THE ESTABLISHMENT AND CHARACTERIZATION OF A NOVEL PLANT IN VITRO CLEAVAGE AND POLYADENYLATION ASSAY SYSTEM

by Jun Zheng

Cleavage and polyadenylation are two tightly coupled reactions required for messenger RNA processing in eukaryotic cells. An in vitro biochemical assay is critical for studying this process, but lack of such an assay system in plants hampered the study of plant mRNA 3’-end formation for the last two decades. This thesis reports the establishment and characterization a plant in vitro cleavage and polyadenylation assay system, in which soluble nuclear protein extracts from Arabidopsis cell cultures can accurately cleave different pre-mRNAs at expected poly(A) sites. The cleavage reaction was performed at optimal condition of 3.5mM Mg$^{2+}$ by an endoribonuclease activity, with a Km of 15.3nM. When supplemented with yeast poly(A) polymerase, about 150-nucleotide poly(A) tracts were added specifically to the newly cleaved 3’-ends. This long-sought system will serve as a platform for further studies of plant mRNA 3’-end formation mechanisms and other RNA processing events.
THE ESTABLISHMENT AND CHARACTERIZATION OF A NOVEL PLANT IN VITRO CLEAVAGE AND POLYADENYLATION ASSAY SYSTEM

A Thesis

Submitted to the
 Faculty of Miami University
 in partial fulfillment of
 the requirements for the degree of
 Master of Science

Cellular, Molecular and Structural Biology Graduate Program

by

Jun Zheng

Miami University
 Oxford, Ohio
 2010

Advisor _____________
 Dr. Qingshun Q. Li

Reader _____________
 Dr. Chun Liang

Reader _____________
 Dr. John Hawes
# TABLE OF CONTENTS

## CHAPTER 1
**Introduction** ................................................................................................................. 1  
Mesenger RNA 3’-end formation and polyadenylation ...................................................... 1  
Current understanding of eukaryotic pre-mRNA cleavage and polyadenylation machinery ................................................................. 2  
The importance of an *in vitro* assay systems in the study of plant polyadenylation mechanism ......................................................................................................................... 6  

## CHAPTER 2
**Materials and Methods** ............................................................................................... 9  
*Arabidopsis* suspension cell cultures ...................................................................... 9  
Nuclear protein extraction ....................................................................................... 9  
Labeling of RNA substrates for *in vitro* assays .................................................. 9  
*In vitro* cleavage and polyadenylation assay ........................................................ 10  
RNA substrate circulation and image quantification ............................................. 10  
Sequencing of cleavage and polyadenylation products ........................................ 11  

## CHAPTER 3
**Results** ......................................................................................................................... 12  
Validation of plant pre-mRNA cleavage activity from soluble nuclear proteins of *Arabidopsis* suspension cells ................................................................. 12  
The cleavage activity was carried out by an endoribonuclease ............................... 13  
The nuclear extract from WT cells culture had the cleavage activity ................. 13  
Validation of the cleavage activity using an *Arabidopsis* Rbcs gene ................. 14  
Optimization of cleavage reaction conditions ......................................................... 14  
The kinetics of cleavage reaction ........................................................................ 15  
Complementation of polyadenylation activity by yeast PAP ............................. 15  
The kinetics of polyadenylation reaction ............................................................... 16  

## CHAPTER 4
**Discussion and Conclusions** ....................................................................................... 17  
An *in vitro* cleavage and polyadenylation assay system for plant mRNA 3’-end formation was established in *Arabidopsis* ................................................................. 17  
The unique feature of this successful cleavage and polyadenylation assay system ................................................................................................................................. 17
The missing polyadenylation activity from plant nuclear protein extract and the complementation of yeast PAP .................................................. 18
Significance and future directions ............................................................... 19

References ........................................................................................................ 20
List of Tables

Table 1: PCR primers used in establishing a plant *in vitro* cleavage system 36
List of Figures

Figure 1: A mammalian pre-mRNA cleavage and polyadenylation model
Figure 2: Western blot verification of the purity of nuclear extracts
Figure 3: A schematic of the CaMV STS pre-mRNA substrate used in this study
Figure 4: Cleavage of 5’-end-labeled STS using Nu-CI
Figure 5: Cleavage of 3’-end-labeled STS by Nu-CI
Figure 6: Endonuclease activity assay of the nuclear extracts by cleavage of linear and circular STS pre-mRNA
Figure 7: Cleavage of uniformly-labeled STS pre-mRNA by nuclear protein extracts from wild-type cells (Nu-WT)
Figure 8: Cleavage of Arabidopsis endogenous Rubisco small subunit (rbcS) gene (At5g38420) 3’-UTR by Nu-CI.
Figure 9: Conditions of cleavage reaction
Figure 10: The cleavage reaction kinetics
Figure 11: Reconstitution of cleavage and polyadenylation coupling reaction by supplementation of yeast PAP
Figure 12: The polyadenylation reaction kinetics
I appreciate my advisor, Dr. Q. Quinn Li, for his mentoring and supporting to my study and research. I am also grateful to my committee members Dr. Chun Liang and Dr. John Hawes, for their valuable times, encouragements and comments to my work. Especially thanks to Dr. Hongwei Zhao, who had used a lot of time to initiate this research for the last few years, and he also showed and instructed me how to do the assays step by step. Thanks to Dr. Denghui Xing and Dr. Man Liu, for their helpful discussions, and thanks to Zhaoyang Liu for pleasant collaboration in both lab and classes. I would also like to thank Dr. Arthur Hunt for great suggestions, thanks to the Center of Bioinformatics and Functional Genomics for technical supports. Finally, I would like to thank my wife Guixia Huang, whose unconditional love and support makes this work possible.
CHAPTER 1

Introduction

Messenger RNA 3’-end formation and polyadenylation

Gene expression in eukaryotes requires the transcription of DNA into messenger RNA (mRNA) in the nucleus. The newly transcribed mRNA precursors (pre-mRNAs) undergo extensive processing, such as 5’-end capping, splicing and 3’-end polyadenylation, before they are ready to be transported to the cytoplasm for translation into proteins. At the 5’-end, a cap of 7-methylguanylate is added to the first base of pre-mRNA; in the middle, introns are spliced and exons are ligated together; cleavage and polyadenylation happens at the 3’-end, adding a poly(A) tail to the pre-mRNA.

