Cellular & Molecular Biology
Cell. Mol. Biol. 2015; 61 (8): 123-127
Published online December 30, 2015 (http://www.cellmolbiol.com)
Received on November 3, 2015, Accepted on December 26, 2015.
doi : 10.14715/cmb/2015.61.8.20

Cloning, Prokaryotic Expression and Purification of *CpfS1* Gene from *Arabidopsis Thaliana*

Q. He✉, A. Y. Fu✉, G. C. Zhang, T. J. Li and J. H. Zhang

Key University Laboratory of Biotechnology and Utilization of Bio-resource of Shandong, College of Life Science, Dezhou University, Dezhou 253023, China

Corresponding author: Q. He, Key University Laboratory of Biotechnology and Utilization of Bio-resource of Shandong, College of Life Science, Dezhou University, Dezhou 253023, China. Email: manuscript_sub@126.com and A.Y. Fu, Key University Laboratory of Biotechnology and Utilization of Bio-resource of Shandong, College of Life Science, Dezhou University, Dezhou 253023, China. Email: aiyun_fu@126.com

Abstract

CpfS1 Gene cloned from *arabidopsis thaliana* was expressed in *Escherichia coli* DH5α. A cDNA fragment about 320 bp was amplified from the total RNA of *arabidopsis thaliana* seeds by reverse transcription PCR (RT-PCR) with a pair of specific primers based on the sequences of the AtCpfS1 gene. The recombinant prokaryotic expression vector pET30a-AtCpfS1 was constructed by inserting the cDNA fragment encoding the mature peptide into the prokaryotic expression vector pET30a, and then transformed into *E. coli* DH5α. Sequence analysis showed that the fragment length was 346 bp containing a full coding region of 332 bp encoding 76 amino acid residues with a molecular mass of 21.5 kD. The SDS-PAGE electrophoresis analysis showed that the best expression was induced by 21°C and 3.6×10^{-3} mol/L IPTG, under which a relative molecular weight of 82.5 kD recombinant protein was produced. The nickel chelating resin was used to purify the protein in size exclusion chromatography (SEC) and the results indicated that AtCpfS1 protein was present in the form of tetramer.

Key words: AtCpfS1 gene, prokaryotic system, protein purification, SEC.

Introduction

Flowering is one of the most important processes of plant growth and transition from vegetative growth to reproductive growth. It is necessary for the development of plant to accomplish this transition. Through the regulation of blossom time, plant can achieve interspecific synchronous cross and produce as many seeds as possible, which is an adaptation of plants to environmental conditions in the process of long-term evolution. In the last decades, remarkable research achievements have been made by using *arabidopsis thaliana* as a model plant to investigate the regulation mechanisms and signal transition pathway of blossom time of advanced plants (1-3). Currently, the regulation mechanisms of plant blossom time were learned mainly through the study of the genetics of *arabidopsis thaliana*.

Blossom time of *arabidopsis thaliana* is affected by many factors, among which the light (including light quality, light intensity, photoperiod) and temperature are main external factors (4,5), Autonomous pathway factor and gibberellin (GA) are primary internal factors (6). In addition, the physiological status of the plant (such as age, plant size), stress conditions (such as drought, lack of nutrition, disease, congestion, pole temperature), plant hormone, salicylic acid, carbohydrate, vitamin C, glutathione, hydrogen peroxide, Ca^{2+} concentration, microRNA also have certain effects on the blossom time (7-9). Usually various signals assemble together to regulate the development process of shoot apical meristem (SAM) (10,11). This fine regulation to blossom time with changes in the internal physiological conditions and external environmental conditions is an adaptive advantage during the long evolutionary process of *arabidopsis thaliana*.

The regulation of *arabidopsis thaliana* to blossom time is achieved through a number of different genetic loci. It has been found that there are more than 80 loci affect blossom time of *arabidopsis thaliana*, more than 20 loci are associated with late blossom time (12). Currently the investigation toward blossom time mainly focuses on the screening of blossom time changed mutants and their cloning and functional study. According to phenotypic and genetic epistasis experiments at different environmental conditions (mainly refers to light, temperature), late flowering mutants can usually be divided into four ways: photoperiod pathway, autonomous pathway, vernalization pathway and gibberellin pathway (13).

