

Research Article



STUDY OF RICE TELOMERE BINDING PROTEIN 1 (RTBP1): AN *IN SILICO* APPROACH

Koel Mukherjee*, Ashutosh Kumar, Dev M Pandey and Ambarish S Vidyarthi

Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India.

*Corresponding author's E-mail: koelmukherjee@bitmesra.ac.in

Accepted on: 05-06-2011; Finalized on: 30-08-2011.

ABSTRACT

Helix Turn Helix is one of the simple structural motifs that reside in the sequence specific DNA binding domain of transcription factors. The interaction of HTH motif with DNA regulates the gene expression. RTBP1 is a sequence specific transcription factor binding protein involved in telomere length regulation in rice (*Oryza sativa*). In this paper, attempts were made to characterise the full protein of 633 amino acid length. We have undertaken a sequential study of this protein by sequence comparison and comparative studies to understand the spatial arrangement of helix – turn - helix motif that is involved in the DNA protein interaction. Knowledge based approaches, combined with the different computational tools focusing on similarity searches, alignments, domain identification, motif prediction and its analysis, detection of specific regions in the motifs, may move us to a new prospective paradigm.

Keywords: Transcription Factor, HTH motif, DNA binding domain, Rice Telomere Binding Protein.

INTRODUCTION

The regulation of gene expression – induction or turning genes on or turning genes off can be accomplished by both positive and negative control mechanism. Modulation of transcriptional activity is fundamental to the regulation of gene expression and it is largely mediated through some proteins that are directly or indirectly interact with the DNA. In the promoter region there are specific sets of short conserved sequences, each of which is recognized by corresponding site specific DNA-binding proteins that helps in the initiation of transcription process. These proteins are called 'Transcription Factor'¹. For the regulation of transcription, Transcription Factors senses the signal through Signal Sensing Domain (optional) followed by the binding of specific site mediated DNA Binding Domain and its activation through Transcription Activation Domain².

Telomere Binding Protein, one major type of transcription Factor is the necessary building blocks of telomere structure. It not only helps in the protection of the DNA ends but also help the telomerase in length regulation. Major advances have been already made in understanding the role of Telomere Binding Proteins in different actions like telomere length regulation, telomere end protection and cancerous growth in yeast, drosophila and man³. It has been also noticed that the TBPs are involved in many aspects of plant development such as flower and root development, cell fate specification. While many putative TBPs have been also isolated from plants but most of them are poorly characterized. The DNA binding domain of Tobacco (NgTRF1)⁴ and Arabidopsis (AtTBP1)⁵ also show the telomere binding activity and helps in length regulation of the plants.

Rice (*Oryza sativa*) is the staple food crop as well as a model system in monocots⁶ and grown all over the world. Despite vast information on Transcription Factors and their regulation at the gene and protein level in eukaryotic system, our knowledge about the conformational arrangement and the interaction of Transcription Factors in rice is still limited. The availability of full rice genome sequence is a great advantage towards computational studies on different Transcription Factors. It is reported 141 families of DNA - binding domains that contain HTH motifs (<http://pfam.sanger.ac.uk/clan/HTH>). Several of them belong to rice Transcription Factors (http://plntfdb.bio.uni-potsdam.de/v3.0/index.php?sp_id=OSAJ). Out of these one interesting, Telomere Binding Protein was detected having 633 amino acid long chain⁷. Ko *et. al.* (2009) reported the solution structure of DNA Binding Domain of RTBP1 (122 amino acid long chain at C-terminus, PDB ID: 2ROH). The DNA binding domain contains a HTH motif. Helix – Turn - Helix motif is the most common DNA – binding motif represented by two alpha - helical segments linked by short non - helical segment⁸. Normally the HTH motifs are composed of 'stabilization' and 'recognition' helices connected by a sharp turn⁹. While analyzing the secondary structure of full RTBP1 protein 6 more helices were detected except the known DNA Binding Domain. So there is considerable interest to identify the pattern / structure and conformational arrangement of HTH motifs in RTBP1 full protein sequence by *In Silico* approach.

MATERIALS AND METHODS

Sequence & Structure retrieval

The sequences of RTBP1 domain (GI: 225733909) and the full protein sequence (GI: 9716453) were collected



from NCBI (<http://www.ncbi.nlm.nih.gov/>). The length of the full protein is 633 amino acids and the 122 amino acid at the C-terminus constitutes the DNA binding domain. The solution structure of DNA Binding Domain of RTBP1 (PDB ID: 2ROH) was taken from PDB (<http://www.rcsb.org/pdb/explore/explore.do?structureId=2ROH>).

