

In silico characterization of a RNA binding protein of cattle filarial parasite *Setaria digitata*

Nirupa Nagaratnam, Eric Hamilton Karunanayake*, Kamani Hemamala Tennekoon, Sameera Ranganath Samarakoon & Karthika Mayan

Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo 90, Cumaratunga Munidasa Mawatha, Colombo 03, Sri Lanka; Eric Hamilton Karunanayake - Email: erick@ibmbb.cmb.ac.lk, *Corresponding author

Received July 08, 2014; Accepted July 11, 2014; Published August 30, 2014

Abstract:

Human lymphatic filariasis (HLF) is a neglected tropical disease which threatens nearly 1.4 billion people in 73 countries worldwide. *Wuchereria bancrofti* is the major causative agent of HLF and it closely resembles cattle filarial parasite *Setaria digitata*. Due to difficulties in procuring *W. bancrofti* parasite material, *S. digitata* cDNA library has been constructed to identify novel drug targets against HLF and many of the cDNA sequences are yet to be assigned structure and function. In this study, a 549 bp long cDNA (*sdrbp*) has been sequenced and characterized *in silico*. The shortest ORF of 249 bp from the isolated cDNA encodes a polypeptide of 82 amino acids and shows an amino acid identity of 54% with the RRM domain of human cleavage stimulation factor-64 kDa subunit (CstF-64). Structure of the protein (sdRBP) obtained by homology modelling using RRM of CstF-64 as template adopts classical RRM topology ($\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$). sdRBP model built was validated by superimposition tools and Ramachandran plot analysis. CstF-64 plays an important role in pre-mRNA polyadenylation by interacting with specific GU-rich downstream sequence element. Molecular docking studies of sdRBP with different RNA molecules revealed that sdRBP has greater binding affinity to GU-rich RNA and comparable results were obtained upon similar docking of RRM of CstF-64 with the same RNA molecules. Therefore, sdRBP is likely to perform homologous function in *S. digitata*. This study brings new dimensions to the functional analysis of RNA binding proteins of *S. digitata* and their evaluation as new drug targets against HLF.

Keywords: Cleavage stimulation factor, RRM domain, superimposition, Ramachandran plot, pre-mRNA polyadenylation, GU-rich downstream sequence element, Molecular docking

Background:

Human Lymphatic Filariasis (HLF) is a neglected tropical disease caused by three types of thread like round worms; *Wuchereria bancrofti*, *Brugia malayi*, and *Brugia timori*. Adult nematodes live for 6-8 years and produce millions of small larvae. The circulating microfilariae are transmitted by blood-sucking arthropods (e.g. *Culex* mosquito). More than 120 million people are currently infected, with about 40 million disfigured and debilitated by the disease. HLF results in an altered lymphatic system and chronic lymphoedema, causing pain and severe disability. Disease transmission is prevented by annual mass drug administration of single doses of two

medicines given together - albendazole plus either ivermectin in areas where onchocerciasis is also endemic or diethylcarbamazine citrate (DEC) in areas where onchocerciasis is not endemic [1]. These drugs eliminate microfilariae from the bloodstream, but cannot clear adult nematodes that lodge in the lymphatic system. In addition, studies have reported that many helminth parasites are developing resistance against albendazole and ivermectin [2].

Therefore, potential research strategies are needed to develop or design other effective pharmacological therapeutics to eradicate HLF globally. Due to difficulties in procuring human

filial parasite material, cattle filial parasite *Setaria digitata* was studied as a model organism to identify novel drug targets against HLF. *S. digitata* closely resembles *W. bancrofti* in many aspects such as morphology, histology, response to drugs and antigenicity. *S. digitata* cDNA library was constructed and several cDNA clones had been sequenced and characterized to identify novel filial proteins [3]. The present work is primarily focused on a selected cDNA sequence which has nucleotide similarity to human filial parasite gene sequence. The structure-function relationship of the hypothetical protein of *S. digitata* is predicted using similarity search, homology modelling, superimposition tools, Ramachandran plot analysis and molecular docking studies *in silico*.

