CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR (CPSF) SUBUNIT INTERACTION IN TOMATO, Solanum lycopersicum

Saba Riaz1,*, Munir Ahmed Sheikh1, Muhammad Shahid1 and Abdul Wahid2

1Department of Biochemistry, University of Agriculture Faisalabad, Pakistan;
2Department of Botany, University of Agriculture Faisalabad, Pakistan.
*Corresponding author's e-mail: sabi_aem@yahoo.com

Cleavage and polyadenylation is an important step in eukaryotic mRNA biogenesis that also controls gene expression. The process undergoes interaction of poly(A) factors with the poly(A) sites and with each other. Among different poly(A) factors, Cleavage and Polyadenylation Specificity Factor (CPSF) has a central role in the process. Tomato, Solanum lycopersicum, is an important member of the Solanaceae family and grown worldwide with high production. We investigated interaction of different CPSF subunits with each other and with some other proteins using TAP (tandem affinity purification) method in tomato. Four CPSF subunit genes were transformed in Agrobacterium followed by transformation in tomato callus. The interaction was studied by proteomics approach. Proteomics results have shown the interactions among CPSF subunits with each other and with other neighboring proteins that may have important implications in gene regulation eventually leading to the control of many biological processes such as flowering time. This further demonstrates that the 3′-end formation of mRNA involves extensive protein-protein interaction network which is coupled with other processes of mRNA biogenesis such as transcription, capping, splicing and translation.

Keywords: CPSF, tomato, poly(A) factors, proteomics

INTRODUCTION

Cleavage and polyadenylation, sometimes referred to as polyadenylation, of pre-mRNA is a crucial step for producing mature mRNA (Tian and Graber, 2012). It is composed of a tract of adenosines (polyA) to the cleaved RNA (Zhao et al., 2009). The polyadenylation is not only an essential step for the production of correctly processed mRNA but also has other crucial functions (Koh and Wong, 2007). The polyadenylation process requires two essential components; the cis-elements on mRNA and trans-elements in the form of protein factors that recognize and interact with the cis-elements to complete the process (Hunt et al., 2008).

A macromolecular machinery comprising of more than twenty proteins bring about 3′-end processing of mRNA (Chan et al., 2011). The trans factors include multiple poly(A) factors organized into complexes. Among these factors CPSF multisubunit complexes have the central role in the polyadenylation process. These subunits are classified on the basis of molecular masses which are CPSF30, CPSF 73, CPSF 100 and CPSF 160 (Cohen et al., 1999). In Arabidopsis the CPSF 73 has been further divided into CPSF73-I and CPSF73-II (Xu et al., 2006). The CPSF subunits in plants are not only involved in cleavage and polyadenylation process but also found to have different roles in plant metabolism (Xu et al., 2004).

The polyadenylation process in plants has mostly been reported from the model plants Arabidopsis and rice that mainly includes bioinformatics analysis rather than experimental work (Li et al., 2008). Tomato, Solanum lycopersicum L., is a model plant of Solanaceae family (Aoki et al., 2010) that comprises more than 3000 species including potato, eggplant, peppers etc (Mueller et al., 2005a). However, many important molecular processes, including 3′-end processing of mRNA, have still not been studied in tomato. Therefore, a project was designed to study CPSF gene expression pattern and CPSF subunits interaction in this important plant, and are reported in this paper.

MATERIALS AND METHODS

Germination of tomato seeds: Tomato (Solanum lycopersicum L.) seeds, var. Roma, were procured from Ayub Agriculture Research Institute, Faisalabad, Pakistan and soaked in 5% sodium hypochlorite for 15 min under sterilized conditions followed by washing with distilled autoclaved water for 6 times at 5 min intervals and then at 20 min intervals twice. The seeds were placed in autoclaved Petri plates having Whatman filter paper soaked in distilled autoclaved water under sterilized conditions and germinated under sterile conditions on wet filter paper at 28°C in growth chamber (Artificial Climat Chamber [RTOP series]).