Among these three pre-mRNA processing events, 3’-end formation and polyadenylation is attracting more and more attentions as it has been found to regulate transcription initiation, elongation and termination; affect intron splicing; promote mature mRNA transportation and translation initiation; and protect mature mRNAs from unregulated rapid degradation (Buratowski, 2005; Moore and Proudfoot, 2009). In addition, studies on alternative polyadenylation show that more than half of human genes can be alternatively processed at different 3’ sites, resulting in distinct mature mRNAs from the same pre-mRNA transcripts (Tian et al., 2005). More recently, the choice of alternative polyadenylation sites at 3’-untranslated regions (UTR) of many genes has been linked to gene expression level control and cancer development in human cells (Mayr and Bartel, 2009). Polyadenylation at 3’-end was also found to regulate synapse and axon development in C. elegans (Van Epps et al., 2010). Thus, a theme of gene expression regulation through pre-mRNA polyadenylation is emerging. Owing to its tight connections to transcription, translation, and RNA decay, mRNA 3’-end polyadenylation appears to act as a hub to fine-tune gene expression and thus to form a complicated regulation network (Danckwardt et al., 2008).
Current understanding of eukaryotic pre-mRNA cleavage and polyadenylation

The 3’-end formation of mRNA includes two coupled steps: the cleavage at a specific site at the 3’-untranslated region (UTR) by a nuclease activity and the addition of a poly(A) tract to the newly formed 3’-end by poly(A) polymerase.

The biochemical process of polyadenylation is relatively well studied in mammals and in yeast. Using bioinformatics, genetics, biochemical and structural methods, it has been showed that the 3’-end cleavage and polyadenylation reaction is guided by the sequence elements (cis-elements) in the pre-mRNA, and carried out by the enzymatic activities in a large protein complex, which recognize those sequence elements, cleave and polyadenylate at the specific site at the 3’-end. In this apparently simple process, in mammals several cis-elements as well as more than 14 protein factors are involved (Mandel et al., 2008). These 14 protein factors can be divided into several sub-complexes, such as cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I (CF Im), and cleavage factor II (CF IIm). Meanwhile, poly(A) polymerase (PAP), poly(A) binding protein-II (PABP-II), symplekin, and C-terminal domain of the RNA polymerase II (Pol II CTD) have also been found to be necessary in this 3’-end processing complex (Mandel et al., 2008).

In mammals, the pre-mRNAs contain three primary sequence elements. The first element is the highly conserved hexamer AAUAAA polyadenylation signal (PAS); the second element is the cleavage site itself which is preferred after CA dinucleotides; the third element is the G/U-rich downstream element (DSE). At the same time there are two auxiliary sequence elements located both upstream and downstream (Zhao et al., 1999). The polyadenylation signal (PAS) is highly conserved in mammals, the AAUAAA hexamer was found in about 60% of sequenced mRNAs, and the AUUAAA variation in the second position, was found in about 15% of all sequenced mRNAs (Beaudoing et al., 2000; Tian et al., 2005). The guidance effect of PAS has been confirmed by point mutations experiments in the PAS which produced a significantly decreased level of polyadenylated mRNA and a significantly increased level of unprocessed pre-mRNA (Wickens and Stephenson, 1984). More
interestingly, the distance between the PAS and the cleavage site is conserved too, from 10bp to 30bp (Fitzgerald and Shenk, 1981). The cleavage site (CS) itself is another sequence element in mammals, it lays between the polyadenylation signal (PAS) and the downstream element (DSE). In vertebrate pre-mRNAs, about 70% of all the cleavage sites are located after an adenosine residue, with a preference of A > U > C >> G. Meanwhile, before the adenosine, a cytosine is preferred (Sheets et al., 1990). Thus, the cleavage site is a less conserved element which directs the cleavage and polyadenylation site after the preferable CA dinucleotides. The downstream element is located at about 10-30 nucleotides after the CS, with characters of GU-rich or simply U-rich sequences. The two auxiliary elements, locating both upstream and downstream, are even less conserved. The auxiliary upstream element often contains a U-rich sequence, while the auxiliary downstream element often contains a G-rich sequence. They are found to enhance the efficiency of cleavage and polyadenylation at the pre-mRNA 3’ end, maybe by the facilitating the binding of other polyadenylation factors to the cleavage site (Brackenridge and Proudfoot, 2000; Gilmartin, 2005).

Under the guidance of sequence elements (cis-elements), the cleavage and polyadenylation machinery recruit a large protein complex which contains the enzymes that directly cleave and polyadenylate the 3’ end at the cleavage site. There are more than 14 proteins in the complex in mammals, and more than 20 proteins in yeast. This complicated 3’-end processing complex, together with their activities and interactions are one of the most studied areas in the study of mRNA 3’ end formation (Mandel et al., 2008; Zhao et al., 1999).

Two enzymatic activities are needed for the process, one is the cleavage activity, and the other is the polyadenylation activity. The 73kd subunit of the cleavage and polyadenylation specificity factor (CPSF-73) was eventually found to most possibly be the endonuclease that is responsible for the cleavage activity (Mandel et al., 2006), while the poly(A) polymerase (PAP) was found relatively long ago to be the enzyme which adds the poly(A) tract at the newly cleaved 3’ end (Zhao et al., 1999). Together with the help of other protein factors, the 3’-end processing complex can cleave and polyadenylate the pre-mRNA, forming the mature
In mammals, cleavage and polyadenylation specificity factor (CPSF) has five subunits (according to their molecular weights in Kd), they are CPSF-30, CPSF-73, CPSF-100, CPSF-160 and hFip1. All of them are needed in efficient cleavage and polyadenylation reactions (Mandel et al., 2008). Mammalian CPSF-160 is 160Kd and the largest subunit of CPSF which directly binds to the PAS, with a relative high affinity for the perfect PAS sequence (AAUAAA) (Zhao et al., 1999). CPSF-160 also interacts with other protein factors such as CPSF-100, hFip1, and PAP (Murthy and Manley, 1995).

