International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

Insilico Analysis of a Rice SR Related Protein, SRCTD-6 Reveals a Splicing Function

Simmanna Nakka[1], Uday Bhaskar Sajja[2]

Department of Biotechnology, Dr. B. R. Ambedkar University, Srikakulam, Etcherla, Andhra Pradesh, India-532410

Abstract: ***RS domain containing proteins are associated with binding to the exons of nascent primary transcript and recruiting components of spliceosome for precise recognition of the splice junctions. This is achieved with the help of RRM domain by which the protein binds with the RNA. However there are proteins which share the domain characteristics of SR proteins but lacking the RRMs. These proteins are grouped under SR related proteins. Here we described a SR Related CTD associate factor-6 (SRCTD-6) which contains unique set of domains. Insilco analysis of this protein identified a hydrophilic signature with coiled domains and 2 distinct helices. Domain characterization revealed the presence of an RPR domain, an RS domain in the C-terminal and stretches of Prolines and Glutamines in the N-terminus. Structural analysis of this protein hypothesise that it is recruited at the splice junction by Heptad repeats of C-terminal domain of RNA polymerase II. We predict that SRCTD-6 has a general role in identifying the precise 5^I and 3^I splice junctions.***

Keywords: spliceosome, RPR-Domain, homology modelling, hydropathacity

1. Introduction

In most of the Eukaryotes including Plants majority of the genes are interrupted with Introns which have to be precisely excised from Primary m-RNA transcript to give rise to a functional mature m-RNA transcript. This process is achieved with the help of a RNA-Protein complex called spliceosome [1], [2]. Multiple transcripts from a single gene can be achieved by a regulated mechanism called Alternative splicing. Alternative splicing events in rice (OryzaSativa) in more than 50% of genes produce a variety of transcripts [3]. A number of associative proteins act as both positive and negative regulators of splicing. The Serine-Arginine (SR) splicing factors are highly conserved family of RNA-binding proteins which participate in spliceosome assembly at the splice junctions [4]. The N Terminus of these proteins typically have one or two RNA Recognition Motifs (RRMs). The C-terminal regions of these proteins have RS domain enriched in Arg-Ser (or Ser-Arg) repeats [5]. In addition to SR family, other proteins have been identified which may or may not have a RRM domain, but containing an RS domain. These proteins are collectively referred as SR-Related Proteins [6].

Biochemical and Bio-Informatics approaches have demonstrated the presence of Exonic Splicing Enhancers (ESEs) in the exonic and intronic sequences. It has been proposed that RS domain of SR proteins binds to ESE and interacts directly with RS domain of other splicing factors and there by recruiting the spliceosomal components such as U1snRNP to 5^I splice site or U2AF to 3^I splice site [7].

The processing of the 1^{st} exon is thought to be mediated by the interaction between the cap binding complex and the spliceosome. The processing of the last exon is thought to be mediated by the interaction of Poly adenylated complex and the spliceosome [8]. Hence, capping, splicing and 3^I end processing are interconnected. CarboxyTerminal Domain (CTD) of RNA Polymerase-II achieves all these processes by recruiting all essential factors. CTD of Eukaryotic RNA Polymerase-II consists of conserved Heptad repeats with consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser [9]. These repeats are the major sites for reversible phosphorylation events. CTD of RNA Polymerase-II interacts with a plethora of proteins having role in transcription, processing, and termination of primary transcripts. Many of the interacting proteins contain a conserved domain called RPR-domain (Nuclear RNA processing Region).

In this study, we are describing a unique protein containing are Arginine and Serine de-peptides in the C-terminal region. However, this protein does not contain a RRM to be classified under SR proteins. Hence, this protein is called an SR-Related protein. Instead, this protein contains an RPR domain which in previous studies have been showed to be interacting with CTD of RNA Polymerase-II. The presence of RPR domain and SR rich region suggest a unique role for this protein in the splicing process. Hence, an attempt was made using Bioinformatics tools to elucidate the possible cellular function of this protein.

2. Methodology

Datasets-SR Related CTD associated factor-6 protein sequence from rice (OS06g0682700) and other sequences used in this study were retrieved from the public databases, http://www.ebi.ac.uk and http://www.ncbi.nlm.nih.gov. For experimentally determining the 3D structure of SR Related CTD associated factor-6, structural homologous subsets were retrieved from PDB (Protein Data Bank).