CPSF subunit interaction study by proteomic approach: Construction of TAP-tagged vectors with CPSF genes: The TAP (Tandem Affinity Purification) tag (TAPI) vectors (Rohila et al., 2004) were very kindly gifted by Quinn Li, Department of Botany, Miami University, Oxford, Ohio,
USA. It was Gateway-compatible binary vector. The CPSF gene sequences (cDNA) were first cloned into pENTR vector (Invitrogen) following manufacturer's protocol. To a 1.5 mL microcentrifuge tube 1 µL of cDNA sequence (50 fmol) was added to 2 µL pDONR 221 vector (150 ng/µL), 4 µL 5X BP clone reaction buffer and made up to 16 µL with TE buffer, pH 8.0. Added 4 µL of the BP clone enzyme, mixed well and incubated at 25°C for 60 min. Added 2 µL of 2 µg/µL protease K and incubated at 37°C for 10 min. The ligated mixture was transformed in competent E. coli cells and selected on kanamycin. The ENTRY clones containing the CPSF genes were subjected to LR recombination in nTAPl and cTAPl vectors. The plasmids were isolated with the help of GeneJet Plasmid miniprep kit (Fermentas) following manufacturer's protocol.

Transformation of CPSF-TAP tagged vectors in Agrobacterium tumefaciens: The CPSF-TAP-tagged vectors were transformed in Agrobacterium tumefaciens LBA4404 by electroporation in AARI, Faisalabad.

Transformation of vectors in tomato: To prepare tomato suspension culture approximately 5-10 mm diameter segments of the tomato leaves were excised and transferred to the callus induction medium containing Murashige and Skoog basal salts, Nitsch's vitamins, myo-inositol (100 mg/L) and 3% sucrose supplemented with plant growth regulators 2,4-dichlorophenoxyacetic acid (2 mg/L) and 6-benzylaminopurine (0.1 mg/L) and solidified agar. The suspension culture was prepared following Lu et al. (2008). The transformation was done using Agrobacterium transformation by employing the method reported by Zhao et al. (2009).

TAP purification and MS analysis: TAP purification was done after Rohila et al. (2004). Briefly, about 100 grams of cells (fresh weight) were ground in liquid nitrogen with acid washed quartz to a fine powder. Extraction was carried out in extraction buffer (20 mMTris/HCl, pH 8.0, 150 mM NaCl, 0.1% NP-40, 2.5 mM EDTA, 10 mM β-mercaptoethanol, 1 mM phenylmethylsulphonyl fluoride (PMSF), 2 mM Benzamide, 20 mM NaF, and 0.1% [V/V] of protease inhibitors cocktail [Sigma]). The mixture was passed through Ig G beads and calmodulin affinity beads. Finally, the proteins were eluted with 1 to 1.5 mL of elution buffer (10 mM Tris/HCl, 150 mM NaCl, 1 mM Mg-acetate, 1 mM imidazole, 2 mM ethylene glycol tetraacetic acid (EGTA), 0.1% NP40 (V/V), 10 mM β-mercaptoethanol, 1 mM PMSF and 1/1000 protease inhibitor cocktail, pH 8.0) into the desired number of fractions, which were monitored by absorbance at 280 nm. Eluted proteins were precipitated by trichloroacetic acid (TCA) and sodium deoxycholate (DOC) before loading onto a SDS-PAGE gel. The proteins were analyzed by mass spectrometry after trypsin digestion from the proteomics facility at Danforth Plant Science Center, USA.

