© 2016 IJSRST | Volume 2 | Issue 4 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Bioinformatics Tools for Protein Analysis

Mahin Ghorbani[1]*, Fatemeh Ghorbani[2], Hamed Karimi[3]

[1]Department of Biotechnology, Fergusson College, F.C. Road, Pune, Maharashtra, India
[2]Centre for continuing education, John Abbott College, Montreal, Quebec, Canada
[3]Department of computer engineering, Azad university of Farsan, Chaharmahal va bakhtiari, Iran

ABSTRACT

Understanding of the relationship between amino acid sequence and three dimensional structure of protein is one of the main objects of bioinformatics, if the relationship is known; it can be helpful for prediction of the protein structure from its amino acid sequence. In this paper, you will learn the fundamentals of structures of proteins, and we will discuss the basic tools and databases for protein visualization and classification. On the basis of such reviews and discussions, you will be able to understand how to predict protein structure and function from its amino acid sequence. The protein structure databases discussed in this paper are such as Protein Data Bank, NCBI Structure Database (MMDB). The Protein structure visualization databases and tools discussed here are such as Cn3-D, Chemscape Chine, Rasmol and Protein Explorer, SWISS PDB Viewer, Mage and Kinemages, and PDBsum. The protein structure alignment discussed are such as VAST and DALI Server. The Domain Architecture databases discussed here are CDD, CDART. Discussed tools for plotting protein-ligand interaction is Ligplot and the protein classification approaches are SCOP and CATH.

Keywords: Bioinformatics Tools, Protein Prediction, Protein Visualization, Protein Classification, Protein Function.

I. INTRODUCTION

A. Protein Structure Overview

Proteins are biomolecules made up chains of 20 different amino acid residues.For proper performance of particular biochemical function, a certain number of amino acid residues are essential and appearance of lower limit for a functional domain size is around 40-50-residues. The range for size of the protein is from this lower limit to several hundred residues in multi-functional proteins. Proteins normally consist of thousands of atoms arranged in a 3-dimentional structure that is specific for each type of protein. While a protein is formed,it folds itself into a complicated 3-dimentional shape. Each protein consist of one folded shape, and regularly folds into it, normally in less than one second.That complex folded shape directs how the protein functions and also how it performs its interaction with others molecules. For making up of each protein, a gene in the DNA of living cells should code the particular sequence of amino acid. Synthesis of protein is not done without its mRNA being present but persistence of a protein is possible in the cell when its mRNA is no longer present. However the abundance of mRNA may be present but there is no message translation into protein unless genetically required. There is thus no proper correlation between protein and its mRNA in a living cell at any given time. Protein synthesis is a very complex process. Ribosomes employed as protein factories, RNA forms bridge in ribosomes and work as support structures as well as protein forming machinery.

There are four level of organization during protein structure analysis: primary structure is the chain of amino acid sequence of the protein. Every protein owns a unique amino acid sequence. Secondary structure :spatial arrangement of protein framework without its side chain conformation. Tertiary structure is three dimentional structre of entire protein. Quaternary structure refers to 3-D structures of proteins that consist of two or more polypeptide chains, referred as sub-units. In primary structure of protein, the sequence is held together by covalent peptide bond, in 3-D structure of

protein, the structure held together by variety of bonds such as: ionic bonds, hydrogen bonds, covalent bonds and hydrophobic interactions. Proteins and their structures are made up of polypeptides and domains(motifs). A short conserved region in a protein is called as motif and domain refers to a distinct portion of a protein estimate to fold independently of the rest of the protein and has its own function.

In primary structure of proteins, amino acids are linked together via the peptide bonds that are made up of the reaction of alpha –carboxyl group of one amino acid with alpha –amino group of another amino acid. For characterization of a protein its unique amino acid is an important part. For obtaining information about primary structure of protein we can use PROWL. This site will provide you with information about amino acid properties, their allowed bond geometry, probability for being interior residue versus exterior residues and which amino acid substitution are likely to maintain function and structure. This site also gives other information required for prediction of protein structure and function based on primary structure.

In secondary structure variety of structural elements are present such as alpha-helix, Beta sheet and random coil. In addition loops and turns, folds and motifs are present. Each of these structures can be predicted using different software which will be discussed in the next step.