Pattern Recognition-Analysis of protein conserved domains and motifs were carried out using Interpro scan based on the PROSITE and Pfam data bases (http://www.ebi.ac.uk/inter) [10], [12]. The protein was further analysed using the tools available in the EXPasy server (http://www.expasy.org/tools [13]. A Physiochemical property of the selected protein was determined using the protparam tools. Hydropathy analysis of the selected protein was done based on Kyte and Doolittle [14], [15].

Homology modelling-PELE program was used to performthe secondary structure analysis (SDSC Biology workbench) (http://workbench.sdsc.edu). GLOBPLOT analysis based on Lindngi values was used to identify intrinsic disorders in the protein sequence [16]. Presences of Nuclear Localization Signals were checked using the NucPred program [17] SWISSPDB homology modelling was used to generate Protein models after aligning to the structural homologues. PROCHECK was used to check the quality of the protein model [18] For visual display of the models SWISSPDB viewer was used [19].

3. Results and Discussions

Primary protein properties of SR-Related CTD associated factor-6 (SRCTD-6): Insilico analysis of SRCTD-6 primary protein using PROSITE program reveals that it is Proline rich protein which constitute up to 17.2% if the total amino acids. These proteins are observed particularly in N-terminal region of the protein. Towards the C-terminus the protein is rich in amino acids serine and Arginine with RS (Arginine/Serine dipeptides) repeating twice. It further reveals that it is an unstable protein and estimated the half-life of 30 hours (Table-1).

Table 1: Summary table showing the SR related CTD associated factor 6 primary protein physiochemical properties

	Property	Summary
1	Number of Amino acids	627
2	Theoritical Pi	6.02
3	Amino acid composition (Three most abundant Amino acids)	Pro (P) 108 – 17.2% Ala (A) 51 – 8.1% Ser (S) 47 – 7.5%
4	Negatively charged residues (Asp+Glu)	61
5	Positively charged residues (Arg+Lys)	49
6	Instability index	66.80 (Protein is unstable)
7	Estimated half life	30 hrs
8	Grand Average of hydrophathicity (GRAVY)	- 0.627

Previous studies have shown that the regions containing stretch of Prolines and Glutamines participate in Protein-Protein interaction. SR domain towards the C-terminus of the protein is largely involved in interaction with other SR domain containing proteins. These observations suggest that this Protein might form a complex with other proteins. Interpro scan carried out revealed in addition to SR domain the protein also has an RPR domain (aminoacids 234-264) (figure 1).

Figure 1: Protein sequence of SR related CTD associated factor 6 showing stretches of Glutamin's and Prolin's (Red and Blue respectively) in the N-Terminus, RPR domain from position 234-364 shown in brown and SR amino acids in the C-Terminus shown in green.

The presence of SR domain suggests SRCTD-6 is a splicing factor. When SRCTD-6 was scanned for Nuclear localized signals using NucPred programme none were found (Figure 2). Though this protein does not have recognisable Nuclear localized signals as it is a splicing factor it is possible that it is localized to the nucleus on Piggy-back with other Nuclear localizing proteins.

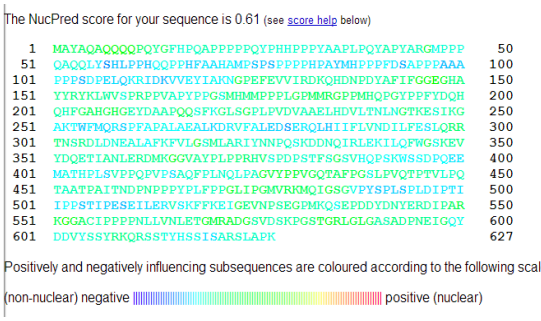

Figure 2: SR related CTD binding factor 6 was scanned for Nuclear Localization Signals using the NucPred program. However no NLS were found in the sequence.

SRCTD-6 Protein is Hydrophilic: Hydropathy plot revealed that SRCTD-6 protein is showing Hydrophilic nature with more than 70% of residues falling in that region with a negative score (figure 3). The overall average of Hydropathacity (GRAVY) of SRCTD-6 is-0.625, suggesting that this protein is hydrated in aqueous environment.