The most important subunit of CPSF is CPSF-73, which is found to be the endonuclease that is directly responsible for the endoribonuclease activity in the cleavage step of pre-mRNA 3’-end formation. First of all, UV cross-linking of nuclear extracts revealed that CPSF-73 binds directly to the cleavage site in an AAUAAA-dependent manner (Ryan et al., 2004). Secondly, by biochemical studies, the expressing and purification of human CPSF-73 from bacteria showed ribonuclease activity on pre-mRNA substrates (Mandel et al., 2006). Meanwhile, the 3-D crystal structure of CPSF-73 was solved which showed that its N-terminal region contains a metallo-\(\beta\)-lactamase domain (Mandel et al., 2006). The metallo-\(\beta\)-lactamase domain and the so-called \(\beta\)-CASP superfamily are features of nuclease (Dominski et al., 2005a). In all, the CPSF-73 has the endoribonuclease activity which interacts with the pre-mRNA at the cleavage site, indicating its role in pre-mRNA 3’ end cleavage.

CPSF-100 is a close relative to CPSF-73, with a recognizable sequence homology (23% sequence identity and 49% similarity for their metallo-\(\beta\)-lactamase), but the structure did not contain the zinc-binding residues, and therefore cannot bind zinc ions and catalytically inactive (Callebaut et al., 2002). Yeast two-hybrid experiments indicated that, while the N-terminal of CPSF-73 is responsible for the endonuclease activity, the C-terminal of CPSF-73 interact with CPSF-100 (Dominski et al., 2005b), and through which to the whole protein complex.
CPSF-30 contains five CCCH zinc finger motifs followed by a CCHC zinc knuckle (Barabino et al., 1997). Because both of the CCCH and CCHC motifs are found to be RNA-binding, CPSF-30 was confirmed to bind to RNA with a preference of PolyU (Barabino et al., 1997). The recombinant CPSF-30 from *Arabidopsis* was showed to have an endonuclease activity that might also be involved in the cleavage reaction (Addepalli and Hunt, 2007).

hFip1 is about 66kd, and contains two distinct segments near the N terminal, the first one is the acidic segment that mediates interactions with PAP, the other one is the highly conserved segment of about 70 residues which functions in interaction with CPSF-30. hFip1 also functions to interact with CPSF-160, CstF-77 and CF Im, and through which bring PAP close to the polyadenylation site (Venkataramanan et al., 2005).

Cleavage stimulation factor (CstF) contains three subunits CstF64, CstF77, and CstF50. CstF-64 has a conserved RNA-recognition motif (RRM) at its N-terminal, which is sufficient for RNA-binding showing preference for the G/U-rich downstream element. There is a hinge region of about 100 residues next to the RRM, which functions in mediating protein-protein interactions and binds to CstF-77 and symplekin (Hatton et al., 2000). CstF-64 contains a C-terminal domain, covering the last 50 amino acids, that is more conserved than the RRM but the functions is still unknown (Mandel et al., 2008). CstF-77 contains 12 repeated HAT (half a TPR) motifs at the N-terminal. Similar to the tetratricopeptide repeat (TPR) motifs which often mediates protein-protein interactions, the HAT-containing CstF-77 is found to interact with CPSF-160 and CTD of Pol II by yeast two-hybrid assays (Bai et al., 2007). CstF-50 contains seven WD-40 repeats that mediates its interaction with CstF-77 (Zhao et al., 1999).

Cleavage factors Im (CF Im) contains 25-, 59- and 68-kDa subunits, and functions to recognize and define of the proper polyadenylation site (Brown and Gilmartin, 2003). CF IIm contains hPcf11 and hClp1 (de Vries et al., 2000), and functions to interact with the CTD of Pol II in regulating transcription termination (Hollingworth et al., 2006).
Poly(A) polymerase (PAP) and Poly(A)-binding protein (PABP) are also important in the polyadenylation process. PAP is an induced-fit enzyme; its N-terminal domain can coordinates either Mg$^{2+}$ or Mn$^{2+}$. Under physiological conditions and in the absence of other factors, PAP has only a very low level of unspecific activity (Zhao et al., 1999). The rate of poly(A) addition is significantly elevated when manganese is substituted for magnesium in the reaction. The C-terminal domain of PAP binds hFip1 and CPSF-160 (Mandel et al., 2008). PABP facilitates the quick and correct addition of poly(A) tail. The polyadenylation firstly occurs without PABP, and then PABP binds to the nascent poly(A) when it is 11–14bp in length and promotes its elongation and controls its length (Meyer et al., 2002). After binding of PABP, the polyadenylation reaction becomes highly processive.

In summary, in the mammalian system, CPSF-160 recognize the upstream PAS and CstF-64 recognize the G/U-rich DSE, they help to bring the whole 3’-end processing machinery to the right position on the pre-mRNA. The endoribonuclease (CPSF-73) is placed close to the cleavage site. Meanwhile the PAP is recruited to the complex by CPSF-160 and hFip1. The endoribonuclease CPSF-73 cleaves at the cleavage site, and PAP add about 200-300 adenyl at the newly formed 3’-end with the help of PABP (see Figure1; (Mandel et al., 2008). Interactions between polyadenylation factors and other RNA processing factors couple 3’-end polyadenylation to transcription, 5’-end capping and splicing.