MADS box gene is a transcription factor, which is a kind of transcription regulatory factor to play a very important role in the process of determining flowering time and flower morphology. In the process of blossom, MADS protein participate in many processes for regulation of blossom (14) through interacting with other proteins (15). However, it is unclear that whether *CpfS1* Gene only functions the formation of heteromeric complexes (16) after combining other transcription factors, or it still contains biological functions after the formation of homologous polymers. Previous studies have demonstrated that MADS protein combines with the specific DNA to form homologous polymers and performs their functions (15,16). However, the formation mechanisms of homologous polymers from MADS protein and the relationship between their functions are still unclear. This study attempts to analyze the methods for purification of soluble protein by in vitro expression, clone the *CpfS1* Gene in different domains to the prokaryotic expression vector pET-30b, followed by induced expression, and obtain high quality protein by purifica-

tion. This work provides an important foundation for expression and purification of plant protein in heterologous prokaryotic expression system.

Materials and Methods

Materials

Columbia wild *arabidopsis thaliana* were grown in Key University Laboratory of Biotechnology and Utilization of Bio-resource of Shandong, College of Life Science, Dezhou University. *Escherichia coli* DH5α, BL21, expression strain *E. coli* and expression plasmid *pET30a* were preserved in Key University Laboratory of Biotechnology and Utilization of Bio-resource of Shandong, College of Life Science, Dezhou University. Poly-Gel DNA extraction kit was purchased from Hangzhou Fode Biotech Co Ltd. M-MLV reverse transcriptase, *Taq* DNA polymerase, *Eco*RI and *Sal*I DNA restriction enzyme were purchased from BaoSheng (TaKaRa) Biotechnology (Dalian) Co Ltd. DNA ligase was purchased from New England company. DNA purification kit and gel recovery kit were purchased from Shanghai Ruijie Biotech Co Ltd. Ni-NTA superflow cartridges were purchased from Qiagen Company. TRIzol reagent was purchased from Invitrogen Company. The grade of other reagents were analytical pure.

According to the database provided by Genbank database, the full length of the *arabidopsis thaliana* AtCpfS1 gene was designed with Primer 5.0 software to produce a pair of PCR primers amplified by the database, and the primers were synthesized by Shanghai Bioengineering Co., ltd. The sequence was described as follows, upstream primer P1:5'-GATATCATG GCCGACCACAGCGCCGACG-3', the downstream primer P2:5'-GTGCA**C**TTAATGGGATACGAAACTTGGGTG-3', before the start codon ATG of the upstream primer and before the stop codon TAA of the downstream primer, *Eco*RI and *sal*I restriction enzyme cutting sites were added and displayed with the bold fonts.

Construction of pET30a prokaryotic expression vector

The seeds of *arabidopsis thaliana* were selected, and loaded in a vermiculite dish. TRIzol reagent (1 ml) was added, fully homogenized and placed at room temperature for 5 min, then 0.2 ml of chloroform was added in each ml of Trizol reagent, after intense oscillating for 15 s, the solutions were placed at room temperature for 10 min, at 4 °C, followed by centrifugation for 15 minutes at 12,000 r/min. The upper aqeous phase was transferred into a new centrifugal tube, and equal volume of isopropyl alcohol was added, placed at room temperature for 10 min to precipitate RNA. The solutions were then centrifuged at 4 °C and 12,000 r/min for 10 min, the supernatant was removed and dried in air for 15-20 min, the precipitates was dissolved in water without RNase. After extracting plasmid from the correct strain verified by sequencing, the plasmid obtained by *Eco*RI and *sal*I double-enzyme digested extraction and pMD18-T vector were loaded into DNA gel recovery kit to recover enzyme digested products, the fragments were connected with vector overnight at 16°C then transformed into *Escherichia coli* DH5α, PCR-test bacteria liquid followed by bacterial strain selection, determine base sequence of strain samples, clone to obtain pET30a prokaryotic expression vector of fragments in different structural domains.

Soluble conditions screening for recombinant proteins

The induction temperature is closely related to the solubility of prokaryotic expression protein. Under the condition of high temperature, strain rapid expressing target protein is easy to make the protein form incorrect folding and finally present in the form of inclusion body, which is not favorable for obtaining soluble protein. Correct recombinant plasmid verified by sequencing is transformed in expression host strain *E. coli* DH5α, cultured 12.0-15.0 h at 37°C, single colony is selected and transferred to 150 mL LB liquid nutrient medium containing 60.0 $mg.L^{-1}$ kanamycin, cultured 5.0 h at 37°C. When the concentration of bacteria is 0.5-0.7 at absorbance OD_{600}, isopropyl thiogalactoside (IPTG) at final concentration of 1.0×10^{-4} mol/L is added, the strains are cultured at 37°C, 3 h, and 21°C, 12 h, respectively. Determine the induction temperature for train culture.