Methodology:

cDNA Sequencing

Twenty recombinant clones from *S. digitata* cDNA library were randomly selected and amplified. Plasmid DNA was isolated by alkaline lysis method and digested by EcoR1 and Xho1 restriction enzymes. Clones having inserts of more than 500 bp in length were chosen for DNA sequencing. Sequencing reactions were performed using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) with pBK-CMV vector specific primers T3 and T7. Products of sequencing reactions were resolved by the genetic analyser; Applied Biosystems® 3500 Dx and data were viewed by BioEdit Sequence Alignment Editor.

Nucleotide analysis

Segments of vector origin were removed from insert sequences by VecScreen program [4]. Obtained cDNA sequences were used as query for searching homologous sequences using BLASTn [5]. cDNA which contains uncharacterized filial parasite protein and shows only partial similarity to human proteins was selected for further bioinformatics characterization.

Amino acid sequence analysis

All possible open reading frames (ORFs) for the selected cDNA were identified by ORF finder [6]. Amino acid sequences derived by conceptual translation of each of the ORFs were used as query for searching homologous sequences using BLASTp [5]. Amino acid sequence which contained putative conserved domain and shows sequence similarity to human filial parasite protein was selected for structural modelling.

Homology modelling and structural analysis

Three dimensional (3D) structure of *S. digitata* hypothetical protein was built by SWISS-MODEL Workspace of ExPASy Proteomics Server [7] using RRM (RNA recognition motif) of cleavage stimulation factor 64 kda subunit (CstF-64) (PDB ID: 1P1T_A) as template and viewed by UCSF Chimera tool [8]. Amino acid sequences of both *S. digitata* and the homologous protein were aligned using MultAlignViewer function of UCSF Chimera and active site residues, organization of protein secondary structures and amino acid sequence of functional domains were analysed.

Structure validation and structural comparison

Quality of sdrBP model built was validated by Ramachandran's map using PROCHECK [9]. 3D structure of CstF-64 (PDB ID: 1P1T) was downloaded from RCSB-PDB [10]

and RRM domain of CstF-64 was separated using UCSF Chimera. For comparative analysis, RRM of CstF-64 was also submitted to PROCHECK program and plot statistics were compared. sdrBP and RRM of CstF-64 were superimposed for structural similarity using Matchmaker function of UCSF Chimera and RMSD (root mean square deviation) value was obtained.

Automated Molecular Docking

3D structures of two octameric RNAs (PDB ID: 1SA9 and 472D) were retrieved from RCSB-PDB. RNA duplexes were separated into single strand RNAs (1SA9: 5'-R(*GP*GP*CP*GP*AP*GP*CP*C)-3', 472D_A: 5'-R(*GP*UP*GP*UP*UP*UP*AP*C)-3', and 472D_B: 5'-R(*GP*UP*AP*GP*GP*CP*AP*C)-3') by UCSF Chimera and docked with sdrBP by Hex 8.0.0 Cuda software [11]. Docking results were viewed by UCSF Chimera. Default parameters were used for docking process and Energy (E) values of each docking event were obtained. For comparative analysis RRM of CstF-64 was docked with the same RNA molecules and E values were compared.

Validation of various physiological parameters

Structure-function relationship of sdrBP was further validated using ProtParam tool of ExPASy Proteomics Server [12] for various parameters such as theoretical pI, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