In Tertiary structure, the arrangement of the polypeptide in 3-D far from its linear sequence is represented. This arrangement is as result of interactions between R-groups via van der waals, hydrogen and hydrophobic, ionic bonding. Motifs and domains are appeared in this structure level. Some examples are: Hairpin Beta-motif,Greek Key motif, Beta-alpha-beta motif.

Domains also appear in tertiary structure level. Polypeptide chains are more than 200 amino acid in length and fold into two or more compact globular cluster are known as domain. usually there are three types of domains: alpha domain, Beta domain and alpha/Beta domains.In quaternary structure, those proteins with more than one polypeptide chains are considered like haemoglobin.

Classification of protein based on structure: There are four types for classification of protein based on the structure such as type 1: alpha (Predominantly or core, exclusively alpha helices) example: Bundle and non-bundle, Beta(predominantly or core, primarily beta strand) example: Roll Barrel,Sandwich, single sheet, Alpha/Beta (Predominantly alternating Alpha-helix and beta-stand) example :Beta Alpha Beta motif, Alpha and Beta (Alpha helices and beta strand regions as separate grouping) example : Anti –parallel Beta sheet.[1-6].

II. METHODS AND MATERIAL

A. Protein structure Databases

Protein Data Bank (PDB) : PDB is a very large universal storage place of processing and distribution of 3-dimensional structure data of macromolecules. the information in PDB derived from variety tools and experiments like NMR, X-ray crystallography, microscopy, cryoelectron and theoretical modeling,. Accommodations of the database for users are access to structural data, providing methods for visualizing the structure and downloading structural information.[7]

NCBI Structure Database (MMDB): It includes database of 3D structure of biomolecules which experimentally determined.Most of these data derived from X-ray crystallography and NMR spectroscopy. The database provide biologists with a broad information on biological functions of proteins, on mechanisms related to their functions and on relationship between biomolecules and their evolutionary history.Additionally this database provide biologists with comparative analysis of 3D structure of proteins. NCBI also called as MMDB (molecular modeling database) and includes 3D structure of macromolecules and visualization tools for comparative analysis of proteins.[8]

Database and tools for protein structure visualization:
Cn3-D : "see in 3-D" is a viewer of structural sequence alignment for MMDB database. It facilitates viewing of 3-Dstructure and alignment of sequence –structure of structure-structure. It serves as a helper application for the browser. Files can be downloaded to the pc and the application can be launched.[9]

SWISS PDB Viewer:

It facilitates and network for analysis of several proteins simultaneously. The proteins lay over each other in order to analyze structural alignment and provide comparison of their active sites, their amino acid mutations angles, distances and H bonds between their atoms. This viewer is joined to Swiss-Model server. [10]
Chemscape Chime, Rasmol and protein explorer:

This tool is one of the usual tools for visualization of protein structure.It can read molecular structure files from PDB. Chemscape chime serves as a plug in to permit structure visualization with browser. Protein explorer serves as a plug in to permit viewing of protein structure with our browser. Both of these application namely Chemscape chime and protein explorer are primary derivation of Rasmol.[11]

Mage and Kinemages:

It is another tool for protein structure visualization. It is able for rotation of entire image in real time, displaying of parts by turning off and on them, selection of points for their identification and animation of change between different forms.[6]

PDBsum :
It is a database that facilitates a large illustrated graphic summary of the main information on each biomolecular structure from the protein data bank. It consists of images of structure, detailed structural analysis derived from PROMOTIF program, schematic graphs of interactions, summary PROCHEK results [12]

Protein structure alignment tools:

VAST (vector alignment sequence tool): it is a tool produced by NCBI and provides identification of similar proteins with 3D structure. So it is structure similarity and search service. [13].

DALI : It is an computational protein structure alignment tool used for comparison of protein structure in 3D.[14]

B: Domain architecture Database:

Conserved Domain Database :(CDD) : is a database contain sequence alignment and profiles, showing protein domain conserved during molecular evolution course.[15]

CDART: (Conserved Domain Architecture Retrieval Tool) used for searching protein having similar domain architectures.[16]

C. Bioinformatics tools for plotting protein –ligand interactions:

Ligplot : It is used to find out interaction between protein and ligand also hydrogen and hydrophobic contacts can be represented in this tool.[17].