RESULTS

Proteomics approach was used to study the CPSF subunit interaction. Tandem Affinity Purification (TAP) system of (Rohila et al., 2004) was applied to study the interaction. A TAP tag is attached with the gene of interest and transformed in plant cells. It helps purification of protein complexes containing proteins of interest attached to the TAP-tag. The TAP tag comprises of Protein A domains having affinity with IgG-agarose, A tobacco etch virus (TEV) protease site so that the fusion protein may be cleaved from the IgG-agarose and A calmodulin-binding protein domain. The protein complexes bound to the bait protein fused to TAP-tag are isolated from the plant cells transformed with the TAP-tagged vectors. The tandem protein A domains attached to the TAP-tagged protein complexes bind to IgG-agarose, and the unbound proteins are washed out. The TAP-tagged protein complexes are separated from IgG-agarose by cleaving with TEV protease at the TEV-protease cleavage site. The remaining TAP-tagged protein complexes are applied to the calmodulin column where calmodulin-binding domain of the TAP-tag binds in the presence of Ca²⁺. This gives further purification of the desired protein complexes as the unbound proteins are washed out again. The TAP-tagged protein complexes are relieved from calmodulin binding with the help of EGTA that chelates out Ca²⁺ ions, hence the interaction between calmodulin and TAP-tagged protein complexes is disrupted. The protein complexes thus isolated are subjected to mass spectrometric analysis to identify the proteins in the protein complexes. The CPSF subunit genes were introduced in TAP vectors nTAPl and cTAPl (Fig. 1). The CPSF100 was fused to TAP tag at N-terminus whereas all other subunit genes were ligated in cTAPl. The CPSF subunit genes were under the control of cauliflower mosaic virus (CaMV) 35S promoter and CaMV terminator. The CPSF subunits in the TAP vectors were transformed in E. coli followed by miniplasmid preparation. The plasmids were digested with HindIII for confirmation of the ligation of CPSF subunits in the plasmids. All the plasmids showed positive results as expected fragment sizes (Table 1) were obtained after the restriction digestion except for CPSF73II that could not be transformed due to some unknown reason.

Table 1. CPSF gene sizes and restricted fragment sizes after miniplasmid preparation

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Size (bp)</th>
<th>Plasmid</th>
<th>Fragment size after HindIII restriction (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPSF30</td>
<td>753</td>
<td>CTAPi-CPSF30</td>
<td>2514</td>
</tr>
<tr>
<td>CPSF73</td>
<td>2338</td>
<td>CTAPi-CPSF73</td>
<td>4099</td>
</tr>
<tr>
<td>CPSF100</td>
<td>2417</td>
<td>NTAPi-CPSF100</td>
<td>4178</td>
</tr>
<tr>
<td>CPSF160</td>
<td>4329</td>
<td>CTAPi-CPSF160</td>
<td>6090</td>
</tr>
</tbody>
</table>
After confirmation the plasmids were transformed in *Agrobacterium tumefaciens* LBA4404 and selected with spectinomycin antibiotic. The positive colonies were picked and the CPSF subunit genes were confirmed by colony PCR. The successful transformants were transformed in tomato cells by electroporation and the TAP-tagged protein complexes were isolated by passing the cell lysate over IgG-agarose column, cleaved at TEV cleavage site and then passing over calmodulin-Ca$^{2+}$ column. The protein complexes were sequenced by mass spectrometric analysis from USA. The poly(A) factors and other proteins identified were tagged with each of the TAP-tagged vector are given in Table 2.