Induction expression of pET30a protein

Monoclonal colonies were selected from the constructed expression vector - pET30a/AtCpfS1 transformed *Escherichia coli* BL21 expression of competent cells and inoculated in LB culture medium and shake at 37°C for overnight, when switching to expand training to bacteria OD_{600} reached 0.5 to 0.6, IPTG at a final concentration of 1.0×10^{-3} mol/L was added, after induction for 1, 3, and 5 h at 37°C, polyacrylamide gel electrophoresis (SDS-PAGE) analysis was conducted followed by sampling. After 5 h, thalli dissolved in solution bacteria buffer (0.05 mol/L NaH_2PO_4 0.3 mol/L NaCl, 0.01 mol/L imidazole, pH=8.0) was collected, and processing ultrasonic cell disruption, centrifuged at 12,000 r/min for 20 min, the supernatant and precipitate were taken respectively for SDS-PAGE analysis.

Purification and renaturation of recombinant pET30a protein

pET30a protein expressed in the form of inclusion bodies was purified using the Qiagen nickel column purification procedure under the denatured conditions. The purified pET30a protein was denatured according to an urea gradient dialysis method with the urea concentrations of 8, 6, 4, 2 and 0 mol/L. First, the purified and denatured pET30a protein was diluted 5 times using renaturation binding buffer (8 mol/L urea, 0.1 mol/L NaH_2PO_4, 0.1 mol/L Tris-HCl, pH=8), then loaded into dialysis bags, and placed in the dialysis liquid (0.1 mol/L Tris-HCl, 0.1 mol/L NaH_2PO_4, 1×10^{-3} mol/L EDTA, 2×10^{-3} mol/L glutathione, 2×10^{-4} mol/L oxidized glutathione, 20% glycerol, 1% glycine, pH=8). The solutions were dialyzed at 4 °C for 12 h under magnetic stirring. The dialysis solution containing low concentrations of urea was replaced every 12 h, other components remained unchanged, finally pET30a protein solution was dialyzed against water, centrifuged to remove precipitates. After freeze drying, the renaturated pET30a protein was detected by 12% SDS-PAGE and stored at -80 °C.

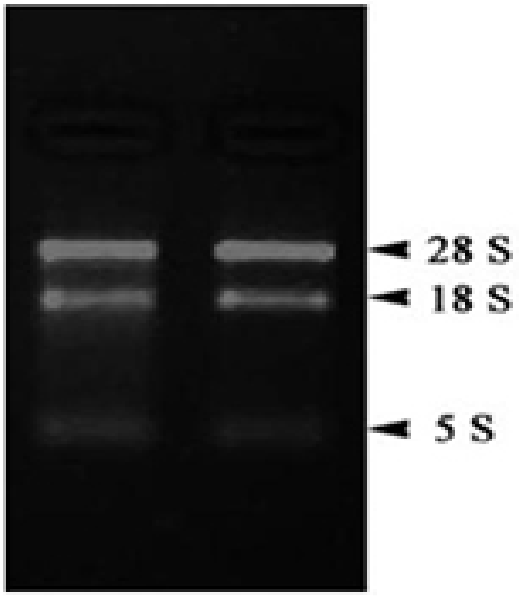

Figure 1. Total RNA of *arabidopsis thaliana* AtCpfS1 gene.

Results

Cloning and characterization of enzyme digestion of PCR products

The A_{270}/A_{290} ratio of total RNA extracted from *arabidopsis thaliana* AtCpfS1 gene is equal to 1.7, suggesting that the total RNA was highly pure and did not degrade (see Figure 1). The cDNA used as a template was obtained by reverse transcription from this RNA, through PCR amplification, a single and specified band with approximately 320 bp was yieled, which is in agreement with the expected size of target gene (see Figure 2).