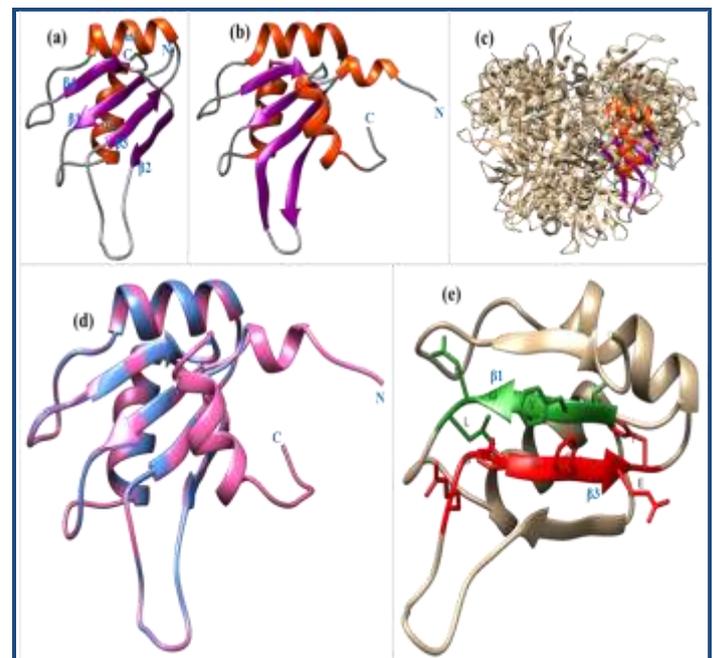


Figure 1: (a) Homology model of the hypothetical protein, sdrBP built by SWISS-MODEL Workspace. (b) NMR structure of the N-terminal RRM domain of human CstF-64 (PDB ID: 1P1T_A). (c) NMR structure of human CstF-64 (PDB ID: 1P1T) retrieved from RCSB-PDB (d) Superposition of sdrBP (blue) with the RRM of CstF-64 (pink) showing how similar these proteins are even in the orientation of the loops, with one significant exception at two N- and C-terminal helices of the RRM of CstF-64 which are not found in sdrBP. (e) Proposed RNP1 (RGFGFCEF) and RNP2 (VFVGNL) sequences of sdrBP are located in the two central β -strands, β_3 and β_1 respectively

exposing three aromatic residues (RNP sequences have been extended in the loops as well).

Results & Discussion:

cDNA sequencing

Upon plasmid isolation and restriction enzyme digestion of the twenty recombinant clones from *S. digitata* cDNA library, eighteen clones had inserts of more than 500 bp in length and their nucleotide sequences were obtained.

Nucleotide sequence analysis

In nucleotide database searches, a 549 bp long cDNA showed 82% nucleotide identity to complete coding sequences of *B. malayi* putative RNA binding protein (bm-rbp-1) and *W. bancrofti* putative RNA binding protein. The full-length cDNA containing a poly A tail and a hypothetical polyadenylation signal sequence (AATAAA) was named sdrbp for *S. digitata* RNA binding protein. The entire sdrbp sequence contains four ORFs and sequence analysis of the shortest ORF (249 bp) reveals an initiation codon (ATG) located within the Kozak sequence context, GAAAACATGT).

Amino acid sequence analysis

The shortest ORF encodes an 82 amino acid polypeptide which contains a RRM domain. In protein database searches, the polypeptide shows 100% identity to RNA binding protein - identical [*B. malayi*] and hypothetical protein - partial [*W. bancrofti*] and 54% identity to the RRM of human CstF-64.

Homology Modelling and structural analysis

Homology modelling of the hypothetical protein, sdrBP by SWISS-MODEL Workspace resulted in three PDB files, out of which the best quality model (Figure 1a) built using RRM domain (Figure 1b) of human CstF-64 (Figure 1c) as template was chosen based on QMEAN scoring function and sequence identity. sdrBP adopts classical $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$ RRM-topology forming four stranded β -sheet packed against two α -helices.

Structure validation and structural comparison of sdrBP

sdrBP model built was validated using PROCHECK server which produced Ramachandran plot calculations and plot statistics for Φ and Ψ distributions for non-glycine non-proline residues. Similar results were obtained for RRM of CstF-64 template (1P1T_A) suggesting sdrBP model built is reliable. For structural comparison, sdrBP was superimposed with RRM of CstF-64 (Figure 1d) and an RMSD of 0.084 Å was obtained. Reasonably, a low RMSD value indicates proper structural and functional similarities between the two proteins; so it is considered as the best model.

