Plant Science 321 (2022) 111309

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Functional diversity of *Medicago truncatula* RNA polymerase II CTD phosphatase isoforms produced in the *Arabidopsis thaliana* superexpression platform

Akihito Fukudome [a], Yasuhiro Ishiga [b,c], Yukihiro Nagashima [d], Katherine H. Davidson [d], Hsiu-An Chou [a], Kirankumar S. Mysore [b,e,f], Hisashi Koiwa [a,d,*]

[a] *Molecular and Environmental Plant Sciences, Texas A&M University, College Station, TX 77843, USA*
[b] *Noble Research Institute, LLC., Ardmore, OK 73401, USA*
[c] *Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan*
[d] *Vegetable and Fruit Development Center, Department of Horticultural Sciences, Texas A&M University, College Station, TX 77843, USA*
[e] *Institute for Agricultural Biosciences, Oklahoma State University, Ardmore, Oklahoma 73401, USA*
[f] *Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, Oklahoma 74044, USA*

ARTICLE INFO

Keywords:
Medicago truncatula
CTD phosphatase
RNA polymerase II
Plant recombinant protein
Arabidopsis thaliana

ABSTRACT

Medicago truncatula is a model system for legume plants, which has substantially expanded the genome relative to the prototypical model dicot plant, *Arabidopsis thaliana*. An essential transcriptional regulator, FCP1 (transcription factor II*F*-interacting RNA polymerase II *c*arboxyl-terminal *p*hosphatase 1) ortholog, is encoded by a single essential gene *CPL4* (CTD-phosphatase-like 4), whereas *M. truncatula* genome contains four genes homologous to FCP1/AtCPL4, and splicing variants of MtCPL4 are observed. Functional diversification of MtCPL4 family proteins was analyzed using recombinant proteins (MtCPL4a1, MtCPL4a2, and MtCPL4b) produced in Arabidopsis cell culture system developed for plant protein overexpression. In vitro CTD phosphatase assay using highly purified MtCPL4 preparations revealed a potent CTD phosphatase activity in MtCPL4b, but not two splicing variants of MtCPL4a. On the other hand, in planta binding assay to RNA polymerase II (pol II) revealed a greater pol II-binding activity of both MtCPL4a variants. Our results indicate functional diversification of MtCPL4 isoforms and suggest the presence of a large number of functionally specialized CTD-phosphatase-like proteins in plants.

1. Introduction

Regulatory phosphorylations of RNA polymerase II (pol II) occurs at the carboxyl-terminal domain (CTD) of its largest subunit, RPB1. The CTD contains tandemly repeated heptads with a consensus sequence of Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (Nawrath et al., 1990). The number of repeats and degree of deviation from the consensus sequence correlates roughly with evolutionary complexity; mammals have 52 repeats (Corden et al., 1985), drosophila has 42 repeats (Zehring et al., 1988), and Arabidopsis CTD contains 34 repeats (Dietrich et al., 1990; Hajheidari et al., 2013). The CTD undergoes waves of phosphorylation and dephosphorylation during the transcription cycle by the function of various position-specific CTD kinases and phosphatases (Dahmus, 1996). Studies in animals and fungi indicated that all residues in CTD heptads except prolines, i.e., Tyr1, Ser2, Thr4, Ser5, and Ser7, are targets of regulatory phosphorylations (Hsin and Manley, 2012). In addition, *cis-trans* isomerization of Pro in CTD heptads modulate dephosphorylation of CTD by phosphatases (Werner-Allen et al., 2011). Phosphorylation marks placed on different residues in each repeat generate CTD-codes uniquely associated with pol II transcription status (Hajheidari et al., 2013; Eick and Geyer, 2013). Phosphorylated CTD is a target of CTD-associating proteins that recognize various CTD-codes and differentially regulate transcription as well as co-transcriptional RNA processing (Hsin and Manley, 2012). Ser2 and Ser5 phosphorylation

Abbreviations: CTD, carboxyl-terminal domain; CPL, CTD-phosphatase-like; pol II, RNA polymerase II; FCP1, Transcription Factor II*F*-interacting *C*TD-*p*hosphatases 1; AMP, Asp-based metal-dependent phosphatases; TAP, tandem-affinity purification.

* Correspondence to: Molecular and Environmental Plant Sciences, Department of Horticultural Sciences, Vegetable and Fruit Development Center, Texas A&M University, College Station, TX 77843, USA.
E-mail address: hisashi.koiwa@ag.tamu.edu (H. Koiwa).

https://doi.org/10.1016/j.plantsci.2022.111309
Received 1 January 2022; Received in revised form 25 April 2022; Accepted 29 April 2022
Available online 10 May 2022

have multiple roles during mRNA transcription, including capping and elongation of transcripts (Hsin and Manley, 2012).

Proper CTD dephosphorylation, too, is essential for the transcription cycle. Prototypical CTD phosphatase is Asp-based metal-dependent phosphatases domain (AMPs) (Thaller et al., 1998), termed FCP1 (Transcription Factor II*F*-interacting *C*TD-*p*hosphatases 1). The signature motif of the AMPs is ϕϕϕDxDxT/Sxxϕϕ (ϕ indicates aliphatic amino acids), with the first aspartic acid residue functioning as a phosphoryl acceptor and the second stabilizing the leaving phosphate group (Kobor et al., 1999; Wang et al., 2002). Our current understandings of plant CTD-phosphatase functions are primarily obtained from studies using the model plant *Arabidopsis thaliana*. Arabidopsis has multiple CTD-phosphatase-like (CPL) AMPs that are classified into three groups, CPL1-like, CPL3-like, and SCP1-like small phosphatases (SSPs) (Koiwa, 2006; Bang et al., 2006). The CPL1-like proteins contain additional C-terminal double-stranded RNA-binding domains (Koiwa, 2004; Koiwa et al., 2002). The CPL3-like-proteins including CPL4 resemble prototypical CTD phosphatase FCP1 and contain a C-terminal BRCT (Breast cancer 1 C-terminus) domain (Bang et al., 2006; Koiwa et al., 2002). The CPL5 and SSP proteins contain only a phosphatase catalytic domain (Jin et al., 2011; Feng et al., 2010). CPL1-like proteins can dephosphorylate CTD-Ser5-PO_4 specifically (Koiwa et al., 2004), whereas SSPs contain both Ser5- PO_4-specific and Ser2/Ser5-PO_4-specific members (Feng et al., 2010).