Construction of pET30a-AtCpfS1 prokaryotic expression vector

PCR products and pMD18-T products were transformed into cDNA DH5α host bacteria. The positive cloning products were selected, and the specific target bands were identified by agarose gel electrophoresis. The pMD18-T/AtCpfS1 plasmid with *Eco*RI and *Sal*I was digested then yield the insert fragments (see

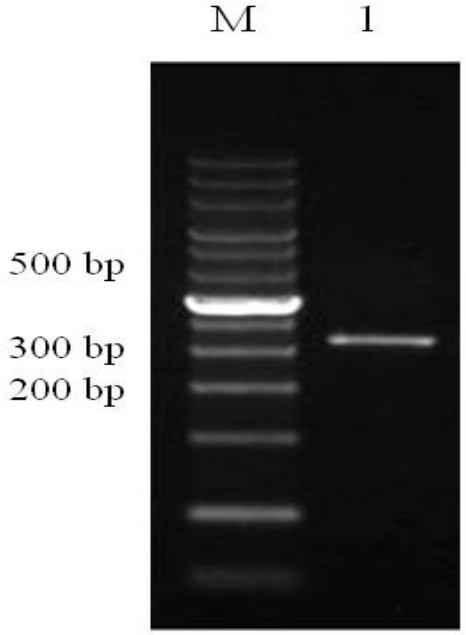

Figure 2. RT-PCR result of AtCpfS1 gene.

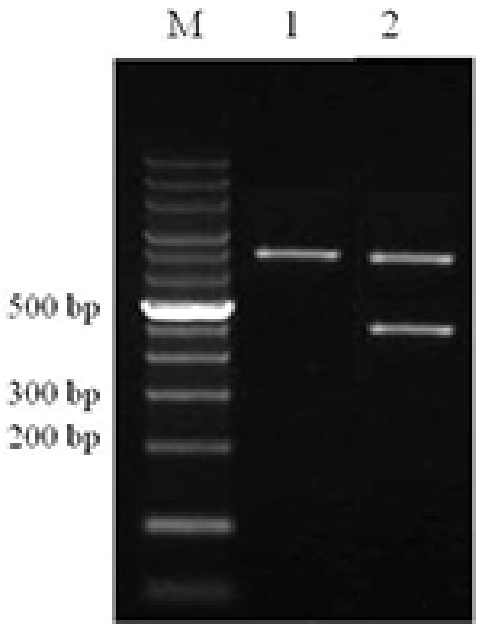

Figure 3. pMD18-T/AtCpfS1 plasmid double digested by *Eco*RI and *Sal*I.

Figure 3) after electrophoresis, which was around 320 bp. The correct recombinant plasmid was constructed and sequenced, the result showed that the obtained sequence was completely consistent with the sequence of AtCpfS1 cDNA in the GenBank database.

Culture conditions Optimization for recombinant protein

The experimental results showed that OD_{600} of bacteria liquid was equal to 0.5 after culture for 6 h. Therefore, the optimized experimental conditions for fusion expression of plasmid pET30a-AtCpfS1 transformed *Escherichia coli* DH5α are given as follows. *Escherichia coli* DH5α was singly inoculated in 3 ml of liquid LB culture medium containing 100 μg/mL of ampicillin, oscillating at 37 °C for overnight incubation. Another 500 μL was inoculated in 50 mL liquid LB culture medium containing 100 μg/mL of ampicillin in 100 mL sterile erlenmeyer flask, culture for 6 h on a shaking table at 37 °C and 250 r/min for oscillation until OD_{600} = 0.5.

IPTG induction and experimental conditions optimization

SDS-PAGE electrophoresis showed that there were no significant differences in the protein produced by pET30a-AtCpfS1 and pET30a before and after induction by various concentrations of IPTG at 16, 18, 22 and 24 °C. At 20 °C, there were also no significant differences in the protein produced by pET30a-AtCpfS1 and pET30a before and after induction by 3.0×10^{-4} mol/L and 5.0×10^{-4} mol/L IPTG. When IPTG concentration increased to 1.0×10^{-3} mol/L, a distinct protein difference displayed in Figure 4 for pET30a expression vector before and after induction, and the same difference was observed for pET30a expression vector before and after induction with IPTG concentration of $1.0\text{-}6.0\times10^{-3}$ mol/L, the molecular weight was around 55 kDa, which was the intein expressed from pET30a expression vector with molecular weight of 55 kDa induced by a certain concentration of IPTG. However, to fusion expression vector pET30a-AtCpfS1, there was no significant difference before and

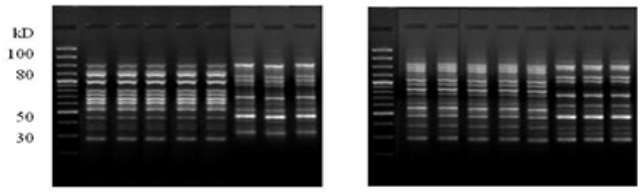

Figure 4. Identification of AtSEP3 protein by mass spectrometry (A) and Mascot Score histogram of mass fingerprinting (B).