Assessment of structure-function relationship

The template protein, cleavage stimulation factor (CstF) is a heterotrimeric protein complex which plays a key role in polyadenylation of mRNA precursors. In vertebrates, polyadenylation site comprises two elements: AAUAAA hexamer and a more diffuse GU-rich sequence. GU-rich sequence is recognized by the 64 kDa subunit of CstF (CstF-64). CstF-64 is a multi-domain protein and the RRM is a 104 amino acid long domain located at the N-terminus of the protein. Studies have revealed that CstF-64 forms more stable complexes with GU-rich sequences containing at least two consecutive us [13].

Classical RRM domain is identified at the primary sequence level as a 90 amino acids long polypeptide containing two conserved sequences of eight and six amino acids called RNP1 ([R/K]-G-[F/Y]-[G/A]-[F/Y]-[I/L/V]-X-[F/Y]) and RNP2 ([I/L/V]-[F/Y]-[I/L/V]-X-N-L) located in the two central β -strands, $\beta 3$ and $\beta 1$ of the motif respectively exposing three conserved aromatic residues [14]. In the proposed RNP1 (RGFGFCEF) and RNP2 (VFVGNL) sequences of sdrBP, all amino acids except for one residue are conserved and expected to form the primary RNA binding surface (Figure 1e & 2a). Four key amino acids of the RRM of human CstF-64 reported to be essential for RNA binding activity are conserved in sdrBP in similar positions (Figure 2b).

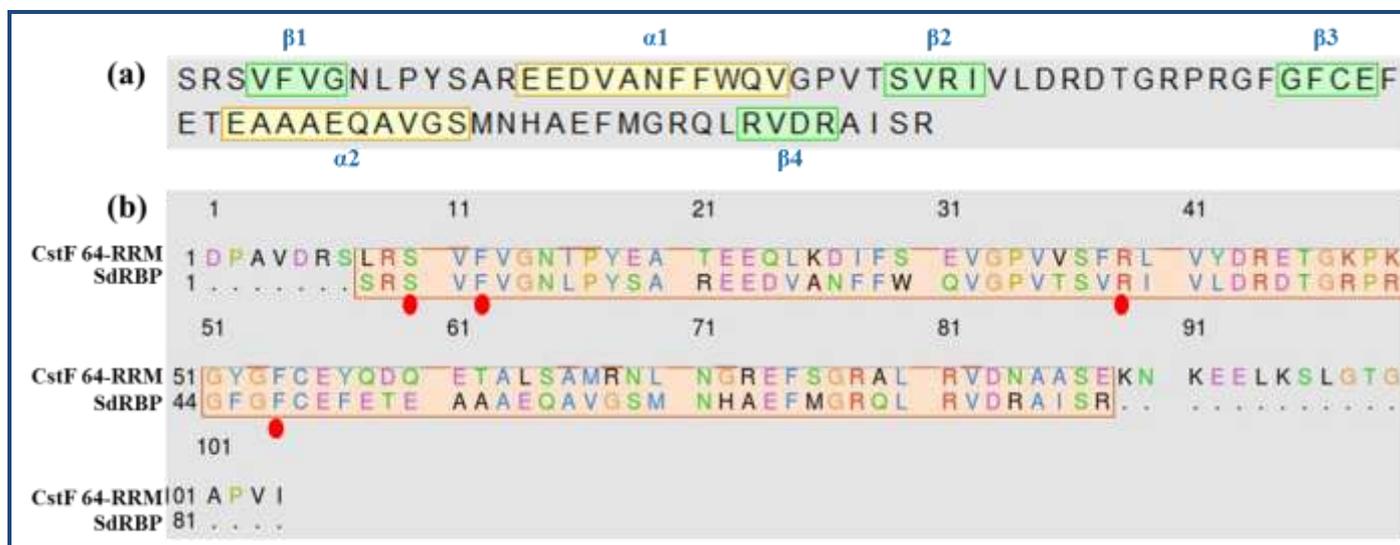


Figure 2: (a) 82 amino acid polypeptide (sdrBP) derived by the conceptual translation of the shortest ORF (249 bp) showing the locations and sequences of four β strands (green) and two α helices (yellow). (b) Multiple alignment of sdrBP with CstF-64 using UCSF chimera results in 54% amino acid identity and the four key amino acids of the RRM of CstF-64 (S10, F12, R39, and F54) reported to be essential for RNA binding activity are conserved in sdrBP (conserved positions are indicated in the alignment by red spots).