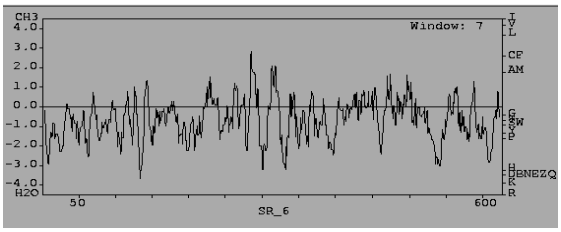

Figure 3: Hydropathy analysis of predicted protein SR related CTD associated factor 6 based on Kyte and Doolittle values. The residues showing hydropathy value below zero are hydrophilic in nature. Hydropathy plot shows residues from 233-364 in box representing the presence of a CID domain in that region.

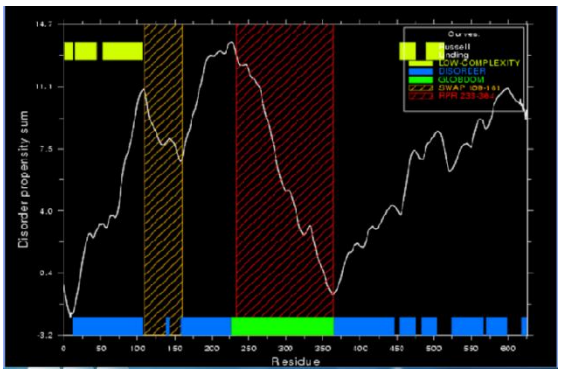

Figure 5: GLOBPLOT analysis of SR related CTD interacting factor 6 showing the regions of intrinsic disorders. Also conserved RPR domain (CID) from 233-264residues in the sequence

The structure prediction programme and intrinsic disorder analysis suggests that SRCTD-6 protein is a well refined structure at the given resolution. Theresidues in the most favoured region of SRCTD-6 is 86.4%. Thus, SRCTD-6 model deduced is a good hypothetical protein model. The Homology model of SRCTD-6 that is generated will aid to determine the mechanistic functions of SR related protein class.

Conserved Domains in the SRCTD-6 protein: This protein has RPR domain (position 224-264) which in previous studies shown to be interacting with C-terminal tail of RNA Polymerase-II (Figure 5). Aspartic acid (D) present in the DSI motif of RPR domain forms a covalent bond with Tyrosine (Y) present in the Heptad repeats of C-terminal tail of RNA Polymerase-II. DSI motif in SRCTD-6 is conserved except for the third amino acid where Glutamine (E) is present instead of Isoleucine (I). As Aspartic acid is intact in the motif we can predict that this RPR domain might interact with Heptad repeats of CTD of RNA Polymerase-II.

The C-terminal region of this protein is fairly rich in Arginine and Serine amino acids. Two Arginine-Serine dipeptides (RS) were observed with in this region. These dipeptides are essential for binding to the ESE cis element present in the Exons or for interacting with other proteins containing RS domain. As this protein lacks an RRM it is highly possible that this protein might not be interacting directly with the primary m-RNA transcript. However, the presence of RPR domain, RS domain and the N-terminal region rich in proteins suggest that this protein might form a recognition complex having a role in precise recognition of splice junctions of nascent primary transcript.

Homology Modelling: At a similar resolution, the homology model validation with PROCHEK essentially satisfied stereo chemical parameters with well refined structures. This distribution of residues in the most favoured region of SRCTD-6 is 86.4%. Thus, the hypothetical protein model predicted here for SRCTD-6 is fairly good. The Homology model of SRCTD-6 thus generated in this study will help in determining the the mechanistic functions of SR related protein class.