CPL3 and CPL4 are highly homologous to human and fungal FCP1 (Fukudome et al., 2014). While several studies characterized CPL1-like proteins in different plant species, studies on CPL3-like phosphatase are scarce and limited to *Arabidopsis thaliana*. CPL3 is dispensable for completing the plant life cycle but is required for ABA/pathogen-responsive gene expression, whereas CPL4 is essential for the growth and development of Arabidopsis. Knocking down CPL4 expression by RNAi (CPL4RNAi) induces global CTD hyperphosphorylation and upregulation of more than 200 genes, indicating that CPL4 functions as a major CTD phosphatase in Arabidopsis (Fukudome et al., 2014). In this study, we investigated the diversity of CPL functions, particularly CPL4, in the model legume plants, *Medicago truncatula.* The MtCPL family is more expanded than the AtCPL family and has four paralogs, even for the single CPL4 subclass. To overcome the challenge in producing recombinant proteins in the heterologous non-plant systems, we used the Arabidopsis superexpression platform to produce recombinant MtCPL4 proteins and asked if expanded MtCPL4 created isoform-specific biochemical functions. The results uncovered distinct characteristics of MtCPL4 isoforms indicating the regulatory role of CPL family proteins in plant gene expression.

2. Materials and methods

2.1. Plant materials and the growth condition

Arabidopsis thaliana Col-0 rdr6–11 was used for plant expression host. For floral transformation, plants were grown on a growth medium (Sunshine Mix #1 Fafard-1 P, Sun Gro Horticulture) in a growth chamber set to 23 °C and 16-h-light (PHILLIPS F96T12HL41) versus 8-h-dark conditions.

2.2. Overexpression of CPL4 in Arabidopsis cells

Cloning of MtCPL4a1 (Medtr2g048610) and MtCPL4b (Medtr6g092010) cDNA fragments from M. truncatula cDNA library will be described elsewhere. cDNA fragment of the splicing variant MtCPL4a2 was derived from MtCPL4a1 using QuikChange™ protocol with primers (QCMtCPL4aF, ACATACACAAGGGACTAAGA cttcatgaggaggaaatttctagagt;.

QCMtCPL4aR, TCTTAGTCCCTTGTGTATGTatccaaatgtcaaaccagaatcac). MtCPL4a2 and MtCPL4b was cloned in Gateway entry plasmid pEnEOimCherryFSGThsp (GenBank accession number KF537341) upstream of 3xFLAG-tag and a TAP-tag (Van Leene et al., 2008). MtCPL4a2 was cloned in pEnEOimCherryF3SG-GFP-Thsp. The resulting expression cassettes encode C-terminally TAP-tagged MtCPL4a1/CPL4b, and C-terminally TAP-GFP tagged MtCPL4a2. The CPL4 entry plasmids were recombined with pMDC99 (Curtis and Grossniklaus, 2003) and then introduced into Agrobacterium tumefaciens GV3101 for Arabidopsis flower transformation. Selection of transformants and induction of cell culture were performed as described previously (Fukudome et al., 2014; Jeong et al., 2018). Tandem affinity purification was performed as described previously (Fukudome et al., 2014) with 20 g cells as starting material.

2.3. Subcellular localization analysis of GFP-fusion proteins

pEnEOiMtCPL4bF3SG-GFP-Thsp was prepared by inserting GFP between TAP tag and stop codon in pEnEOiMtCPL4bF3SG-Thsp. This was recombined with pMDC99 and introduced into GV3101. Preparation of GFP-tagged AtCPL4 and Agroinfiltrations were performed as described previously (Fukudome et al., 2014).

2.4. CTD phosphatase assay

For in vitro CTD phosphatase assays with recombinant CTD protein substrates, phosphatase reaction mixtures (10 µl) containing 50 mM Tris-acetate [pH 6.5], 10 mM $MgCl_2$, 1 µg of phosphorylated GST-AtCTD, and indicated amount of TAP-purified CPL4 were incubated for 90 min. At the end of incubation, reactions were stopped by the addition of 3.35 µl 4x SDS loading buffer plus 36.5 µl 1x SDS loading buffer and heating at 100 °C for 5 min. The Ser5, Ser2, and Ser7 phosphorylation status were analyzed by immunoblotting analyses as described (Koiwa et al., 2004; Fukudome et al., 2014).

