Udayakumar Mani et al /J. Pharm. Sci. & Res. Vol.5(7), 2013, 150 - 152

ISSN:0975-1459
Journal of Pharmaceutical Sciences and Research
www.jpsr.pharmainfo.in

PDBMut: A Protein Mutant Identification Server

Udayakumar Mani*, Sadhana Ravisankar, Sai Mukund Ramakrishnan

Department of Bioinformatics,
School of Chemical and Biotechnology,
SASTRA University, Tanjore – 613 401,
Tamil Nadu, India

Abstract

Proteins are large biomolecules comprising of α – L amino acids connected by peptide bonds. They fold into a three dimensional structure to become functionally significant. Most of the information related to protein folding is embedded within the primary peptide sequence itself. With the advent of structural genomics, prediction of biological function from protein structure has become one of the major goals of structural biology. The structure data comes in the form of a PDB file, listing the three-dimensional (3D) coordinates of atoms in the protein. The "site" record which is present sometimes in the PDB file can be used for marking the ligand-binding sites, metal-binding sites or active sites. Analysis of the PDB files suggests that a limited number of protein folds/families occur in nature. This limitation is probably due to physicochemical constraints on protein folding that favour particular packing arrangements. There are many methods from which protein function can be inferred with respect to its structure. Important among these techniques is identification based on homologous sequences. Homologous proteins within species are similar in sequence and are likely, but not guaranteed, to have a similar function. Comparison of an un-annotated sequence to known homologous sequences is therefore a good starting point for predicting function. Centered on this concept, we have developed an online server PDBMut, for comparing protein structures with respect to their primary sequence. The tool is unique to itself as it identifies similar protein structures as well as structures of proteins with one or more differing amino acids.

Keywords: Homologous proteins, Mutation, Protein structure, Protein function, Structure comparison.

INTRODUCTION

Proteins are assembled from amino acids using information encoded in genes. Each protein has its own unique amino acid sequence that is specified by the nucleotide sequence of the gene encoding this protein. Protein folding is the process by which a protein structure attains its functional shape or conformation. Proteins fold into unique 3-dimensional structures known as the native conformation. The 3-dimensional structure is what controls the basic function of the protein**.** Since the advent of structural genomics, deciphering biological function from structure has become a major goal in structural biology and bioinformatics **[1]**. Assignment of biological function provides a valuable first step towards experimental characterization of the cellular and physiological roles of gene products. Ultimately, this assignment would improve genome analysis and annotation, and aid in the design of proteins with novel or modified function. At present (May, 2013) the Brookhaven Protein Databank (PDB) **[2]** contains nearly 90,424 (as on May 2013) protein entries and the number is increasing by approximately by 200 a month. Some proteins have structural similarities with other proteins and, in many cases, share a common evolutionary origin **[3]**. Protein homology refers to the biological homology between proteins, meaning that the proteins are derived from a common "ancestor". Inferring a protein's function by homology is a convenient methodology for biologists. Since the two well-defined secondary structural units, the alpha-helix and the beta-sheet are abundant in proteins, families of protein structures have been classified by a system based upon the alpha-helix and beta-sheet topologies. Every different topology may be considered as a fold. To date, most folds have had a homologous family associated with them **[4]**. Within homologous protein families, it is expected that members of the family will have related functions. However, this is not always the case as considerable diversity has been observed within homologous super families. Till date, there are a considerable number of tools and online servers like ProFunc Server **[5]**; PredictProtein Server **[6]** that infers the protein functions by performing structural analysis. Streamlining towards the ideology of procuring protein function form structure we have put together an online server, PDBMut that performs structural comparison of proteins with respect to the deviation in the number of amino acids in the protein and aids in the inferring of its function.