Phylogenetic Analysis

To analyze the whole protein sequence, full RTBP1 protein sequence was retrieved from NCBI (<http://www.ncbi.nlm.nih.gov/>) protein database. BLASTP 2.2.24+¹⁰ was performed and sequences having > 90% query coverage was retrieved. These sequences were used for Multiple Sequence Alignment by ClustalX (1.83)¹¹ separately. To view the putative evolutionary relationship among these sequences, the ClustalX output file was analyzed by PHYLIP 3.65 package (<http://evolution.genetics.washington.edu/phylip.html>). It consist different interdependent tools to draw a phylogenetic tree. First, PRODIST computes distance between pair of protein sequences and estimates the total branch length between them. From the output of the previous step a distance matrix is calculated by NEIGHBOR¹². Lastly the rooted tree was drawn by DRAWGRAM (interactively plots a cladogram - or phenogram - like rooted tree) on the basis of distance matrix calculated by NEIGHBOR.

2-D Structure Prediction

3-D structure of 122 amino acids long DNA binding domain of RTBP1 at the C-terminus is already reported⁷ with PDB ID : 2ROH. But the Secondary structure of full protein has not been yet reported. Therefore 2-D structure of full RTBP1 protein was predicted using Jpred3 (v3)¹³ (<http://www.compbio.dundee.ac.uk/www-jpred/>). It uses the Jnet algorithm in order to make more accurate 2-D structure predictions of proteins as well as prediction on Solvent Accessibility and Coiled – coil regions (Lupas method) basis (Fig 2-A, B, C, D). Output of the Jpred was further used to draw an average distance tree using Blossum62.

Domain & Protein architecture Prediction

Domains were identified in the predicted 2 – D structure of full RTBP1 protein. Online available software CDART: Conserved Domain Architecture Retrieval Tool¹⁴ (<http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi>) was used for domain prediction (Fig 3A). The criteria of Probability value that a domain hit must exceed 0.01 was given and low complexity regions were ignored. CDART defines a domain by a Position – specific scoring matrix (PSSM), based on probabilistic values of amino acids exist at each position of the domain.

Parallel to domain prediction the full protein architecture was also predicted (UniProtKB / Swiss-Prot

Q9LL45 (TBP1_ORYSJ) (Fig - 3B). Conserved Domains were searched against known motifs through online available software **InterProScan**, provided by the European Bioinformatics Institute (<http://www.ebi.ac.uk/Tools/InterProScan/>). This search is comprehensively performed against various databases and provides sophisticated graphical output.

Motif Identification

Identified domains were scanned for the presence of conserved motifs by MEME software¹⁵ (<http://meme.nbcr.net>) (Fig 4) considering the total number of possible motifs per session manually. MEME searches statistically significant motif by local, multiple sequence alignment, for repeated, ungapped sequence patterns that occur in the DNA or protein sequences (Fig 5).

RESULTS AND DISCUSSION

The results were obtained at each step of computer-assisted search and analysis of the protein in interest. From the BLAST result of the full protein we chose 12 sequences which are having the query coverage of more than 90%. To predict the conserved regions among these protein sequences ClustalX was done. A good conserved region was identified from the ClustalX result which reveals that sequences having a pretty good similarity. So to check the phylogenetic relationship among these sequences PHYLIP package was used. The tree was generated but it was noticed that the RTBP1 protein was making a single branch and the nearest protein was TBP1 of *Gallus gallus*. All the other plant TBP proteins were far distant from RTBP1 (Fig 1). Phylogenetic tree predicted RTBP1 full protein as in a singleton cluster.

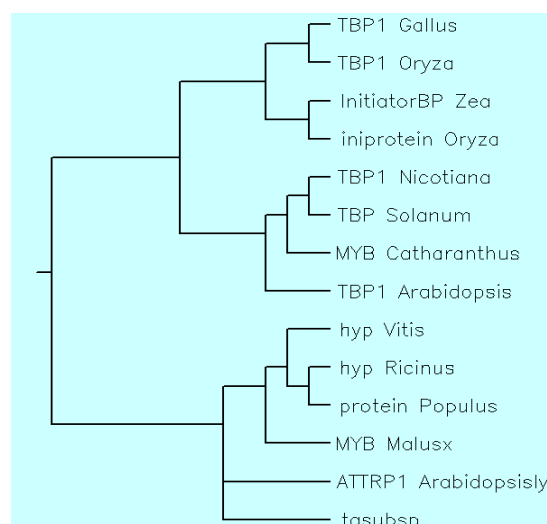


Figure 1: Phylogenetic analysis of full protein RTBP1 with 633 amino acids long chain and other 12 sequences from BLAST result by PHYLIP software. The concerned protein is represented in the figure as TBP1 Oryza.