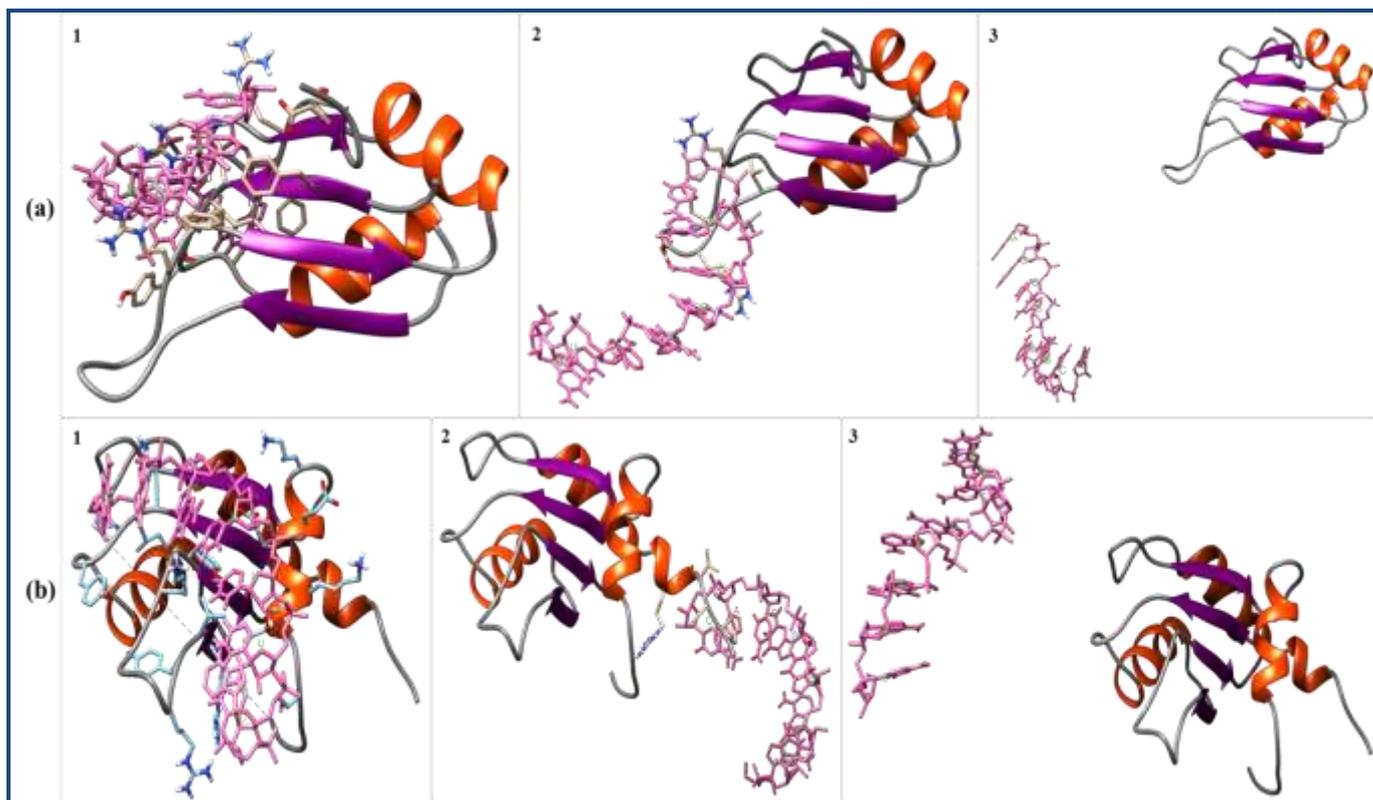


Figure 3: (a) Automated molecular docking of sdrBP with three types of RNA octamers shown in pink (1: 5'GUGUUUAC-3' [472D_A], 2: 5'GUAGGCAC-3' [472D_B], 3: 5'- GGCGAGCC)-3' [1SA9]) using Hex 8.0.0 Cuda software (viewed by UCSF Chimera) showing sdrBP has greater binding affinity to GU-rich RNA which comprises of consecutive Us (a.1). (b) Automated molecular docking of RRM of CstF-64 with the same three types of RNA octamers results in comparable results as obtained for sdrBP revealing the fact that sdrBP has significant functional similarity with human CstF-64.