METHODOLOGY

PDBMut is a structure cum sequence analysis server useful for analyzing amino acid sequences in a protein, their structures and mutational co-variations. The server predicts structural similarities based on the number of amino acids deviating in between the sequences. End users can get identical structures if they input the query deviation entry as 0. Similar structures can also be obtained by giving a deviation range of 1,2,3,4 etc. PDBMut is built using Hypertext Preprocessor language implementing relational database management systems using MySQL **[7]**. The software application is user friendly and has been tested in various computer platforms using internet browsers like Google Chrome, Mozilla Firefox. Navigation of the server starts on entering a valid PDB identifier (protein identifier/PDB ID), selecting the protein chain of interest, and specifying the mutation count (Figure: 1 (A)). Based upon the deviation range and the chain specified, the PHP script

fetches sequences with the same length as the query sequence from the database. The output is displayed in a table, which displays the PDB file name, ID, length, differing residues and their position (Figure: 1 (C)). Multi-structure comparison is a feature unique to the tool. The query sequence is the template against which the user has the option of selecting 'n' number of structures to be aligned. Secondary Structural Elements (SSE) of the input PDB identifier can be visualized graphically using the SVG module **[https://developer.mozilla.org/en-US/docs/SVG/Tutorial]** (Figure: 1 (C)). A structural alignment viewer of the input protein and a selected output structure can also be viewed using RCSB Structural alignment view (Figure: 2). PDBMut can be used to identify molecules of similar functions with divergent residues.

CASE STUDY:

Two proteins structures 118L and 109L were taken, with both having a sequence length of 164 in chain A. A ClustalW alignment was done between the 2 protein structures inferring sequence divergence at 2 positions 44 and 130 (Figure: 1 (C)). The server also predicts the same output when the user queries with a deviation range of 2 amino acids in the input. The identification was procured that, the proteins were functionally similar but on the contrary are structurally different. Using PDBMut researchers can extend their horizon of identifying protein function from sequence based on homologous structures.

Figure: 1. (A) Prerequisites for PDBMut. **(B)** ClustalW alignment between 118L_A and 109L_A showing the divergence in 2 residue positions. **(C)** Tabulated depiction of the similar structures for 118L_A differing by 2 residues. The SSE of the input protein is depicted graphically in red (**S**- Sheets) and green (**H**-Helix)

Structure Sequence of 118L_A:

MNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSL
NAAKSELDKAIGRNTNGVITKDEAEKLFNQDVDAAVR
GILRNAKLKPVYDSLDAVRRAALINMVFQMGETGVA
GFTNSLRMLQQKRWDEASVNLAKSRWYNQTPNRAKR
VITTFRTGTWDAYKNL

Structure Sequence of 109L_A:

MNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSL
NAAKKELDKAIGRNTNGVITKDEAEKLFNQDVDAAV
RGILRNAKLKPVYDSLDAVRRAALINMVFQMGETGV
AGFTNSLRMLQQKRWDEAAVNLAKSRWYNQTPNRA
KRVITTFRTGTWDAYKNL