**The importance of an *in vitro* assay system in the study of plant polyadenylation mechanism**

When investigating the mechanisms of mRNA 3’-end formation in mammals and yeast, *in vitro* assay systems were critical in identifying and characterizing components of the cleavage and polyadenylation complex (Shi et al., 2009; Zhao et al., 1999). These *in vitro* assay systems simplified the isolation and identification of new protein factors involved in this process, making it much easier to perform subsequent functional studies of each protein factors and test the functions for *cis*-elements. With the *in vitro* assay system, tentative cleavage and polyadenylation factors can be purified and concentrated by following the active
fractions in chromatography purification. The identities and functions of interesting protein factors and cis-elements were revealed too (Yoshio and Manley, 1997).

In mammals, the first successful in vitro polyadenylation system which coupled with RNA transcription was reported in 1978. Cooper and Marzluff showed that under suitable conditions, soluble nuclear protein extract from mouse cells can transcribe double strand DNA to single strand pre-mRNA, and the subsequent cleavage and polyadenylation at the 3’ end can be detected too in vitro (Cooper and Marzluff, 1978). After that, in vitro RNA assay system in which polyadenylation was uncoupled from transcription was developed by using exogenous pre-RNAs, with this system, synthesized pre-mRNAs were accurately cleaved and polyadenylated at the expected poly(A) sites by nuclear protein extracts (Manley, 1983; Moore and Sharp, 1985). After that, cleavage and polyadenylation machinery in mammals has been extensively studied. It was found that, for the in vitro cleavage and polyadenylation assay in mammals, all 13 protein factors except PABP are required in the cleavage step, while only CPSF, PAP and PABP are required in the polyadenylation step (Zhao et al., 1999). At the same time, the functions of each cis-elements and protein factors were studied as discussed in the previous sections.

Until now, however, an in vitro assay system that can accurately cleave and polyadenylate pre-mRNA in plants has never been established, and the lack of such a system has affected the study of polyadenylation machinery in plants (Li and Hunt, 1997; Rothnie, 1996). To address this, numerous efforts were made in establishing such a system by using the nuclear extracts from different plants, e.g., peas, cauliflower, and Arabidopsis, but without success. Over the past few years, through bioinformatics, phylogenetics, protein-protein interactions, and proteomic analysis, the conserved polyadenylation factors in plants have been identified (Herr et al., 2006; Hunt et al., 2008; Simpson et al., 2003; Zhao et al., 2009). However, the biochemical efficacy of these proteins largely remains elusive.

Recently, while studying the proteomics of plant polyadenylation factors, a number of Arabidopsis CPSF subunits were overexpressed using a Tandem Affinity Purification (TAP)
tagging approach (Zhao et al., 2009). Among the TAP-fused proteins was an *Arabidopsis* homologue of mammalian CPSF73, termed AtCPSF73-I. The nuclear extracts from AtCPSF73-I overexpression cell culture showed a cleavage activity for pre-mRNA 3’-end (Zhao, 2008).

In this thesis, I will demonstrate the experiments I did to validate the pre-mRNA cleavage activity, to optimize the assay conditions, and to reconstitute the full cleavage and polyadenylation activity by supplementing with yeast poly(A) polymerase to the plant *in vitro* assay system. Finally a fully reconstituted mRNA cleavage and polyadenylation assay system in plants is established. Wild-type nuclear protein extracts from *Arabidopsis* suspension cells have also been shown to carry out the committing cleavage step, and the polyadenylation activity was reconstituted by a yeast PAP too. This assay system will be useful for studying plant mRNA 3’-end formation mechanisms as well as other related RNA processing events.
CHAPTER 2

Materials and Methods

*Arabidopsis* suspension cell cultures

*Arabidopsis* thaliana suspension cell culture (MM1 from Landsberg erecta) conditions and AtCPSF73-I -TAP fusion constructs and transformation were as described (Zhao et al., 2009).

Nuclear protein extract preparation

Nucleus isolation from *Arabidopsis* suspension cultures was modified from Escobar et al. (Escobar et al., 2001) and Folta et al. (Folta and Kaufman, 2006). About 50g suspension cells were ground in liquid nitrogen with acid-washed sand (Sigma) and then resuspended in extraction buffer, filtered through 2 layers of Miracloth twice. The filtrate was centrifuged at 25,000xg for 5 min. The pellet was collected, resuspended in 30% Percoll (GE Healthcare), then overlaid on the top of 30% and 80% Percoll double layers, centrifuged again at 2,000xg for 30 min using a swing rotor. The middle layer between the 30% and 80% Percoll layers was collected, washed twice with Buffer B (25mM Tris-HCl pH8.0; 10mM MgCl$_2$; 0.46M sucrose; 0.5mM PMSF; 6mM β-ME; 0.5% Triton X-100), and the nuclei were resuspended in 5 ml Buffer C (25mM Tris-HCl pH8.0; 10mM MgCl$_2$; 0.46M sucrose; 0.5mM PMSF; 6mM β-ME; 75% Percoll). The nuclei were then collected by centrifuge at 5,000xg for 30 min.

The concentrated nuclei were broken by slowly adding 2.0M ammonium sulfate solution to a final concentration of 0.5M. Lysed nuclei were centrifuged at 13,000rpm for 30 min, and the supernatant (soluble nuclear protein extracts) were recovered and stored at -80°C freezer for future use.

Labeling of RNA substrates for *in vitro* assays

The 3’-UTR of CaMV 35S RNA STS clone (Mogen et al., 1990) was a gift from Dr. Arthur Hunt (University of Kentucky). The target region (Fig. 3) was amplified with a
primer fused with a T7 promoter sequence at the 5'-end. The PCR product was purified and used as a template for in vitro transcription using the AmpliScrib T7 High Yield Transcription Kit (Epicentre, Inc.), according to the manufacturer’s instructions. The same kit was used for the transcriptions of cold and [α-32P]ATP-labeled (using 1/10th of cold ATP) RNA. The cold STS RNA was 5'-end labeled by [γ-32P]ATP using RNA kinase (Epicentre, Inc.), and 3'-end-labeled by [5'-32P] pCp using T4 RNA ligase (New England Biolabs). The labeled RNAs were purified using 7M urea 6% polyacrylamide gel. The corresponding gel bands were auto radiographed and cut out, eluted using RNA elution buffer (50mM Tris-HCl pH7.5; 10mM EDTA; 0.1% SDS; 0.3M NaCl) over nigh under room temperature with shaking, precipitated using cold ethanol, washed, dissolved to proper concentration and stored for further use. Glycogen can be used to aid precipitation, if desired.