2.5. Analysis in vivo pol II phosphorylation status

Callus tissues (0.2 g) were harvested to 1.5 ml microcentrifuge tubes, frozen in liquid nitrogen, and ground to a fine powder. Cellytic P buffer (Sigma) with proteinase inhibitor cocktail (Sigma) was added to the powder (2 ml/gFW) and mixed thoroughly. The homogenate was centrifuged at 25000xg for 5 min at 4 °C. The supernatant was mixed with an equal volume of 2x SDS-PAGE loading buffer and boiled for 5 min. For detection of RPB1 and the phosphorylation status, 12 µg of proteins from each sample were resolved by 5% SDS-PAGE and transferred to nitrocellulose membrane by semi-dry blotting with transfer buffer containing 25 mM Tris-base, 192 mM Glycine, and 0.05% SDS. The membrane was washed with distilled water twice and blocked with TBS buffer containing 6% skim milk for 1 hr. The filter was washed twice with TTBS for 5 min. The first antibodies used were rabbit anti-RPB1 (Kang et al., 2009), 3E10 (αSer2P, rat, Millipore), H14 (αSer5P, mouse, Covance), 4E12 (αSer7P, rat, Active Motif). The blots were incubated with the first antibody at four °C for overnight. After washing four times with TTBS for 5 min, the blots were incubated with HRP-conjugated second antibodies for 1 h at room temperature. After washing four times with TTBS for 5 min, chemiluminescence signals were developed with SuperSignal West Femto Chemiluminescent Substrate (Pierce), and images were acquired using EMCCD camera (Photometrics) and processed by WinView software (Roper Scientific).

2.6. Affinity co-precipitation assay

Crude cell extracts were prepared from approximately 0.6 g of cells as described above. Crude extracts (1 ml, 8 mg protein) were mixed with IgG sepharose (30 µl bead volume, prewashed with Cellytic P buffer) and incubated for 2 h at 4 °C with constant rotation. Beads were recovered by centrifugation at 1500 xg for 1 min and washed once with Cellytic P buffer and then twice with TEV cleavage buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5% NP-40, 0.5 mM EDTA, and 1 mM DTT). TAP-tagged

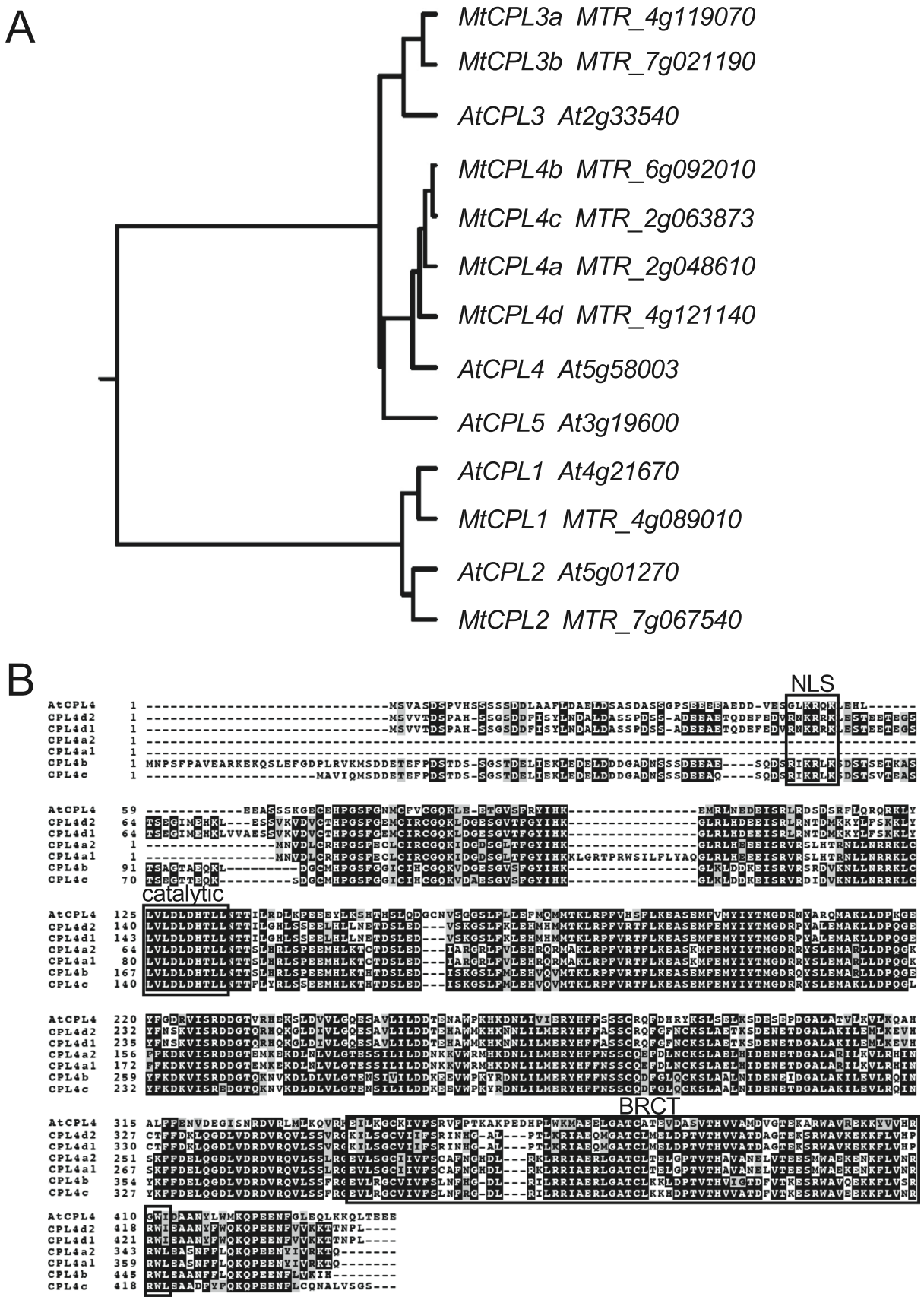

Fig. 1. CPL families in Medicago and Arabidopsis. A) Phylogenetic tree of Medicago and Arabidopsis CPL family proteins. B) Amino acid sequence alignment of CPL4 proteins. Putative nuclear localization signal (NLS), acid phosphatase catalytic motif, and BRCT domain are boxed.