To identify the exact position of helices, strands and coils in the full protein sequence, secondary structure

prediction is the best way which was carried over by Jpred tool. The DNA binding domain residing in the C – terminal (506-615) consist of four helices to arrange a HTH motif. But the prediction also reveals that there are more 6 helices (helix1 : 38-52, helix2 : 70-80, helix3 : 284-296, helix4 : 328-337, helix5 : 370-390, helix6 : 408-415) out of the DNA binding domain that may help in the interaction with DNA (Fig 2 A,B,C,D).

The complete sequence of RTBP retrieved from NCBI evaluates only one domain named as SANT around C terminal region by the CDART online tool (Fig 3A). Domain analysis is the most useful level to understand the protein function and structure. To be sure with the domain analysis result full protein architecture was also predicted by **InterProScan**, provided by the European Bioinformatics Institute. There also show the SANT domain, HTH myb domain¹⁶ and DNA binding part (Fig 3B). But all these were in the C - terminal region (529-615). One new information was revealed that in the region of 351 – 430 there was also another domain '**Ubiquitine - like**' domain¹⁷. From the secondary structure prediction it was already available that helix5 and helix6 was residing in Ubiquitine – like domain part and may be have the property to build a new HTH motif.

Protein sequence motifs are signatures of protein families and can often be used as tools for the prediction of protein function. For RTBP1 sequence we search the motifs by MEME where we keep the condition width of 50, 100, 150. The minimum length of motif was 5 and maximum was 50. According to MEME result we got the best report where the width was 100. All the 12 sequence was given as input. Total 15 motifs were found out of which 12 motifs for RTBP1 (Fig 4). Motif1 was the biggest motif and situated in the C - terminal area. But surprisingly we found another motif which was second largest and also falling in the region of 320 – 410 of the sequence where we already found the helix5 and helix6. Depending on the MEME result again one phylogram was drawn to find the relation of RTBP1 with other sequences (Fig 5).

Ubiquitin like domain was the new finding with respect to *In silico* full sequence analysis of RTBP1 protein. This domain constituted with helix5 and helix6 connected with strand and coil and MEME result also proved the presence of the motif. This new findings can help us to reveal the knowledge of particular amino acids which are interacting with other protein or DNA during transcription. Another important point also came forward that although the full protein sequence was having a good sequence similarity with other plant TFB protein sequences but the RTBP1 in phylogenetic analysis always remain in one different branch. Finally to conclude, RTBP1 protein is having very different properties which can help in resolving the plant aging problem.

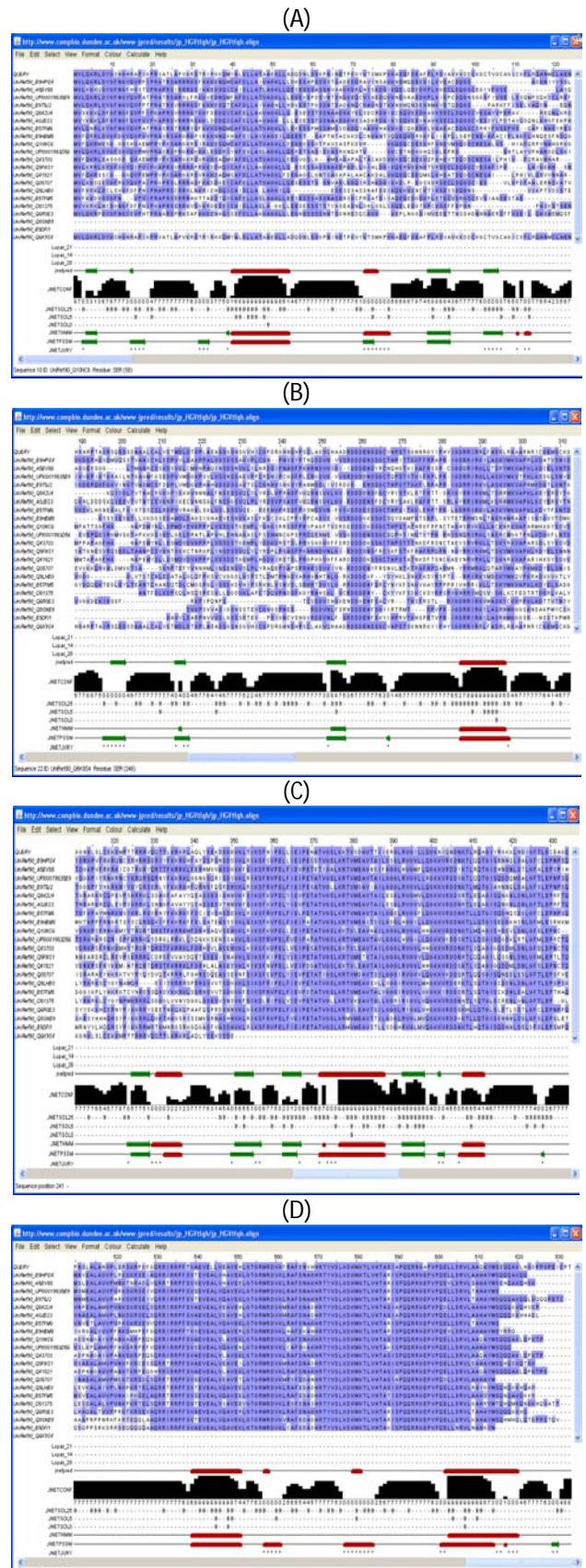
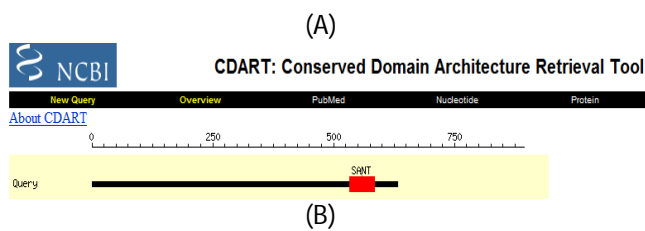