**In vitro cleavage and polyadenylation assay**

For an in vitro assay, in a 0.6 ml Eppendorf tube, 4.5μl of cleavage buffer (5mM MgCl2; 41.67mM phosphocreatine disodium; 1.67mM ATP; 3.3% glycerol; 0.8% polyvinyl alcohol; 3.3mM HEPES pH 8.0), 0.5μl RNaseOut (an RNase inhibitor; Invitrogen), and 0.5μl nuclear protein extracts (about 0.5μg/μl, either dialyze or not), diluted pre-mRNA substrate (1000~2000 cpm/reaction), and water were added to a total final volume of 7.5μl. Incubate the reaction at 30°C for 2 hours. For polyadenylation assays, add 1μl diluted (0.01X, 6 units) yeast poly(A) polymerase (US Biochemical). When the reaction was done, directly add 7.5μl 2X RNA gel loading buffer, load all 15 μl to a 6% sequencing gel, and run at 800-1000V for 1.5 to 2 hours. Transfer the gel to Whatman filter paper, dry, autoradiograph and scan by a PhosphorImager Scanner (Molecular Dynamics).

**RNA substrate circulation and image quantification**

Hot linear STS RNAs were ligated at 16° C overnight with T4 RNA ligase (New England Biolabs), and these partially ligated RNAs were used in the assay as combination of linear and circular RNAs. After cleavage, both linear and circular cleavage products were imaged and
then quantified using Imagequant (Molecular Dynamics) software, by selecting equal rectangle area which sufficiently cover the bands among different lanes.

**Sequencing of the cleavage and polyadenylation products**

The polyadenylation products of cold STS RNA were reverse transcribed with oligo(dT)$_{18}$VN primer with an adaptor sequence. The reverse transcribed cDNA was then amplified by using an STS-specific primer and a primer annealed to the adaptor sequence. The amplified fragments were cloned into pTopo TA cloning vector (Invitrogen), and transformed to *E. coli* Dh5α competent cells. The positive clones were selected, and their plasmids were extracted and sequenced using M13 primer located in the cloning vector to reveal the cleavage and polyadenylation site.
Chapter 3

Results

Validation of the plant pre-mRNA cleavage activity from soluble nuclear proteins of Arabidopsis suspension cells

While studying the proteomics of plant polyadenylation factors, a number of Arabidopsis CPSF subunits were overexpressed and isolated by using a Tandem Affinity Purification (TAP) tagging approach (Zhao et al., 2009). Among the TAP-fused proteins was an Arabidopsis homologue of mammalian CPSF73, termed AtCPSF73-I. As the mammal CPSF73 was shown to be the endonuclease that plays a direct role in the cleavage step of pre-mRNA 3’-end formation in mammals (Dominski et al., 2005a; Mandel et al., 2006), the function of AtCPSF73-I in plant polyadenylation was also examined. It is found that the nuclear extract from Arabidopsis suspension cells overexpressing AtCPSF73-I had a pre-mRNA cleavage activity (Zhao, 2008).

To validate the plant pre-mRNA cleavage activity, soluble nuclear protein extract were prepared from nuclei of AtCPSF73-I overexpressed suspend cells, free from cytoplasm or organelle contamination (Fig. 2). This extract (called Nu-CI) was used for the initial establishment of the assay conditions.

A 3’-UTR from the pre-mRNA of cauliflower mosaic virus (CaMV) transcript named STS was used as the substrate. The cleavage and polyadenylation profile of STS has been extensively studied (Mogen et al., 1990; Rothnie et al., 1994). Based on conventional genetic studies, it has been determined that STS contains three well-defined cis-elements: the far-upstream element (FUE), the near-upstream element (NUE), and the polyadenylation sites (Fig. 3). Accordingly, the nuclear extract was tested for its cleavage activity of the uniformly [α-32P]ATP-labeled STS in assay conditions as described in the Methods. Indeed, the nuclear proteins could cleave STS into about 182 nt, the expected size range as
authenticated in vivo assays (Rothnie et al., 1994). To rule out random nuclease activity cleavage as a cause of cleavage activity, a 5’-end-labeled STS was generated and subjected to the same assay conditions. Again, the cleavage reaction left a product that was ~182 nt in length (Fig.4). In agreement with the in vivo experiments, the results showed that the cleavage occurred at the same sites (Rothnie et al., 1994), producing two different lengths, 180 and 182 nt (not seen as two bands on the gel due to resolution).

The cleavage activity was carried out by an endoribonuclease

To rule out that the cleavage activity was from an exonuclease activity, cold STS was 3’-end-labeled by [5’-32P] pCp, using T4 RNA ligase, and then subjected to the same assay. As shown in Fig. 5, a 3’-end cleavage product of about 20 nt was detected. This result excludes the possibility that such products were generated by an exonuclease activity (which could only have generated free nucleotides, not a 20-nt fragment).

To further prove that the cleavage activity was carried out by an endoribonuclease, uniformly labeled circular and linear STS were both subjected to the assay. As indicated in Fig.6, when both circular and linear forms were added to the same assay reaction, they both were cleaved at a similar rate during a time course study, suggesting the involvement of an endoribonuclease. Thus, an in vitro assay system for pre-mRNA cleavage was established by using the nuclear proteins from cells overexpressing AtCPSF73-I.