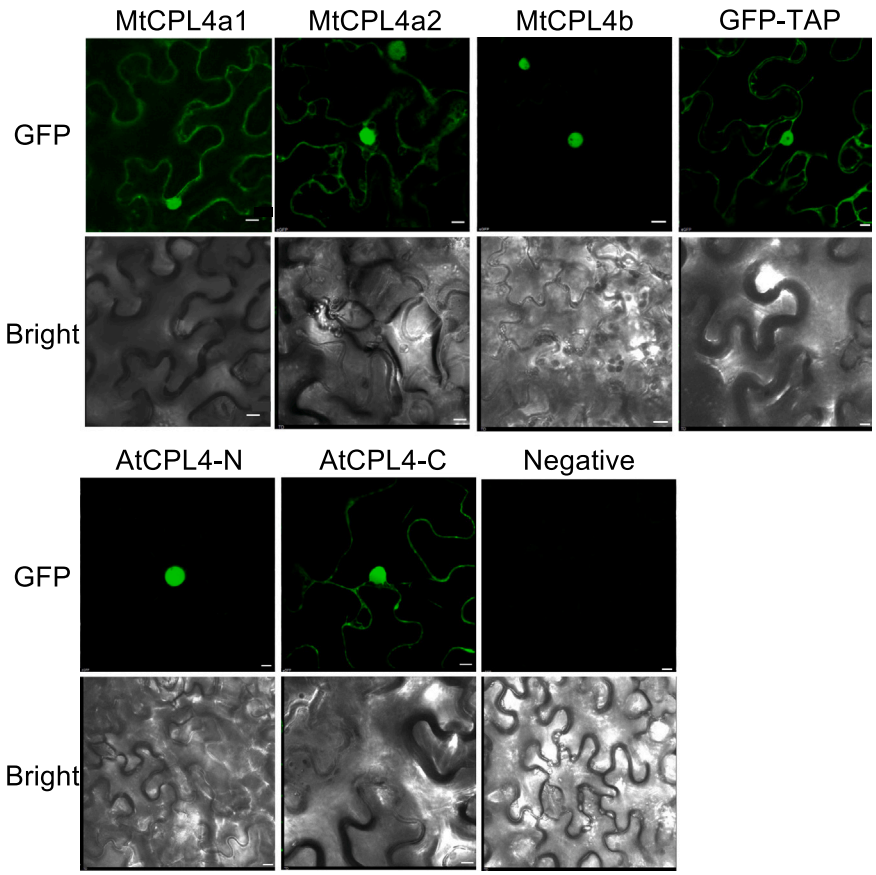

Fig. 2. Distinct subcellular localization patterns of MtCPL4 paralogs. MtCPL4-GFP fusion proteins were transiently expressed in *Nicotiana benthamiana* using Agrobacterium infiltration. GFP fused to N-terminal or C-terminal fragment of Arabidopsis CPL4 (AtCPL4-N, AtCPL4-C) or TAP tag (GFP-TAP) were used for references for nuclear (AtCPL4-N) or nuclear-cytoplasmic (AtCPL4-C, GFP-TAP) locations. Negative control indicates leaves infiltrated only with helper Agrobacterium expressing P19 protein. Bars indicate 10 μm.

proteins were eluted from the beads by cleavage with 32 μg tobacco etch virus (TEV) protease (Blommel and Fox, 2007) in 50 μl TEV cleavage buffer overnight at 4 °C. The eluted proteins were precipitated with TCA precipitation and dissolved in 20 μl 1x SDS-PAGE loading buffer. Recovered proteins were analyzed by anti-RPB1 immunoblot.

3. Results

3.1. CPL family in M. truncatula

In the Medicago genome sequence, homologs for all four major classes of CTD-phosphatase-like proteins (CPL) are present, of which CPL4 class forms a small gene family with four members (CPL4a to CPL4d) (Fig. 1A). Splicing variants have been detected for MtCPL4a and CPL4d. The longer CPL4 isoforms (isoform 1) contains 16 and 3 extra amino acids, respectively, compared to isoform 2 for CPL4a and CPL4d, respectively (Fig. 1B). Acid phosphatase catalytic motif ϕϕϕDxDxT/Sxxϕϕ site and C-terminal BRCT (breast cancer C-terminal) domains are conserved in all paralogs and isoforms (Fig. 1B), and their overall sequence are highly homologous. However, there are several critical differences in the amino acid sequences. First, the annotated MtCPL4a sequence has a truncation at N-terminus and is approximately 70 amino acids shorter than other MtCPL4s. Inspection of the upstream genomic region up to 4 kbp did not find sequence homologous to the truncated sequence, indicating MtCPL4a shown here is indeed a full-length protein. The second difference is 16 amino acid insertion in MtCPL4a1 relative to MtCPL4a2 due to alternative splicing. The N-terminal truncation and the insertion occur in the conserved catalytic domain, potentially impacting the phosphatase activity in MtCPL4a. The N-terminal truncation also eliminated the putative nuclear localization signal (R^{75}IKRLK^{80} in MtCPL4b, Fig. 1B). To test if MtCPL4 paralogs show distinct subcellular localization patterns, GFP fusion proteins for MtCPL4a isoforms and MtCPL4b were transiently expressed in *Nicotiana benthamiana* leaves. MtCPL4b and AtCPL4N (amino acid 1–296: nuclear control) produced the typical nuclear-localized expression pattern, whereas MtCPL4a1, MtCPL4a2, AtCPL4C (amino acid 297–440), and GFP-TAP tag (see below for TAP structure: nuclear-cytoplasmic control) showed nuclear-cytoplasmic GFP signals characteristic of the protein without specific targeting signals (Fig. 2). These results indicate MtCPL4 family proteins have distinct subcellular localization patterns despite a common predicted function in transcription regulation.