D. Approaches for classification of proteins:

Classification of proteins b several databases usually is on the basis of their structural similarities. Both structural and evolutionary relationship is factors of their classification. In hierarchy of proteins several levels exist but the main level considered are such as Family, superfamily and fold

Family: In this level proteins are grouped together into family having clear and known evolutional relatedness so called as clear evolutionarily relationship level.

Superfamily: In this level proteins are with low sequence identities but their structural and functional characters suggest a common evolutionary origin so the level called as probable common evolutionary origin. This proteins positioned in superfamily level.

Fold: In this level the proteins are not having evolutionary origin but structural similarities derived from physics and chemistry of proteins facilitating certain chain topologies and packing arrangements. So this level also called as major structural similarity level.
SCOP: It is a database for structural classification of proteins. It provides comprehensive classification of structural and evolutionary relationships between those proteins with known structures.[18].

CATH: (Class, Architecture, Topology and Homologous superfamily): This database facilitates a hierarchical classification for domain structures of proteins, which cause clustering of proteins at four different levels: C, A, T, H means Class, Architecture, Topology and Homologous superfamily, respectively.[19].

III. RESULT AND DISCUSSION

Understanding and analysis of proteins structure and function is one of most important goal of bioinformatics as proteins are important key in biological science research and they are directly and indirectly related to development of diseases, evolution. Mutations and drug discovery. Knowing their structure and function as well as structure-function relationship are very important and helpful to biologists as experimental tools and technologies are not fully support research studies on proteins and their role in disease and other issues like evolution, mutation an drug discovery, investigators help bioinformatics tools for solving and u deep understanding of the biological problems.

Experimental technologies possess disadvantages of time and cost so bioinformatics tools as an alternative method support biologists to process their research studies for prediction of protein structures and their functions.

For example research studies on many proteins like CDKs, Aquaporins, Ion channels, G -protein coupled receptors, Biomarkers and so on for drug discovery, cancer treatment, classification of cancer and so on depend on prediction of structure or structure-function relationship of such biomolecules so for such new and emerging research studies these bioinformatics tools as accessory or alternative tools are very helpful. These tools in combination of many other bioinformatics tools can extent scope of research studies in many aspects for example tools for gene finding, biological pathways, SNP detection, bioinformatics tool of microarray technology and biotechnology tools. [20-29].

IV. CONCLUSION

In this paper, the fundamentals of different structures of protein discussed as basic information. Protein visualization and classification tools help biologists to predict protein structure from its amino acid sequence. So for conducting the projects related to prediction of protein structure or understanding protein function structure, bioinformatics tools can be used in order to provide precise, fast, low cost results. This paper helps biologists to understand and learn basic information of protein structure and bioinformatics tools for identification, classification and visualization of proteins.

V. REFERENCES

[1] David Lee, Oliver Redfern & Christine OrengoPredicting protein function from sequence and structure Nature Reviews Molecular Cell Biology 8, 995-1005 (December 2007)

[2] J. Westbrook, Z. Feng, L. Chen, H. Yang, H.M. Berman,The Protein Data Bank and structural genomics,Nucleic Acids Res, 31 (2003), pp. 489–491

[3] Benkert P, Biasini M and Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27(3): 343–350.

[4] Haas J, Roth S, Arnold K et al. (2013) The protein model portal – a comprehensive resource for protein structure and model information. Database (Oxford) 2013: bat031.

[5] Hildebrand A, Remmert M, Biegert A and Söding J (2009) Fast and accurate automatic structure prediction with HHpred. Proteins 77(Suppl 9): 128–132.

[6] Rastogi.S. C., Rastogi P and Mendiratta N., 2008. Bioinformatics Methods and Applications: Genomics.Proteomics And Drug Discovery PHI Learning Pvt. Ltd

[7] Berman, H. M. (January 2008). "The Protein Data Bank: a historical perspective" (PDF). Acta Crystallographica Section A A64 (1): 88–95. doi:10.1107/S0108767307035623. PMID 18156675

[8] 8Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, Bryant SH. MMDB and VAST+: tracking structural similarities between macromolecular complexes. Nucleic Acids Res. 2014 Jan 1;42(1):D297-303. Epub 2013 Dec 6. doi: 10.1093/nar/gkt1208.