The nuclear protein extract from WT cells culture had the cleavage activity

Since we found the cleavage activity based on nuclear extracts, we further asked if the overexpression of AtCPSF73-I was required for this activity. To address this question, we generated nuclear extracts from non-transformed, wild-type culture cells (WT), and we tested for their cleavage activity. Surprisingly, the WT nuclear protein was also able to correctly cleave the pre-mRNA too (Fig. 7). Therefore, our results indicated that specific pre-mRNA cleavage activity could be found in both the AtCPSF73-I overexpressing and the WT nuclear protein extracts. Although the overexpressed materials helped us in the establishment of the
assay system at first, overexpression of the plant polyadenylation factors was not essential for this activity. So most of the following experiments were done using nuclear protein extracts from AtCPSF73-I overexpressing cells, but were repeated using WT nuclear extracts as indicated.

Validation of the cleavage activity using an *Arabidopsis Rbcs* gene

Furthermore, we tested if the cleavage activity also works on *Arabidopsis* endogenous genes, in contrast to one from viral genome (CaMV), a representative 3’-UTR of a Rubisco small subunit gene (At5g38420; encoded by a nuclear gene) was chosen. When this pre-mRNA was processed *in vitro* using our cleavage assay, two bands (222nt and 167nt; Fig. 8) were detected corresponding to the polyadenylation sites supported by ESTs found in The *Arabidopsis* Information Resources (TAIR) databases. Considering the fact that the assay system also depended on the correct poly(A) signals (Zhao, 2008), the cleavage activity we established is an authentic mRNA 3’-end processing component that functions on both viral and native plant pre-mRNA transcripts in a poly(A) signal-dependent manner.

Optimization of cleavage reaction conditions

To optimize the conditions for the cleavage assay, different temperatures and reaction times were tested. The highest cleavage efficiency was observed at 30°C, where about 80% of pre-mRNA was cleaved (Fig. 9A). When a time course of the cleavage reaction was carried out, we found that a significant proportion of cleavage products was detected within 90min and continued to accumulate until 180min (Fig. 9B). Thus, the overall reaction condition can be set at 30°C, for 120min.

It has been shown that divalent metal ions, such as magnesium (Mg$^{2+}$), affect cleavage and polyadenylation in both mammalian and yeast systems (Moore and Sharp, 1985; Ryan et al., 2004; Wahle, 1991). To test the efficacy of these ions in our cleavage system, uniformly labeled STS was incubated with nuclear extracts from AtCPSF73-I overexpressed cells under the optimal reaction conditions described above, except varying concentrations of Mg$^{2+}$. The
cleavage efficiencies were then compared. As shown in Fig. 9C, cleavage efficiency was increased with increasing Mg\(^{2+}\) concentration, until reaching the highest level at 3.5mM. Thus, 3.5mM was used as the standard condition. These results largely agree with those found in the *in vitro* assays of mammalian and yeast systems (Butler et al., 1990; Manley, 1983; Moore and Sharp, 1985), where Mg\(^{2+}\) is important for poly(A) signals specific activities.

### The kinetics of the cleavage reaction

To study the reaction kinetics, a series dilution of STS substrates was assayed using excessive amount of the WT nuclear proteins under the standard cleavage assay conditions. The reaction was stopped after 30 min, and the products were then resolved in a sequencing gel, and the ratios of cleaved products to uncleaved products were measured. Based on the amounts of RNA added, the concentrations of RNA substrate as well as the cleavage velocity were derived. The plot shown in Figure 10 was drawn using substrate concentration vs. average cleavage velocity. This exercise generated a Km for STS pre-mRNA (202 nt) of about 1ng/\(\mu\)l or 15.3 nM.

### Complementation of polyadenylation activity by yeast PAP

Having successfully reconstituted cleavage activity, we extended our search for the polyadenylation activity. However, after many attempts, no detectable poly(A) tail was found in coupling with the cleavage assay. As in mammal and yeast mRNA 3’-end formation, the cleavage step is fully committed and irreversible, and, when compared to polyadenylation, it requires more protein factors (Zhao et al., 1999). On the other hand, the polyadenylation reaction requires PAP and the Poly(A) Binding Protein (PABP-II). Since these were extracts from whole nucleus where polyadenylation occurs, it is likely that these proteins were there, but their activities were either masked or inhibited by unknown factors. To circumvent this and reconstitute the full 3’-end processing assay, we supplemented the cleavage assay by commercially available yeast PAP. To our surprise, the addition of yeast PAP induced the specific addition of about a 150 nt poly(A) tract (Fig. 11, lane 2). To confirm that the
poly(A) addition was cleavage-dependent, the polyadenylation products were cloned and sequenced. The results showed that, indeed, the poly(A) tails were added to the newly cleaved ends. With the addition of cordycepin (3’-dATP), an RNA chain elongation inhibitor, the addition of poly(A) to the newly cleaved product was interrupted, as shown in Figure 11, lane 5, as compared to the normal addition of poly(A) tail by PAP shown in Figure 11, lane 4. Importantly, PAP alone could not add poly(A) to the pre-mRNA without the help of nuclear proteins (Fig. 11, lane 6) under our assay conditions, which indicates that the addition of poly(A) is specific to the RNA that were just generated by the cleavage activity, a result of the cooperation between plant nuclear extracts and the yeast PAP. Such a cross-kingdom cooperative reconstitution of polyadenylation factors between yeast and plants opens up the possibility of complementation of other poly(A) factors, some of which were proved not possible between mammals and yeast (Jenny et al., 1994; Jenny et al., 1996). More importantly, such complementation also implies that the cleavage activity described here is authentic since it can be recognized and cooperated by yeast PAP because only the correct protein-protein interaction may lead to such cooperation, as it is indicated by a recent work in yeast (Meinke et al., 2008).