3.2. MtCPL4b but not MtCPL4a is an active CTD Phosphatase

To investigate the molecular function of MtCPL4, we analyzed the enzymatic activity of MtCPL4. In this study, we focused on MtCPL4a1, MtCPL4a2 and MtCPL4b. Our initial attempts to express MtCPL4a isoforms in *E. coli* did not produce catalytically active proteins. Therefore, we expressed MtCPL4 proteins using the Arabidopsis superexpression platform, which successfully produced AtCPL4 previously (Fukudome et al., 2014) MtCPL4a1 and MtCPL4b was fused to C-terminal tag consisting of 3xFLAG-streptavidin-binding peptide-TEV protease cleavage site-2x Protein G domains (TAP-tag). MtCPL4a2 was fused to C-terminal 3x-FLAG-TAP tag-GFP, a TAP tag improved in the detection of the transgene expression (Fig. 3A). After flower transformation and selection of transformed T1 seedlings using anti-FLAG immunoblots, cell lines expressing MtCPL4a1, MtCPL4a2, or MtCPL4b were established, and recombinant proteins were purified using a tandem affinity purification system (Fig. 3B). As a positive control, a cell line expressing AtCPL4 was prepared as described previously (Fukudome et al., 2014).

When CTD phosphatase activity was measured using GST-CTD-PO4 as substrate, MtCPL4b but neither isoform of MtCPL4a showed phosphatase activities (Fig. 3C). We did not succeed in detecting phosphatase activities of MtCPL4a using a general phosphatase substrate (p-

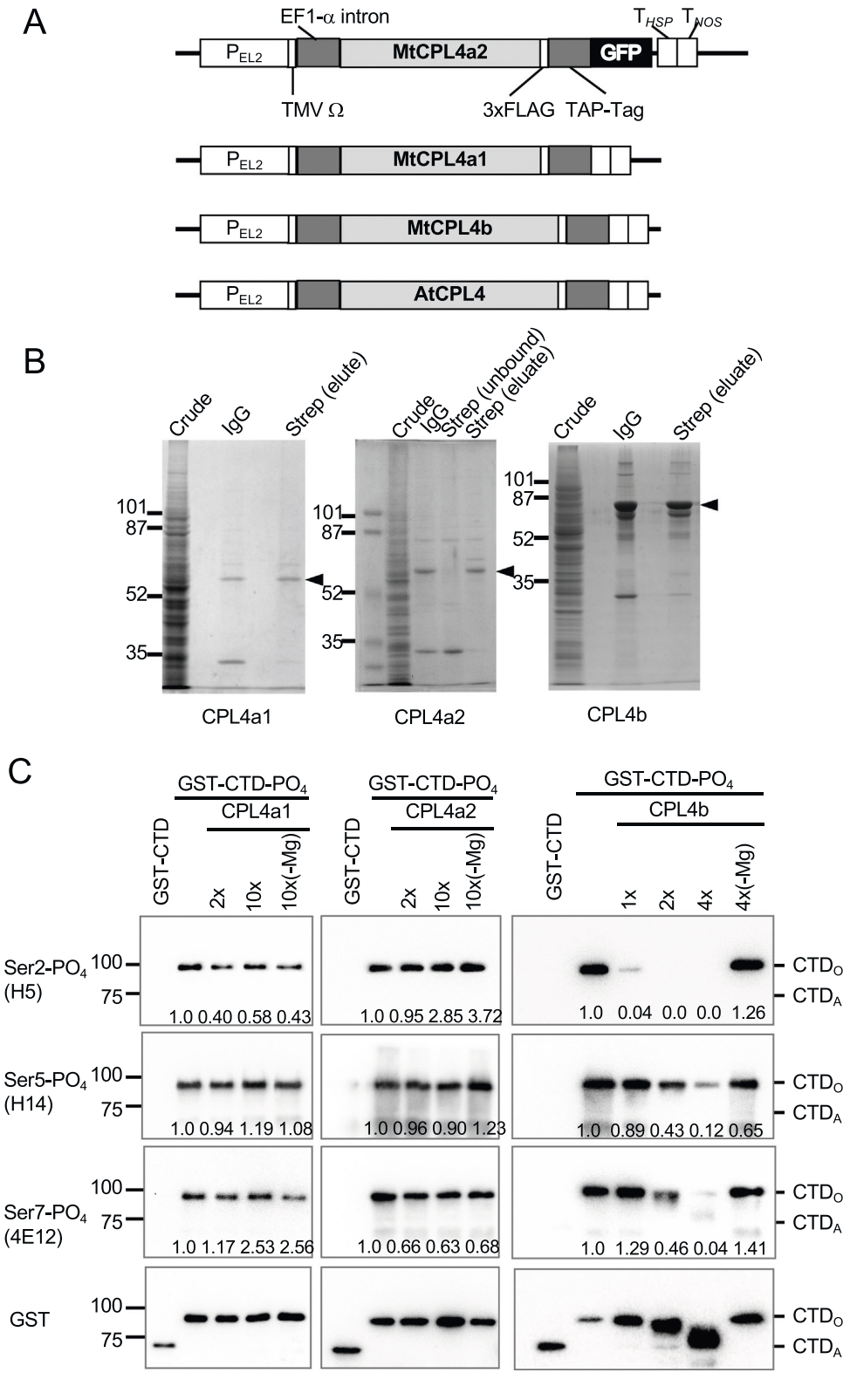

Fig. 3. MtCPL4b but not MtCPL4a produced CTD phosphatase activity. A) Schematic drawing of MtCPL4 and AtCPL4 expression cassettes. B) Tandem affinity purification of MtCPL4a1, MtCPL4a2 and MtCPL4b. Protein samples were taken from the crude extracts, the eluates from IgG-sepharose, and the eluates from streptavidin sepharose and separated by SDS-PAGE. Arrowheads indicate bands for CPL4 peptides. Note that GFP is cleaved from MtCPL4a2 at the elution step of IgG affinity purification. C) CTD phosphatase activity of CPL4 using recombinant GST-CTD-PO_4 substrate. GST-CTD-PO_4 were incubated with the indicated amount of CPL4 for 90 min and resolved on 5% SDS-PAGE gel and subjected to immunoblot to visualize phosphorylation status of GST-CTD. 1x CPL4 corresponds to 0.25 μg purified recombinant CPL4. The total substrate was visualized using HRP-conjugated anti-GST. Relative signal levels were quantified and listed under the image for each Ser-PO_4 blot.