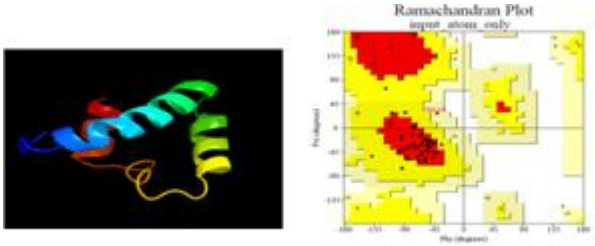

Figure 6: (a) Predicted 3D structure of SR CTD-6 proteins. (b) Ramachandran plot showing the residues of SRCTD – 6 falling in the most favoured region and other allowed regions

4. Conclusions

SRCTD-6 protein is a rice protein with unique domain architecture. The presence of RPR domain in addition to the RS domain gives it a unique role for this protein. The nature of this protein is hydrophilic with predominantly random coil arrangement along with a helical conformation. This protein does not bind to the primary m-RNA transcript because of a lack of RRM domain. However, the presence of RPR domain suggests an interaction with CTD of RNA Polymerase-II between Aspartic acid of RPR domain and Tyrosine of CTD. This possible interaction suggests that it is recruited to the splice junction by the C-Terminal Domain of RNA Polymerase-II. The presence of RS domain and in the splice junction SRCTD-6 might act as a bridge between components of spliceosome and other SR proteins.

References

[1] Roy SW, Irimia M. Splicing in the eukaryotic ancestor: form, function and dysfunction. TrendsEcolEvol 2009, 24: 447–455
[2] Wahl, M. C., Will, C. L. and Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. Cell, 2009, 136, 701–718.
[3] Zhang G, Guo G, Hu X, Zhang Y, Li Q, Li R, Zhuang R, Lu Z, He Z, Fang X, et al. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. Genome Res20, 2010: 646–654.
[4] Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. Biochem 2009, J417: 15–27
[5] Haynes C, Iakoucheva LM. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. Nucleic Acids Res 2006, 34: 305–312.
[6] Boucher L, Ouzounis CA, Enright AJ, Blencowe BJ. A genome-wide survey of RS domain proteins. RNA 2001, 7: pp 1693–1701
[7] Graveley BR. Sorting out the complexity of SR protein functions. RNA.2000; 6: 1197–1211.
[8] Izaurralde E, Lewis J, McGuigan C, Jankowska M, Darzynkiewicz E, Mattaj IW. A nuclear cap binding protein complex involved in premRNA splicing. Cell.1994; 78: 657–668.
[9] Stiller, J. W. & Hall, B. D. *Evolution of the RNA polymerase II C-terminal domain. ProcNatlAcadSci U S A 2002, 99, 6091–6096.*
[10] Mulder N. J., Apweiler R., Attwood T. K., Bairoch A., Bateman A., Binns D., Bradley P., Bork P., Bucher P., Cerutti L., et al. InterPro, progress and status in 2005. Nucleic Acids Res.2005; 33: D201–D205.
[11] De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. (*ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.* Nucleic Acids Res.2006 Jul 1; 34 (Web Server issue): W362-5.
[12] Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. *New and continuing developments at PROSITE* Nucleic Acids Res.2012; doi: 10.1093/nar/gks1067
[13] Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M. R., Appel R. D., Bairoch A.; *Protein Identification and Analysis Tools on the ExPASy Server;*
(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005) pp.571-607

[14] Kyte, J & Doolittle, R. F. A simple method for displaying the Hydropathic characters of a protein. J. Mol. Biol.1982, 157: 105-132
[15] Gaboriaud C, Bissery V, Benchetrit T, Mornon J. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. FEBS Lett.1987; 224: 149–155
[16] Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, et al. Protein disorder prediction: implications for structural proteomics. Structure, 2003, 11: 1453–1459. doi: 10.1016/j. str.2003.10.002
[17] Nguyen Ba AN, Pogoutse A, Provart N, Moses AM. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. BMCBioinformatics.2009; 10: 202. doi: 10.1186/1471-2105-10-202.
[18] Laskowski R A, MacArthur M W, Thornton J M. PROCHECK: validation of protein structure coordinates, in *International Tables of Crystallography, Volume F. Crystallography of Biological Macromolecules*, eds. Rossmann M G & Arnold E, Dordrecht, Kluwer Academic Publishers, The Netherlands, 2001, pp.722-725
[19] Johansson, M. U., Zoete V., Michielin O. &Guex N. Defining and searching for structural motifs using Deep View/Swiss-Pdb viewer BMC Bioinformatics, 2012, 13: 173.