The kinetics of the polyadenylation reaction

After complementation of polyadenylation activity, the reaction kinetics was also studied. A time course of cleavage and polyadenylation reactions using Nu-CI with supplement of yPAP (Fig. 12), showed that the addition of poly(A) by yPAP was progressive at 5-10nt/10min, but faster at the beginning at a rate of 60nt/10min.
**Discussion and Concluding Remarks**

An *in vitro* cleavage and polyadenylation assay system for plant pre-mRNA 3’-end formation was established in *Arabidopsis*

We report here the successful establishment of an *in vitro* cleavage and polyadenylation system in *Arabidopsis*. This assay system will not only promote the biochemical study of plant polyadenylation, but will also shed light on the establishment of other mRNA processing *in vitro* reactions, e.g., splicing and transcription termination. Although, in many cases, genetic studies can explain how genes function, biochemical investigation most certainly will lead to important functional information at the molecular level. With an increasing number of cases documenting gene expression regulation through alternative processing, such as alternative splicing and alternative polyadenylation, of pre-mRNA, *in vitro* studies to understand the mechanisms at the biochemical level would be indispensable.

**The unique feature of this successful cleavage and polyadenylation assay system**

Compared to previous attempts to set up such an *in vitro* assay system, our success may be attributed to some distinct features such as the protein source, the way to prepare nuclear extracts, reaction conditions, and choice of pre-mRNA substrate. First, the nuclear proteins used in our research came from *Arabidopsis* suspension cells, which are actively dividing and growing cells. This compares to materials previously used, including young seedlings, or cauliflower heads, both of which only have a small percentage of actively dividing cells. The gene expression machinery in actively growing and dividing cells is continually active, and their nuclear proteins may therefore be more enriched with polyadenylation factors. Second, the method we used in our nuclear protein extraction may result in more abundant polyadenylation factors. Because our extraction began with a considerable amount of material
(about 50g cells), we can obtain about 0.5-1 ng/μl nuclear proteins, which is a relatively high concentration for *in vitro* assays, even after two centrifugations through Percoll gradients (to ensure purity). Meanwhile, we minimized the volume of nuclear extracts by directly adding ammonia sulfate to a final concentration of 0.5M to break the nuclei. So doing we also eliminated the need of using high concentration of ammonia sulfate to precipitate nuclear proteins, thus possibly helping to preserve functional protein complexes that are required for *in vitro* activity. Third, the assay conditions we used were carefully optimized to achieve maximal activity. In particular, the use of polyvinyl alcohol and glycerol in the cleavage buffer may further increase the effective reaction concentration by reducing reaction volume. Finally, the pre-mRNA substrate we used, CaMV STS pre-mRNA, is well characterized (Mogen et al., 1990; Rothnie et al., 1994). This 200nt substrate contains both strong AAUAAA as NUE, as well as the typical FUE, with an extension of a newly discovered Cleavage Element (Loke et al., 2005) covering both sides of the cleavage site. These *cis*-elements may help cleavage site recognition and processing. Taken together, incremental modifications of all these may contribute to a successful reconstitution of the cleavage assay.

**The missing polyadenylation activity from plant nuclear protein extract and the complementation of yeast PAP**

Our *in vitro* assay system is quite similar to assays used for mammals and yeast, including assay conditions, small volume reaction of about 10μl, reaction over a period of 2h at 30ºC, and similar concentration of Mg$^{2+}$, K$^+$ and Na$^+$ ions. However, we did not include any EDTA, and used much less radiolabelled RNA (only about 2000 cpm/reaction). Still, while our assay conditions yielded nuclear extracts exhibiting cleavage activity, no polyadenylation activity could be observed. As a consequence, we asked why PAP activities, which have been seen before in nonspecific assays (Addepalli et al., 2004; Hunt et al., 2000), were not found in our extracts. One plausible explanation is that while nuclear protein extracts contain all protein factors involved in the polyadenylation steps, PAP activity was masked by inhibitory
factors, such as Putative Polyadenylation Factor-B, which has been shown to inhibit a
nonspecific PAP activity (Forbes, 2004). Alternatively, the phosphorylation state of PAP
could also be altered thus affecting the activity (Bond et al., 2000). Meanwhile, other not yet
identified PAP inhibitors may exist to prevent in vivo polyadenylation from happening. It is
interesting to note that, initially, PAP activity was very low when setting up the yeast in vitro
polyadenylation assay system. However, researchers eventually was able to achieve higher
PAP activity (Butler and Platt, 1988; Butler et al., 1990).

Significance and future direction

With the supplement of yeast PAP, we managed to gain the full processing activity in our
assay system. Functional complementation of plant proteins with yeast protein, or vice versa,
is not uncommon (e.g., among exosome proteins (Chekanova et al., 2000)). However, early
work with polyadenylation factors indicated that such complementation with CPSF 73 and
100 and their counterparts in yeast (Jenny et al., 1994; Jenny et al., 1996) was not successful.
Our results provided direct evidence that at least some yeast polyadenylation-related proteins
can work cooperatively in the plant system. It would be interesting to see if plant
polyadenylation factors can complement the functions of their yeast counterparts.

Clearly, the establishment of such an in vitro cleavage and polyadenylation assay system opens up a
new avenue to study plant polyadenylation. Future directions should include the identification of the plant
polyadenylation factors that are responsible for the activity, and the roles of those potential factors revealed
through bioinformatics and genetic studies, as confirmed by protein-protein interactions (Hunt et al., 2008;
Xing et al., 2008a; Xing et al., 2008b; Xu et al., 2006; Zhao et al., 2009). Several approaches can be
employed to identify the plant polyadenylation factors through this type of biochemical assay. Moreover,
with the broad spectrum of tools we have generated, such as genetic mutants, gene expression constructs,
and protein-protein interaction profiles, the assay system will undoubtedly promote the synergistic study of
plant mRNA alternative polyadenylation, and post-transcriptional gene expression regulation in general.
References


homologue of the 64 kDa subunit of cleavage stimulation factor interacts with the 77 kDa subunit encoded by the suppressor of forked gene. Nucleic Acids Research 28, 520-526.