nitrophenyl phosphate), either, suggesting MtCPL4a is catalytically inactive. MtCPL4b showed position preference for Ser2-PO4 in the CTD heptad but could act on Ser5 and Ser7, albeit less efficiently. These results indicate that the MtCPL4b function is similar to prototypical FCP1 and AtCPL4, but the function of MtCPL4a is likely not direct dephosphorylation of the pol II CTD.

3.3. Both MtCPL4a and MtCPL4b interact with pol II in vivo

To test if MtCPL4 affects pol II in vivo, the cell lines expressing MtCPL4 were evaluated for their pol II phosphorylation status using position-specific anti-phospho-CTD antibodies (Fig. 4). Like the previously reported CPL4 overexpression cell line, MtCPL4 overexpressing lines showed a modest decrease of phosphorylation levels of CTD Ser2,

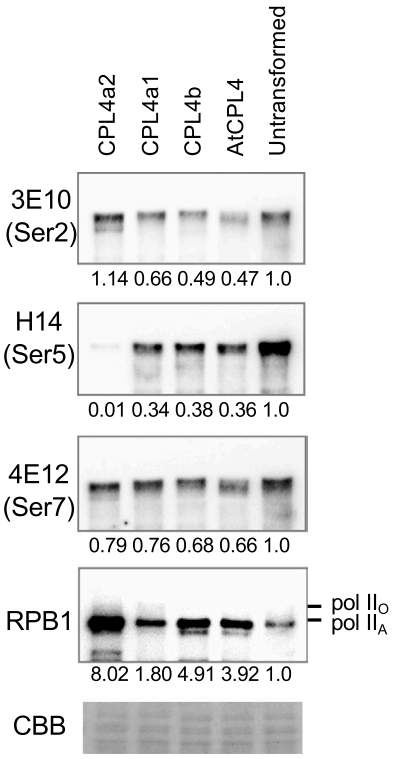

Fig. 4. Overexpression of CPL4 affects CTD phosphorylation status in vivo. Immunoblots of Arabidopsis cells overexpressing MtCPL4a1, MtCPL4a2, MtCPL4b, and Arabidopsis CPL4 as well as untransformed Arabidopsis cells are shown. Twelve micrograms of crude cell-free extracts were resolved on 5% SDS-PAGE gel and probed with anti-CTD-PO_4 antibodies. Anti-RPB1 is specific to the N-terminal region of the pol II RPB1 subunit and shows total RPB1 proteins levels. The bottom two panels represent CBB-stained low molecular weight proteins on the same gel as a loading control. Relative signal levels were quantified and listed under the image for each blot.

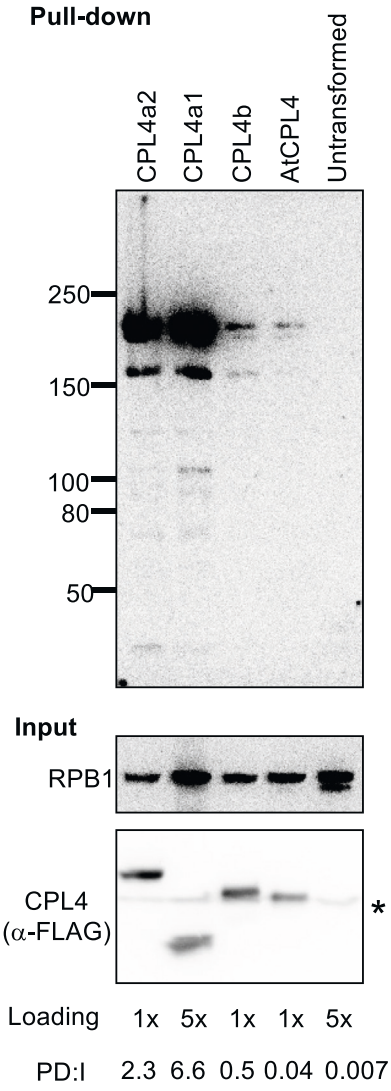

Fig. 5. CPL4 associates with pol II in vivo. Coprecipitation of pol II with CPL4 overexpressed in Arabidopsis cells. Crude extracts of Arabidopsis cells expressing TAP-tagged CPL4 were incubated with IgG sepharose, and bound CPL4 proteins were eluted by boiling in the SDS-PAGE loading buffer. RPB1 subunit of pol II coprecipitated with CPL4 was analyzed by anti-RPB immunoblot. Due to the different levels of pol II accumulation in transformed cells, the loading amount (1x or 5x) was normalized based on the anti-RPB signal levels in the crude extract (see Fig. 4). For coprecipitated RPB1, 1x loading corresponds to pol II precipitated from 0.6 mg crude protein. The level of CPL4 in the input (1x = 9 μg crude protein) was visualized using an HRP-conjugated anti-FLAG antibody. *non-specific band. Note MtCPL4a2 is fused to GFP and appears larger than MtCPL4a1. PD:I, signal ratio of Pull-down: Input for each sample.