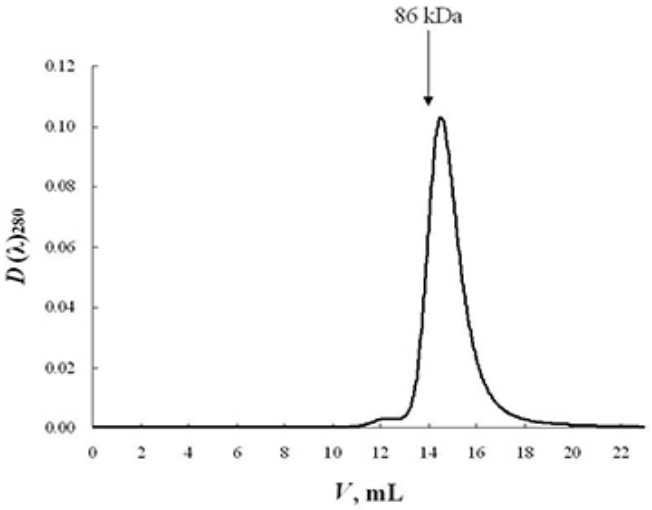

Figure 5. Gel filtration of AtCpfS1 protein.

after induction with IPTG concentration of 1.0-3.0×10^{-3} mol/L, the staining intensity of protein after induction (indicated in Figure 4) might increase a little comparing with that before induction. When IPTG concentration increased to 4.0×10^{-3} mol/L, a distinct protein difference displayed in Figure 4 for pET30a expression vector before and after induction, and the same difference was observed for pET30a expression vector before and after induction with IPTG concentration of 4.0-6.0×10^{-3} mol/L, the molecular weight was approximately 83.1 kDa. In this study, the constructed fusion expression vector pET30a-AtCpfS1 is a fusant combining intein and protein, suggesting that the differential protein is exactly the expected fusion protein. According to the experimental results, the optimal induction temperature of *Escherichia coli* DH5α cells transformed from fusion expression vector pET30a-AtCpfS1 was 20 °C, and the optimal IPTG concentration was 4.0×10^{-3} mol/L.

Purification of the soluble protein AtSEP3B

The results of electrophoresis of the purified product from each stage are shown in Figure 5. The ovalbumin (OVA) solution is used as a reference and present in the forms of monomer and dimer, with molecular weight of 43 kDa and 86 kDa, respectively. The black solid line indicates AtCpfS1, the peak position is close to 86 kDa of the OVA dimer. In Figure 6, SDS-PAGE detects sample purity after the gel filtration chromatography of protein. The arrow indicates target protein, suggesting that highly purified AtCpfS1 protein was obtained. The molecular weight of a single peptide chain is 21 kDa, and the peak in Figure 5 is approximately 86 kDa, suggesting that the purified AtCpfS1 is present in the form of tetramer.

Discussion

A convenient and efficient method to obtain a large

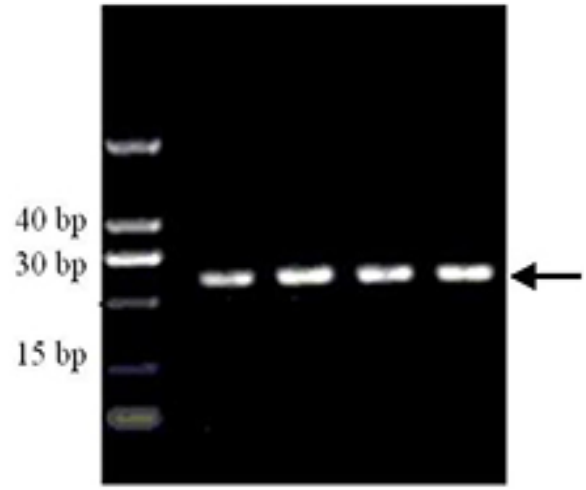

Figure 6. SDS-PAGE result of AtCpfS1 protein after purified by gel filtration chromatography.