Table 1. PCR primers used in establishing a plant *in vitro* cleavage system

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>748</td>
<td>5’-ATTTAGGTGACACTATAGAACA CGCTGAAATCAC-3’</td>
<td>Forward primer, for the cloning of STS DNA used in <em>in vitro</em> transcription,</td>
</tr>
<tr>
<td>711</td>
<td>5’-GTACTGGATTTTGGTTTAG-3’</td>
<td>Reverse primer for the cloning of STS DNA used in <em>in vitro</em> transcription</td>
</tr>
<tr>
<td>803</td>
<td>5’-TAAACGACTCACTATAGGAG TAATCCCCTTTCTGGAATATTC-3’</td>
<td>Forward primer, for the cloning of RbcS gene At5g38420</td>
</tr>
<tr>
<td>804</td>
<td>5’-CAATGTGTGTTTTAAAGGAGCTAA AC-3’</td>
<td>Reverse primer, for the cloning RbcS gene At5g38420</td>
</tr>
<tr>
<td>725</td>
<td>5’-TTCTAGAATTCCAGCATTGC TCTTTTTTTTTTTTTTTTV-3’</td>
<td>Oligo(dT)V+ adaptor, for the reversed transcription of polyadenylated RNA, with an adaptor sequence for PCR priming</td>
</tr>
<tr>
<td>726</td>
<td>5’-TTCTAGAATTCCAGCATTGC TC-3’</td>
<td>Adaptor, for the RT-PCR of polyadenylated RNA with 725</td>
</tr>
<tr>
<td>888</td>
<td>5’-TTGTAADAACGACGGCCAGT-3’</td>
<td>M13 forward primer, for the sequencing of insertions in pTOPO vectors</td>
</tr>
</tbody>
</table>
Figure 1: A mammalian pre-mRNA cleavage and polyadenylation model.

Schematic drawing of the pre-mRNA 3’-end processing complex in mammals, the cis-elements in the pre-mRNA are also indicated (Mandel et al., 2008).
Figure 2: Western blot verification of the purity of nuclear extracts.

*Left panel,* Coomassie blue staining of equal loading of whole cell extracts (WCE-CI), and the nuclear extracts (NU-CI). *Right panel,* The same loading of WEC-CI and NU-CI were detected by Rubisco large subunit (RbcL) antibody (purchased from Agrisera, UK) for any potential cytoplasmic contamination. The expected RbcL band is ~45KD, as seen in the whole cell extract (WCE-CI), but it is absent in the nuclear extract preparation (NU-CI). The protein size markers (KD) are marked on the right.
Figure 3: A schematic of the CaMV STS pre-mRNA substrate used in this study.

The CaMV STS pre-mRNA is the 3’-UTR (untranslated region) of CaMV 35S RNA, which includes the FUE (Far-Upstream Element), the NUE (Near-Upstream Element), and the CE (Cleavage Element, including cleavage site, CS, indicated by arrows), with the size of cleavage products indicated. CDS, coding sequence.
Figure 4: Cleavage of 5’-end-labeled STS using Nu-CI.

Cleavage of 5’-end-labeled STS using nuclear protein extract from cells overexpressing AtCPSF73-I (Nu-CI). The arrow indicates the expected 182nt 5’-end cleavage product.
Figure 5: Cleavage of 3’-end-labeled STS by Nu-CI.

Cleavage of 3’-end-labeled STS by Nu-CI. The arrow indicates the 20nt 3’-end product.
Figure 6: Endonuclease activity assay of the nuclear extracts by cleavage of linear and circular STS pre-mRNA

(A) A time course of the cleavage of both linear and circular STS pre-mRNAs. Circular RNA (**) was generated by ligating the linear one (*) using an RNA ligase. RNA size standards (nt) are shown on the left. (B) Plot of the percentages of each uncleaved pre-mRNAs found in (A). The linear RNA at time 0 was set at 100%.
Figure 7: Cleavage of uniformly-labeled STS pre-mRNA by nuclear protein extracts from wild-type cells (Nu-WT)

Cleavage of uniformly-labeled STS by nuclear protein extracts from wild-type cells (Nu-WT). The arrow indicates the 182nt 5'-end product. Control in each panel refers to reactions using buffer to replace nuclear extracts. RNA size standards (nt) are marked on the right hand side of each panel.
Figure 8: Cleavage of *Arabidopsis* endogenous Rubisco small subunit (*rbcS*) gene (At5g38420) 3’-UTR by Nu-Cl.

The cleavage of *Arabidopsis* endogenous Rubisco small subunit (*rbcS*) gene (At5g38420) 3’-UTR by Nu-Cl, showing it was cleaved at the correct sites indicated by triangles (sizes labeled by *). The same results were obtained with Nu-WT. RNA size markers are in nt.
Figure 9: Conditions of the cleavage reaction.

The reaction conditions of temperature (A), time (B), and Mg$^{2+}$ concentration (C), were tested as indicated.
Figure 10: The cleavage reaction kinetics

The cleavage kinetics plot using different concentrations of substrate ([S]) vs. corresponding cleavage velocity (V), showing that the Vmax is about 0.9ng/μl*30min and the Km of the cleavage reaction of STS by NU-WT is about 1ng/μl, or 15.3nM.
<table>
<thead>
<tr>
<th>Lane</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nu-Cl</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Yeast PAP</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>1h</td>
<td>1h</td>
<td>+</td>
</tr>
<tr>
<td>3'-dATP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1h</td>
<td>-</td>
</tr>
</tbody>
</table>

**Figure 11: Reconstitution of cleavage and polyadenylation coupling reaction by supplementation of yeast PAP.**

Lane 1: control, no protein added; lane 2: addition of both Nu-Cl and yPAP resulted in polyadenylated product (2h reaction); lane 3: Nu-Cl only; lane 4: addition (1h) of yPAP after 2h cleavage reaction produced shorter poly(A); lane 5: addition of cordycepin (or 3'-dATP) inhibited the poly(A) addition; lane 6: without Nu-Cl, yPAP did not add poly(A) to the uncleaved substrate.
Figure 12: The polyadenylation reaction kinetics

A time course of cleavage and polyadenylation reactions (Nu-CI+yPAP), showing that the addition of poly(A) by yPAP was progressive at 5-10nt/10min, but faster at the beginning at a rate of 60nt/10min. RNA size markers are labeled on the right hand side of each panel in nt.