Ser5, Ser7 (Fig. 4), and increase of hypophosphorylated pol II (pol IIA form, Fig. 4, RPB1 panel), which is detected by an antibody specific to N-terminal region of pol II. The increase of pol IIA form was the most evident with the MtCPL4a2 cell line, which also showed the lowest Ser5 phosphorylation levels.

The above results showed that MtCPL4a affects in vivo CTD phosphorylation status even though it lacks in vitro CTD phosphatase activity. We hypothesized that MtCPL4a could have a phosphatase activity-independent function, reported in human FCP1 (Mandal et al., 2002). To test this model, we tested if MtCPL4 proteins can interact with pol II in vivo. Fig. 5 shows the results from CPL4-pol II co-precipitation assays. Because the total pol II levels in the cell lines varied substantially (Fig. 4, RPB1 panel), we adjusted the cell extract input (1x or 5x) depending on the cell line. Arabidopsis CPL4 was used as a positive control (Bang et al., 2006; Koiwa et al., 2002). Surprisingly, both isoforms of MtCPL4a were strongly associated with pol II. MtCPL4b, on the other hand, produced a similar level of signal to AtCPL4, which is consistent with the similar phosphatase characteristics of MtCPL4b and AtCPL4. These results suggest that MtCPL4a isoforms affect pol II CTD phosphorylation status, and in turn, pol II function independent of its catalytic activity.

4. Discussion

Production of recombinant protein phosphatases for biochemical characterization often poses a significant technical challenge due to their tendency to produce inactive proteins in *E. coli* and the requirement of specific cofactors or post-translational modifications from the host cells. Plant-based recombinant protein production offers advantages as it can provide proteins environments like their natural occurrence. Cell culture-based expression platform allowed selection of individuals that show high expression typically disadvantageous to the whole plants, and cell cultures were maintained without expression loss often associated with sexual reproduction. Here we expressed three forms of MtCPL4 in Arabidopsis cells. We also implemented a cleavable GFP tag that improved the selection of transformants with high expression levels. Recombinant MtCPL4b was catalytically active and dephosphorylated Ser2, Ser5, and Ser7-PO_4, whereas MtCPL4a1 and 2 did not show any activities in vitro. The stable expression in Arabidopsis cells also allowed in vivo functional tests of MtCPL4. These results, together with our recent application of the Arabidopsis cell system to

express a recombinant antibody (Kang et al., 2021), indicate that Arabidopsis superexpression is a viable option for in vitro and in vivo analyses of a potentially wide range of proteins. We also noted, however, the limitation of interpreting Arabidopsis system results because we have not succeeded in complementing the lethal phenotype of the T-DNA-inserted *cpl4–2* mutant neither by MtCPL4a1 nor by MtCPL4b, whereas the genomic fragment of AtCPL4 was able to complement *cpl4–2* (data not shown). This suggests that combination of multiple MtCPL4 protein is required for replacing AtCPL4 function in vivo; however, due to the nature of the heterologous complementation, we cannot exclude other possibilities.

Among the Medicago CPL family proteins, CPL4 is the most diversified class with four paralogs and at least two splicing variants. Our results indicated that the expansion of the CPL4 class is also associated with functional diversifications. While the property of MtCPL4b was similar to prototypical FCP1 and AtCPL4 (Fukudome et al., 2014; Hausmann et al., 2003), MtCPL4a proteins showed characteristics much deviated from AtCPL4. Unlike AtCPL4, MtCPL4a isoforms lack putative NLS and did not specifically localized to nuclei but produced nuclear-cytoplasmic location, typical for proteins that enter nuclei by passive diffusion (Timney et al., 2016). The pol II interactions with both MtCPL4a isoforms were even stronger than those detected for AtCPL4 or MtCPL4, which likely contribute to the retention of MtCPL4a proteins in the nuclei. The interaction occurs even though MtCPL4a isoforms do not dephosphorylate pol II, and MtCPL4a1 has a 16-amino acid insertion in the conserved region upstream of the catalytic motif. Given the observation that the overexpression of MtCPL4a could still cause CTD hypophosphorylation of pol II in the host cell, MtCPL4a isoforms may be pseudoenzyme that lack catalytic activity, i.e., pseudophosphatase like STYX (Hinton, 2020; Reiterer et al., 2017), but establish stable complex with pol II and provide a scaffold to other regulators of pol II including other CTD phosphatases. Indeed, human CTD phosphatase RPAP2 requires scaffold proteins RPRD1A RPRD1B that preferentially bind a phosphorylated CTD (Ni et al., 2014). So far, we did not succeed in detecting specific protein-protein interactions between MtCPL4a2 and known Arabidopsis CTD phosphatases, CPL1, CPL3, CPL4, and SSP4b (data not shown), however, other CTD phosphatases in Arabidopsis could be responsible for the observation (Koiwa, 2006; Zheng, 2022). Arabidopsis genome encodes numerous CTD phosphatase-like enzymes, and many CTD phosphatase candidates have not been characterized in plants. Other possibility is that MtCPL4a proteins prevent the CTD kinase action. The stronger CTD Ser5 hypophosphorylation in the MtCPL4a2 cells than in the MtCPL4a1 cells may indicate that alternative splicing fine-tunes the functionality of MtCPL4a; however, the expression level of MtCPL4a1 was relatively lower than that of MtCPL4a2. Indeed, GFP-assisted screening enabled by the improved TAP vector was necessary to identify transformants expressing a high level of MtCPL4a2. Therefore, we cannot exclude the possibility that isoform-specific effect on in vivo phosphorylation levels in MtCPL4a expressing cells was due to the higher expression levels of MtCPL4a2 than MtCPL4a1.