number of recombinant proteins is expression of the target gene by *Escherichia coli*. The method has been applied in protein purification, localization and functional analysis. This study successfully constructed the prokaryotic expression vector pET30a-AtCpfS1 of *arabidopsis thaliana* CpfS1 gene which was expressed in *E. coli*. After induced by IPTG for 1 h, the protein expression was obvious and reached the peak in 3-5 h. *E. coli* for have many advantages in expressing recombinant proteins, such as easy to grow and be controlled, inexpensive materials and higher level of expression. However, the protein expressed in *Escherichia coli* usually form inclusion due to the lack of modification and glycosylation, phosphorylation, etc. In this study, the expression conditions of CpfS1 were optimized, such as decreasing temperature, reducing IPTG concentration, etc., but soluble expression of CpfS1 protein was still not obtained, most of the protein presented in the form of inclusion body. Further complex treatment of the protein by the method of dialysis yielded the soluble protein, which will provide sufficient protein for the preparation of AtCpfS1 antibody and the fundamental for further investigation of *arabidopsis thaliana*.

Prokaryotic expression vector pET30a expressing shear fusion protein can encode an intein containing 76 amino acid residues with a molecular weight of approximately 56 kD. Intein is a section of polypeptide which was expressed with functional protein and self-splicing after expression, it has the characteristics of both protease and protein ligase, and is similar with the manner of RNA self-splicing. In the fusion expression vector constructed by pET30a, intein fuses with N terminal of target protein. In this study, the AtCpfS1 gene was fused into expression vector pET30a to construct fusion expression vector pET30a-AtCpfS1, which was transformed into *Escherichia coli* DH5α for induction. The molecular weight of protein was around 82.5 kD, suggesting that the fusion protein expressed in this study was pET30a-AtCpfS1. The main factors affecting the expression of fusion protein expression vector pET30a-AtCpfS1 include: growth temperature, cell growth rate and IPTG induced concentration. The most important factor for obtaining high-level expression in *E. coli* may be the growth temperature, and a very important step to express foreign protein is the determination of optimum temperature. The results showed that the in-

duced expression of fusion expression vector pET30a-AtCpfS1 was highly efficient at 20 °C. The cell growth rate seriously affects the expression of foreign proteins. Therefore, during the expression of soluble protein in *E. coli*, the induction time and the dosage of the induction agent should be strictly controlled. During shake flask culture, it is appropriate for induction at low *Escherichia coli* concentration due to logarithmic growth phase (OD_{600} = 0.5), active growth, and is good for the expression of protein at low concentration. Growth excessive or too fast will increase the burden on the system of synthetic bacteria. The effect of IPTG concentration on expression level of the fusion expression vector pET30a-AtCpfS1 was also very large. In the experiments, expression of the fusion protein was induced at different concentration gradient. The SDS-PAGE results showed that the protein had no difference before and after IPTG induction at low ($3\text{-}5\times10^{-4}$ mol/L) and medium ($1\text{-}3\times10^{-3}$ mol/L) IPTG concentrations, but the difference was significant at high concentration ($4\text{-}6\times10^{-3}$ mol/L) of IPTG. Therefore, the concentration of 1×10^{-3} mol/L is only a start and can be considered as a "higher" concentration. In the experiment, the concentration should be optimized by expression according to setup of different gradient concentrations. According to the results of this study, the optimal temperature of fusion protein expression vector pET30a-AtCpfS1 was 20°C, and the optimal IPTG concentration was 4×10^{-3} mol/L.

In this paper, AtCpfS1 gene was first cloned into prokaryotic expression vector pET30a, the fusion expression vector pET30a-AtCpfS1 was constructed and successfully expressed in *Escherichia coli* DH5α. The protein was purified in size exclusion chromatography and the results indicated that AtCpfS1 protein was present in the form of tetramer. This study provides a scientific basis for the investigation of *arabidopsis thaliana* protein in plant genetic engineering, biotechnology and biochemistry. Further study may investigate the formation of AtCpfS1 protein tetramer and the relationship between the spatial structure and function through the methods of structural biology.