**DISCUSSION**

The interaction with CPSF subunits was studied by proteomics analysis of the proteins purified from tomato cell suspension after transformation with TAP-tagged-CPSF gene vectors. Each of the TAP-tagged fraction fractionates the proteins that have interaction with the CPSF subunit and with other proteins. As seen in Figure 2 the CPSF30 had interaction with FIPS5 which in turn interacts with poly(A) polymerase (PAP) and stimulates it (Forbes et al., 2006). The CPSF30 also had interaction with CPSF73II that has been shown to have functions other than housekeeping (Xu et al., 2004). The CPSF73II interacts with CPSF73I. The CPSF73I also has interaction with CPSF100, CPSF160, FY protein, CLPS3 (cleavage family complex protein) and PCFS4 (protein 1 of cleavage factor). CPSF100 also has interaction with CPSF160 and FY, and CPSF160 with FY, PAP and SYM5 (symplekin). Symplekin is also involved in post-transcriptional gene silencing by RNA (Herr et al., 2006) and has interaction with CstF64. Another poly(A) factor CLPS3 (CLP1-Similar Protein 3), an ortholog of human CLP1, was also found to be essential for embryo development and female gamete transmission in Arabidopsis (Xiang et al., 2008). (Zhao et al., 2009) investigated CPSF *in vivo* in *A. thaliana* by proteomic studies using tandem affinity purification for isolation of the protein complexes followed by mass spectrometry and Western immunoblots. They found all the above mentioned CPSF subunits plus AtFY and AtFIPS5. They further demonstrated that AtCPSF100 served as a core to which all other factors were associated except AtFIPS5. Similarly, AtCPSF73I has been shown to be an endonuclease having a role in cleavage of the mRNA during 3'end formation (Mandel et al., 2006). CPSF30 has also been shown as an RNA-binding endonuclease associated with other factors (Rao et al., 2009). Moreover, it was found that the AtCPSF30 itself was localized to the cytoplasm when expressed by itself; however, it was found to be localized in nucleus when co-expressed with other factors such as AtCPSF160 and AtCPSF73 probably due to interaction. (Rao et al., 2009) further demonstrated that the CPSF30 was speculated to have multiple novel protein-protein interactions among CPSF factors. (Delaney et al., 2006)
studied CPSF30 interactions using calmodulin binding assay and found that the CPSF30 was not only capable of self-interactions but also potentially involved in possible interplay between splicing and polyadenylation and their regulation. In another study (Addepalli and Hunt, 2007) the CPSF30 was found to have zinc finger motifs, and it was indicated that interactions of the protein were probably via Fip1. The CCCH zinc finger motif of AtCPSF30 has also been shown to have a disulphide linkage (Addepalli et al., 2010). Fip1 (Factor Interacting with poly(A) polymerase) is another protein that is known to provide a bridge between poly(A) polymerase, poly(A) apparatus and RNA in eukaryotes (Forbes et al., 2006; Meinke et al., 2008); rather it was reported that the Fip1 from Arabidopsis had plant-specific interactions between the protein and other factors. The CPSF100 gene is ca 2400 nucleotide long and encodes a polypeptide of 100 kDa (Elliott et al., 2003). The factor was found not only in Arabidopsis but the nuclear extracts from pea and wheat germ also recognized the antibodies raised against the Arabidopsis CPSF100. Moreover, it was found that 60 amino acid domain of AtCPSF100 had an interaction with 220 amino acids of poly(A) polymerase. Dominski et al. (2005) demonstrated that the CPSF73 and CPSF100 belonged to a superfamily of zinc-dependent β-lactamase fold proteins. Several RNA binding proteins other than poly(A) factors have been reported to influence different post-transcriptional processes including splicing, RNA decay, polyadenylation etc (Staiger and Koster, 2011). Among these Fip1 is a subunit of the poly(A) apparatus and acts as a bridge between poly(A) polymerase, other poly(A) subunits and substrate RNA (Forbes et al., 2006; Meinke et al., 2008). The protein was found to have 1196 amino acids, and 137 amino acids had interaction with poly(A) polymerase. Another RNA binding protein FCA has also been shown to interact with 3'end processing machinery of mRNA (Simpson et al., 2010). It was found to be an alternative polyadenylation regulator. The FCA expression was speculated to have its role in regulating flowering time in Arabidopsis. Whole-genome tiling array studies regarding FCA and FPA proteins revealed that the proteins had an important role in 3'end processing and transcription termination in Arabidopsis (Sonmez et al., 2011). The FCA promotes flowering and autoregulates its own expression via alternative polyadenylation (Quesada et al., 2005) in Arabidopsis. Physical interaction of FCA with a 3'end

**Figure 2. Interaction among CPSF subunits and other protein factors in tomato.**

The interactions were determined by the proteomics analysis. The protein factors are drawn to nearest possible scale. Abbreviations of proteins other than CPSF: PAP (Poly(A) polymerase), FIPS5 (factor interacting with poly(A) polymerase, PCFS4 (Pcf1-similar protein 4), SYM5 (Symplekin5), CLPS3 (CLP1-similar protein 3), FY (a homologue of yeast Pfs2p factor).
processing factor (FY) was reported to be mandatory for the FY function in flowering control.

**Conclusion:** Different plant proteins have been found to interact with 3'end processing of mRNA in tomato that may have important implications in gene regulation eventually leading to the control of many biological processes such as flowering time. This further demonstrates that the 3'-end formation of mRNA in the plant involves extensive protein-protein interaction network which is coupled with other processes of mRNA biogenesis such as transcription, capping, splicing and translation. The results would have practical implications in improving tomato cultivars by applying outcome of this research.

**Acknowledgment:** The research was conducted under Higher Education Commission, Islamabad, Pakistan funded Project No. 20-890. The help by Dr Ihsan Ullah from Agricultural Biotechnology Research Institute, AARI, Faisalabad is also deeply appreciated.

**REFERENCES**