[9] 9Wang Y, Geer LY, Chappey C, Kans JA, Bryant SH. Cn3D: sequence and structure views for Entrez. Trends Biochem Sci. 2000 Jun; 25(6): 300-2. PubMed PMID: 10838572]

[10] Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18, 2714-2723

[11] Richardson, D. C.; J.S. Richardson (January 1992). "The kinemage: a tool for scientific communication". Protein Science 1 (1): 3–9. doi:10.1002/pro.5560010102. PMC 2142077. PMID 1304880.
[12] Laskowski RA (Jan 2001). "PDBsum: summaries and analyses of PDB structures". Nucleic Acids Research 29 (1): 221–2. doi:10.1093/nar/29.1.221. PMC 29784. PMID 11125097
[13] Gibrat JF, Madej T, Bryant SH. Surprising similarities in structure comparison. Curr Opin Struct Biol. 1996 Jun; 6(3): 377-85.
[14] Holm L, Sander C. Dali: a network tool for protein structure comparisonTrends Biochem Sci. 1995 Nov;20(11):478-80.
[15] Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015 Jan 28;43(Database issue):D222-2. doi: 10.1093/nar/gku1221. Epub 2014 Nov 20
[16] Geer LY, Domrachev M, Lipman DJ, Bryant SH. CDART: protein homology by domain architecture. Genome Res. 2002 Oct;12(10):1619-23.
[17] Wallace A C, Laskowski R A, Thornton J M (1996). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng., 8, 127-134
[18] Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536-540.
[19] Sillitoe I, Lewis, TE, Cuff AL, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees J, Lehtinen S, Studer R, Thornton JM, Orengo CA,Nucleic Acids Res. 2015 Jan,CATH: comprehensive structural and functional annotations for genome sequences.
[20] Ghorbani M and Karimi H. Cyclin-Dependent Kinases as valid targets for cancer treatment. Journal of Pharmacy Research 2015,9(6),377-382
[21] Ghorbani M, Karimi H, 'Ion Channels Association with Diseases and their Role as Therapeutic Targets in Drug Discovery', International Journal of Scientific Research in Science and Technology(IJSRST), 1(3):65-69,July-August 2015.
[22] Ghorbani M, Karimi H, 'Role of Aquaporins in Diseases and Drug Discovery', International Journal of Scientific Research in Science and Technology(IJSRST),1(3):60-64, July-August 2015
[23] Mahin Ghorbani, Hamed Karimi, 'Role of G-Protein Coupled Receptors in Cancer Research and Drug Discovery', International Journal of Scientific Research in Science and Technology (IJSRST),1(3), pp.122-126, July-August 2015.
[24] Mahin Ghorbani, Hamed karimi, 'Role of Biomarkers in Cancer Research and Drug Development', International Journal of Scientific Research in Science and Technology(IJSRST),1(3), pp.127-132, July-August 2015
[25] Ghorbani M,Karimi H,Ten Bioinformatics Tools for Single Nucleotide Polymorphisms, American Journal of Bioinformatics ;2014:3(2):45-48
[26] Mahin Ghorbani, Hamed Karimi, 'Bioinformatics Approaches for Gene Finding ', International Journal of Scientific Research in Science and Technology (IJSRST),1(4),12-15, September-October 2015.
[27] Mahin Ghorbani, Hamed Karimi, 'Bioinformatics Methods for Biochemical Pathways and System Biology Analysis', International Journal of Scientific Research in Science and Technology (IJSRST),1(4),75-79, September-October 2015.
[28] Mahin Ghorbani, Hamed Karimi, Role of Biotechnology in cancer control, International Journal of Scientific Research in Science and Technology (IJSRST),,1(5),180-185, November-December 2015
[29] Ghorbani M, Karimi H, 'Role of Microarray Technology in Diagnosis and Classification of Malignant Tumours', International Journal of Scientific Research in Science and Technology(IJSRST), 1(3):117-121, July-August 2015