Figure 2: The secondary structure of RTBP1 full protein is shown above by the Jpred Analysis Tool. The alignment shown here of full protein. Helix regions were shown as cylindrical rod in red colour, bold arrow with green colour representing a Beta-strand. The buried amino acids are also represented by 'B' in the JNETSOL25 program within the Jpred tool.





(B)

Molecule processing			
Chain	1 – 633	633	Telomere-binding protein 1
PRO_0000387498			
Regions			
Domain	351 – 430	80	Ubiquitin-like
Domain	529 – 588	60	HTH myb-type
Domain	7534 – 7584	51	SANT
DNA binding	557 – 584	28	H-T-H motif (By similarity)
Region	506 – 615	110	Sufficient for telomeric DNA binding

Figure 3: (A) Domain prediction of full protein by CDART. Query is our sequence and the red rectangle showing the SANT domain in C-terminal region. (B) Full protein architecture was predicted by InterProScan showing different domain names, their lengths and their orientation along the full protein sequence.

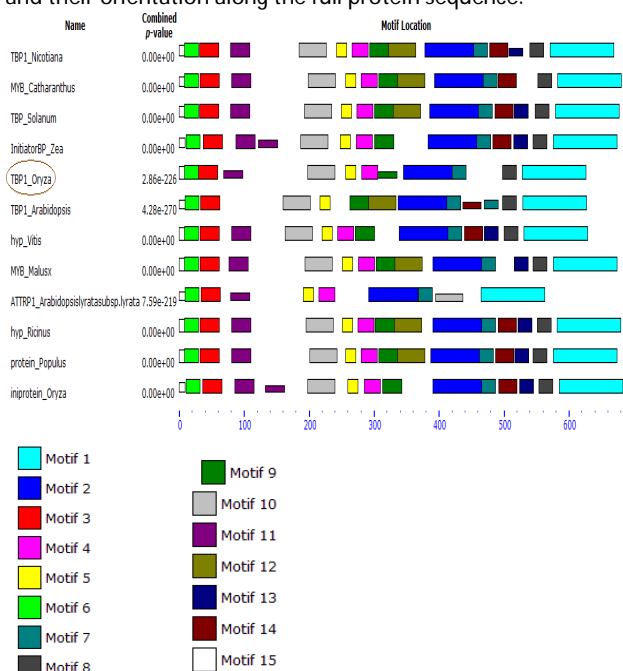


Figure 4: Schematic Motif orientation of TBP1 proteins. The protein of concern is circled by oval shape. 12 numbers of motifs were predicted by MEME. Different colors were used to identify the 12 motifs in TBP1 in rice. Motif 1 is biggest in the length with light blue colour. The second motif is next biggest with navy blue color.