Automated molecular docking

Consistent with structure-function similarities, automated molecular docking studies have demonstrated that sdrBP has a greater binding affinity to GU-rich RNA (**Figure 3a**). Upon docking of sdrBP with the three types of RNA octamers, larger energy values **Table 1** (see **supplementary material**) were obtained for GU-rich RNA which comprises of consecutive Us and similar results were obtained for docking of RRM of CstF-64 with same RNA octamers (**Figure 3b**). Since RRM of CstF-64 and sdrBP share 54% sequence identity and their docking energy values are similar, this is an evidence to believe that sdrBP has similar biological function as that for CstF-64.

Validation of various physiological parameters

The functional similarity was further validated by ProtParam which reveals sdrBP to have a theoretical pI of 5.75 and estimated half-life to be >20 hours (yeast, *in vivo*). The instability index (II) is computed to be 30.60 and this classifies the protein as stable. Grand average of hydropathicity (GRAVY) is -0.293 and aliphatic index of sdrBP is 65.37. As mentioned in the ProtParam tool, aliphatic index is regarded as a positive factor for the increase of thermo stability of globular proteins.

Conclusion:

In silico structural modelling and automated molecular docking of *S. digitata* predicted protein using different bioinformatics tools, suggest that it has similar structural and functional features as human CstF-64 and therefore, predicted

to be involved in pre-mRNA polyadenylation process. Therefore, this study brings new dimensions to the functional analysis of RNA binding proteins of *S. digitata* and their evaluation as new drug targets against HLF.

Acknowledgement:

Financial support from International Programme in Chemical Sciences, University of Uppsala, Sweden and sida Secretariat for Research Cooperation with Developing countries (formerly SAREC) are gratefully acknowledged. We would like to thank Kanchana Senanayake for providing computational assistance and Anoma jayasoma for supporting in DNA sequencing.

References:

- [1] Sharma OP *et al.* *J Vector Borne Dis.* 2013 **50**: 155 [PMID: 24220073]
- [2] Schwab AE *et al.* *Am J Trop Med Hyg.* 2005 **73**: 234 [PMID: 16103581]
- [3] <http://repo.jfn.ac.lk/med/handle/701/925>
- [4] <http://www.ncbi.nlm.nih.gov/tools/vecscreen/>
- [5] https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome
- [6] <https://www.ncbi.nlm.nih.gov/gorf/gorf.html>
- [7] <http://swissmodel.expasy.org/workspace/>
- [8] www.cgl.ucsf.edu/chimera/
- [9] www.ebi.ac.uk/thornton-srv/software/PROCHECK/
- [10] <http://www.rcsb.org/pdb/home/home.do>
- [11] <http://hex.loria.fr/>
- [12] <http://web.expasy.org/protparam/>

[13] Pérez Cañadillas JM & Varani G, *EMBO J.* 2003 **22**: 2821
[PMID: 12773396]

[14] Cléry A *et al.* *Curr Opin Struc Biol.* 2008 **18**: 290 [PMID:
18515081]

Edited by P Kanguane

Citation: Nagaratnam *et al.* Bioinformation 10(8): 512-517 (2014)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Table 1: Energy values obtained after each docking event

Ligand - RNA Octamers	sdRBP	1P1T_A
472D_A: 5'-R(*GP*UP*GP*UP*UP*UP*AP*C)-3'	E: -526.12	E: -502.98
472D_B: 5'-R(*GP*UP*AP*GP*GP*CP*AP*C)-3'	E: -249.51	E: -182.76
1SA9: 5'-R(*GP*GP*CP*GP*AP*GP*CP*C)-3'	E: -25.56	E: -83.75