the "in silico" prediction of function. The flourishing of genomic data has enabled other modes of function prediction independent of the identification of homologous sequences. The function of proteins can be inferred from the study of the similarity of their expression pattern with properties of a system can be explained by but not deduced from its components (such as protein domains).

Biological reality actually indicates just the opposite; the presumption that fold similarities alone are sufficient to identify functional similarity is discredited in numerous cases (Koppensteiner Devos D, Koppensteiner, Skolnick, Karplus , Tramontano, Fischer, Kolinski ,Rost, Flockner , Jones, Kelley, Rychlewski, Skolnick, Valencia A., 2000). Methods of annotation on the basis of sequence similarity (such as BLAST (Smith TF, Zhang X.,1997)) or sequence motifs (such as Blocks, PRINTS, Pfam, and Prosite) have proven successful, they are limited by implicit assumptions underlying their methodology. A number of new sequence analysis challenges have emerged in the genome era. Predicting the function of each newly found protein has been a main focus of genome analysis (Bork P, Dandekar T, Diaz-Lazcoz Y, Eisenhaber F, Huynen M, Yuan Y., 1998). A general approach for functional characterization of unknown proteins is to infer protein functions based on sequence similarity to annotated proteins in sequence databases. This complex and ambiguous process is inevitably error prone (Bork and Koonin, 1998).

Histones are highly conserved proteins that serve as the structural scaffold for the organization of nuclear DNA into chromatin. The histones have an amino terminal tail, a globular domain, and a carboxy-terminal tail. Histone H1, the most common form of linker histone, binds to nucleosomal DNA at the point from which the DNA exits the nucleosome, and is required for higher order packing of chromatin. The four core histones, H2A, H2B, H3 and H4 assemble into the octamer (2