References

1. Wellmer, F., Graciet, E. and Riechmann, J. L. "Specification of floral organs in Arabidopsis" *J Exp Bot*, 2014, **65**: 1-9. doi: 10.1093/jxb/ert385

2. Xu, Y., Lascu, I. and Gursky, S. K. "Binding of Nucleotides to Nucleoside Diphosphate Kinase" *Biochemistry*, 2011, **40**: 4538–4589. doi: 10.1021/bi004232s

3. Coen, E.S. and Meyerowitz, E.M. The war of the whorls: genetic interactions controlling flower development. *Nature* 1991, **353**: 31-37. doi: 10.1038/353031a0

4. Lyons, R., Rusu, A., Stiller, J., Powell, J., Manners, J.M. and Kazan, K., Investigating the Association between Flowering Time and Defense in the Arabidopsis thaliana-Fusarium oxysporum Interaction. *PLoS One* 2015, **10:** e0127699. doi: 10.1371/journal.pone.0127699

5. Henry. Y., Bedhomme, M. and Blanc, G., History, protohistory and prehistory of the Arabidopsis thaliana chromosome complement. *Trends Plant. Sci.* 2006, **11:** 267-273. doi: 10.1016/j.tplants.2006.04.002

6. Lin, C. and Thomashow, M.F., DNA sequence analysis of a complementary DNA for cold-regulated Arabidopsis gene cor15 and characterization of the COR15 polypeptide. *Plant Physiol.* 1992, **99:** 519-525.

7. Jetha, K., Theißen, G. and Melzer, R. "Arabidopsis SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes" *Nucl Acids Res* 2014, **42**: 10927-10942. doi: 10.1093/nar/gku755

8. Arenas-Mena, C., Raynal, M., Borrell, A., Varoquaux, F., Cruz Cutanda, M., Stacy, R.A.P., Pages, M., Delseny, M. and Culianez-Macia, F.A., Expression and cellular localization of Atrab28 during Arabidopsis embryogenesis. *Plant Mol. Biol.* 1999, **40:** 355-363. doi: 10.1023/A:1006219315562

9. Rouse, D.T., Marotta, R., and Parish, R.W., Promoter and expression studies on an Arabidopsis thaliana dehydrin gene. *FEBS Lett.* 1996, **381:** 252-256. doi: 10.1016/0014-5793(96)00051-8

10. Blanc, G. and Wolfe, K.H., Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell.* 2004, **16:** 1679-1691. doi: 10.1105/tpc.021410.

11. Casneuf, T., de Bodt, S., Raes, J., Maere, S. and van de Peer, Y., Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant Arabidopsis thaliana. *Genome Biol.* 2006, **7:** R13. doi: 10.1186/gb-2006-7-2-r13.

12. Assunção, A.G.L., Herrero, E., Lin, Y.-F., Huettel, B., Talukdar, S., Smaczniak, C., Immink, R.G.H., van Eldik, M., Fiers, M., Schat, H. and Aarts, M.G.M., Arabidopsis thaliana transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. *Proc. Nat. Acad. Sci. USA* 2010, **107**: 10296-10301. doi: 10.1073/pnas.1004788107

13. Rohde, P., Hincha, D.K. and Heyer, A.G., Heterosis in the freezing tolerance of crosses between two Arabidopsis thaliana accessions (Columbia-0 and C24) that show differences in non-acclimated and acclimated freezing tolerance. *Plant J.* 2004, **38:** 790-799. doi: 10.1111/j.1365-313X.2004.02080.x

14. Artus, N.N., Uemura, M,. Steponkus, P.L., Gilmour, S.J., Lin, C. and Thomashow, M.F., Constitutive expression of the cold-regulated Arabidopsis thaliana COR15a gene affects both chloroplast and protoplast freezing tolerance. *Proc. Natl. Acad. Sci. USA* 1996, **93:** 13404-13409. doi: 10.1073/pnas.93.23.13404

15. Dunaeva, M. and Adamska, I., Identification of genes expressed in response to light stress in leaves of Arabidopsis thaliana using RNA differential display. *Eur. J. Biochem.* 2001, **268:** 5521-5529. doi: 10.1046/j.1432-1033.2001.02471.x

16. Kimura, M., Yamamoto, Y.Y., Seki, M., Sakurai, T., Sato, M., Abe, T., Yoshida, S., Manabe, K., Shinozaki, K. and Matsui, M., Identification of Arabidopsis genes regulated by high light-stress using cDNA microarray. *Photochem. Photobiol.* 2003, **77:** 226-233. doi: 10.1562/0031-8655(2003)0770226IOAGRB2.0.CO2