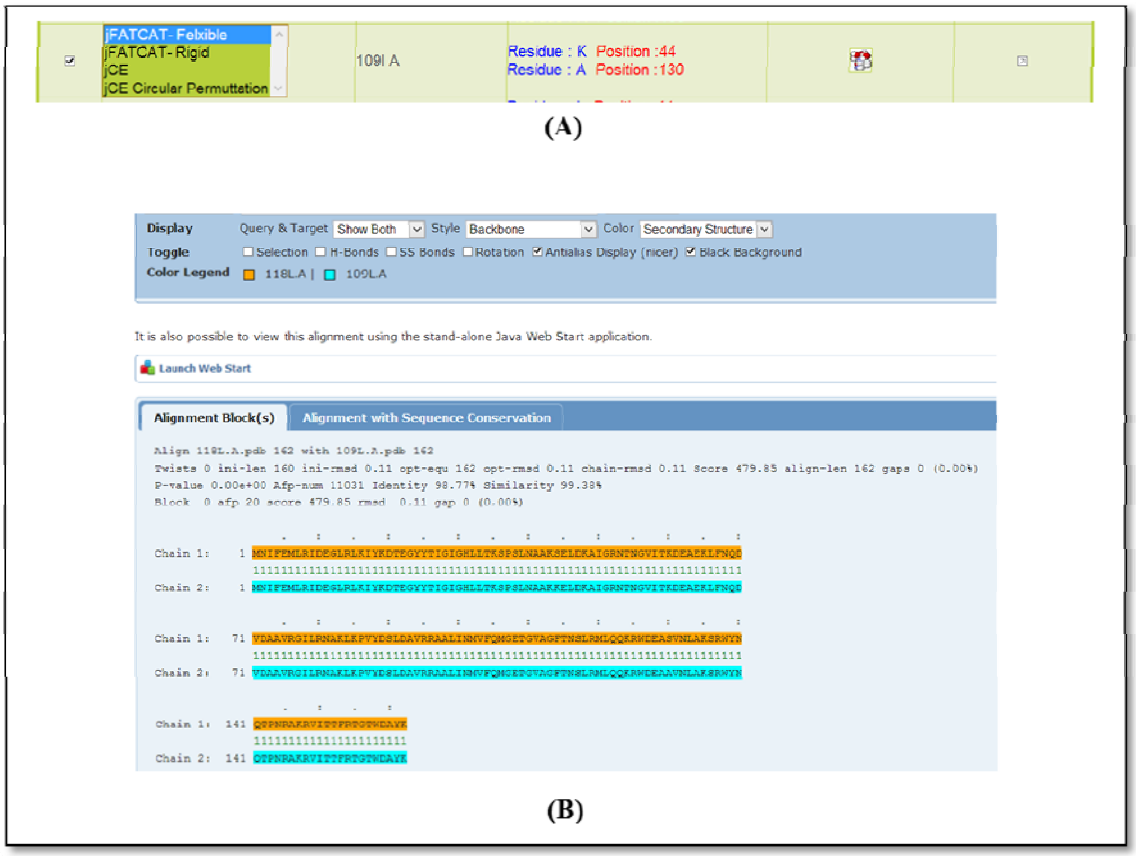

Figure: 2. (A) 109L_A selected for the structural alignment against 118L_A. The various structural alignment algorithms used by PDBMut and PDB are displayed in the drop down. **(B)** Structural alignment of 118L_A and 109L_A.

HARDWARE AND SOFTWARE REQUIREMENTS

Minimum hardware requirement of a dual core processor, 2GB RAM is required to meet the full functionality of PDBMut. Good internet speed and connectivity is required for using PDBMut to its fullest of potential. PDBMut can accessed from http://www.bioindians.org/pdbmut

CONCLUSIONS

PDBMut is by far the only online server for comparison of protein structures to find out the differing residues in the pair of structures. The server will also be useful for structural biologists who find it difficult to compare all the structures at the same time from the PDB to get the result. PDBMut alleviates the burden of researchers, enabling them to quickly compare protein structures and get the required results .The output produced by the server allows the user to figure out the location of the residues in the secondary structural elements. PDBMut is currently incorporated with 70,000 structures from PDB database to test the working process. As PDBMut has proven to provide accurate results the remaining of the PDB database is being added into the PDBMut database. Additions like the structural analyzer, based on domains and motifs and a multiple sequence alignment generator will be added to the later versions of PDBMut.

ACKNOWLEDGMENTS

We thank the Department of Bioinformatics, SASTRA University for providing us the necessary infrastructure to carry out our research work successfully. We also thank Dr. M. Vijayalakshmi for her support and motivation.

REFERENCE

1. Shapiro, L., Harris, T., *Curr Opin Biotech*. 2000, *11(1)*, 31–5.
2. Abola, E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F., Weng, J., *Crystallographic Databases—Information Content, Software Systems, Scientific Applications*, Data Commission of the International Union of Crystallography, Cambridge 1987, pp. 107–132.
3. Loredana L.C., Bart, A., Hubbard T.J.P., Brenner, S.E., Murzin, A.G., Chothia, C., *Nucleic acids res*. 2000, *28(1)*, 257-259.
4. Todd, A., Orengo, C., Thornton, J., *Curr Opin Chem Biol*. 1999, *3*, 548–556.
5. Roman, A.L., James, D.W., Janet, M.T., *Nucleic acids res*. 2005, *33*, W89-W93.
6. Rost, B., Yachdav, G., Liu, J., *Nucleic acids res*. 2004, *32*, W321-W326.
7. Hugh, E., Williams, E., Lane, D., *Web Database Applications with PHP and MySQL*, O' Reilly Media Inc., California 2004.