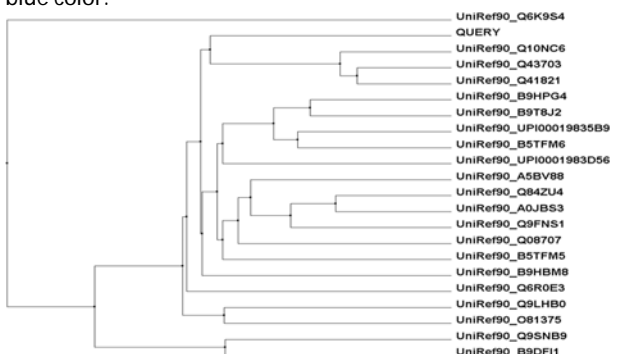


Figure 5: Phylogram of MEME analysis of RTBP1, showing the Average distance tree using Blosum62. The RTBP1 protein is shown here as QUERY. The tree again showing the QUERY to be far distant from the other references proteins.

Acknowledgements: All authors are highly thankful to Birla Institute of Technology, Mesra, Ranchi for the infrastructural facilities. We hereby acknowledge the Department of Agriculture, Government of Jharkhand (Grant No.5/B.K.V/Misc/12/2001), BTISNetSubDIC (BT/BI/04/065/04), from DBT, Govt. of India to our department.

REFERENCES

1. Pabo CO and Sauer RT. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* 1992; 61: 1053–1095.
2. Mitchell PJ, Tjian R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*.1989; 245 (4916): 371–8. doi:10.1126/science.2667136. PMID: 2667136.
3. Hernandez N. TBP, a universal eukaryotic transcription factor? *Genes Dev.* 1993; 7: 1291-1308.
4. Ko S, Jun SH, Bae H, Byun JS, Han W, Park H, Yang SW, Park SY, Jeon YH, Cheong C. Structure of the DNA-binding domain of NgTRF1 reveals unique features of plant telomere-binding proteins. *Nucleic Acids Res.* 2008 ; 36(8):2739-55.
5. Sue SC, Hsiao H hao, Chung Ben C.-P., Cheng YH, Hsueh KL, Mong CC. Solution Structure of the Arabidopsis thaliana Telomeric Repeat-binding Protein DNA Binding Domain: A New Fold with an Additional C-terminal Helix. *Journal of Molecular Biology.* 2006; 356(1): 72-85.
6. Cantrell P R., Hettel PG. New challenges and technological opportunities for rice-based production systems for food security and poverty alleviation in asia and the pacific. *fao rice conference Rome, Italy, 12-13 February 2004.*
7. Ko S., Yu E.Y., Shin J., Yoo H.H., Tanaka T., Kim W.T., Cho H.-S., Lee W. and Chung I.K. Solution structure of the DNA binding domain of rice telomere binding protein RTBP1. *Biochemistry.* 2009;48: 827-838.
8. Brennan RG, Matthews BW. The Helix-Turn-Helix DNA Binding Motif, *The Journal Of Biological Chemistry.* 1989; 264(4): 1903-1906.
9. L.Aravind , Anantharaman V, Balaji S, Babu MM, Iyer M.L. The many faces of the helix-turn-helix domain: Transcription regulation and beyond q L. *FEMS Microbiology Reviews.* 2005; 29(2): 231–262.
10. Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 1997; 25: 3389-402.
11. Jeanmougin,F., Thompson, J. D., Gouy, M., Higgins, D. G. and Gibson, T. J. Multiple Sequence Alignment With Clustal X. *Trends Biochem Sci.* 1998; 23: 403405.
12. Saitou N, Nei M. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* 1987; 4(4): 406-425.
13. Cole C, Barber D J., Barton J G.The Jpred 3 secondary structure prediction server. *Nucleic Acids Research,* 2008; 36: 197–201, doi:10.1093/nar/gkn238.

14. Geer LY., Domrachev M, Lipman DJ., Bryant HS. CDART: Protein Homology by Domain Architecture. *Genome Research*. 2002; 12: 1619–1623.
15. Bailey L.T., Williams N, Mischel C, and Li W W., "MEME: discovering and analyzing DNA and protein sequence motifs", *Nucleic Acids Research*, 2006; 34: 369-373.
16. Baker M A., Fu Q, Hayward W, Lindsay M S., Fletcher M T. The Myb/SANT domain of the telomere-binding protein TRF2 alters chromatin structure. *Nucleic Acids Research*, 2009; 37(15): 5019–5031.
17. Pulido S L., Devos D, Sung R Z., Calonje M. RAWUL: A new ubiquitin-like domain in PRC1 Ring finger proteins that unveils putative plant and worm PRC1 orthologs, *BMC Genomics* 2008; 9(308): 1-11.