Unlike other classes of CPL, the CPL4 class in Medicago was expanded to 4 genes. To date, paralog-specific roles of the CPL4 family in transcription regulation or plant physiology have not been established. AtCPL4, the only CPL4 protein characterized previously, is a single essential protein that controls multiple processes, including xenobiotic resistance, lateral root formation, cell redifferentiation (Fukudome et al., 2014; Fukudome et al., 2018; Fukudome and Koiwa, 2018). At the molecular level, AtCPL4 regulates snRNA to mRNA switching by affecting snRNA transcription termination, likely in addition to the maintenance of general pol II transcription (Fukudome et al., 2017). Considering the diverse function of other CPL class members (Li et al., 2014; Rodriguez-Cazorla et al., 2018; Shen et al., 2021; Thatcher et al., 2018) as well as the cytoplasmic presence of MtCPL4a, it is likely that in Medicago, in vivo physiological function of MtCPL4 also has undergone substantial diversification. It is expected that further genetic studies on MtCPL4 members likely delineate how the group of MtCPL4 coordinates transcriptional activities in the host plant.

CRediT authorship contribution statement

A.F., H.K., and K.M. designed the research; A.F., Y.N., K.D., H.C., U. G., Y.I., S.M., SRU, YT, H.K. performed the experiments; Y.N. and H.K. analyzed the data; A.F., H.K., and K.M. wrote the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by National Science Foundation (IOS-2049642 to H.K.), Noble Research Institute, LLC (K.M.), and in part by BASF Plant Sciences, Germany (K.M.). The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

References

W.Y. Bang, S.W. Kim, A. Ueda, M. Vikram, D.-J. Yun, R.A. Bressan, P.M. Hasegawa, J. D. Bahk, H. Koiwa, Arabidopsis carboxyl-terminal domain phosphatase-like (CPL) isoforms share common catalytic and interaction domains but have distinct in planta functions, Plant Physiol. 142 (2006) 586–594.

P.G. Blommel, B.G. Fox, A combined approach to improving large-scale production of tobacco etch virus protease, Protein Expr. Purif. 55 (2007) 53–68.

J.L. Corden, D.L. Cadena, J.M. Ahearn Jr., M.E. Dahmus, A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II, Proc. Natl. Acad. Sci. USA 82 (1985) 7934–7938.

M.D. Curtis, U. Grossniklaus, A gateway cloning vector set for high-throughput functional analysis of genes in planta, Plant Physiol. 133 (2003) 462–469.

M.E. Dahmus, Reversible phosphorylation of the C-terminal domain of RNA polymerase II, J. Biol. Chem. 271 (1996) 19009–19012.

M.A. Dietrich, J.P. Prenger, T.J. Guilfoyle, Analysis of the genes encoding the largest subunit of RNA polymerase II in *Arabidopsis and soybean*, Plant Mol. Biol. 15 (1990) 207–223.

D. Eick, M. Geyer, The RNA polymerase II carboxy-terminal domain (CTD) code, Chem. Rev. 113 (2013) 8456–8490.

Y. Feng, J.S. Kang, S. Kim, D.J. Yun, S.Y. Lee, J.D. Bahk, H. Koiwa, Arabidopsis SCP1-like small phosphatases differentially dephosphorylate RNA polymerase II C-terminal domain, Biochem. Biophys. Res. Commun. 397 (2010) 355–360.

A. Fukudome, H. Koiwa, Cytokinin-overinduced transcription factors and thalianol cluster genes in CARBOXYL-TERMINAL DOMAIN PHOSPHATASE-LIKE 4-silenced Arabidopsis roots during de novo shoot organogenesis, Plant Signal Behav. 13 (2018), e1513299.

A. Fukudome, E. Aksoy, X. Wu, K. Kumar, I.S. Jeong, K. May, W.K. Russell, H. Koiwa, Arabidopsis CPL4 is an essential C-terminal domain phosphatase that suppresses xenobiotic stress responses, Plant J. 80 (2014) 27–39.

A. Fukudome, D. Sun, Z. Zhang, H. Koiwa, Salt stress and CTD PHOSPHATASE-LIKE 4 mediate the switch between production of small nuclear RNAs and mRNAs, Plant Cell 29 (2017) 3214–3233.

A. Fukudome, J.S. Goldman, S.A. Finlayson, H. Koiwa, Silencing arabidopsis CARBOXYL-TERMINAL DOMAIN PHOSPHATASE-LIKE 4 induces cytokinin-oversensitive de novo shoot organogenesis, Plant J. 94 (2018) 799–812.

M. Hajheidari, C. Koncz, D. Eick, Emerging roles for RNA polymerase II CTD in Arabidopsis, Trends Plant Sci. 18 (2013) 633–643.

S. Hausmann, H. Erdjument-Bromage, S. Shuman, Schizosaccharomyces pombe CTD phosphatase Fcp1: distributive mechanism, minimal CTD substrate, and active site mapping, J. Biol. Chem. 279 (2003) 10892–10900.

S.D. Hinton, Pseudophosphatase MK-STYX: the atypical member of the MAP kinase phosphatases, FEBS J. 287 (2020) 4221–4231.

J.P. Hsin, J.L. Manley, The RNA polymerase II CTD coordinates transcription and RNA processing, Genes Dev. 26 (2012) 2119–2137.

I.S. Jeong, S. Lee, F. Bonkhofer, J. Tolley, A. Fukudome, Y. Nagashima, K. May, S. Rips, S. Y. Lee, P. Gallois, W.K. Russell, H.S. Jung, A. von Schaewen, H. Koiwa, Purification and characterization of Arabidopsis thaliana oligosaccharyltransferase complexes from the native host: a protein super-expression system for structural studies, Plant J. 94 (2018) 131–145.

Y.M. Jin, J. Jung, H. Jeon, S.Y. Won, Y. Feng, J.S. Kang, S.Y. Lee, J.J. Cheong, H. Koiwa, M. Kim, AtCPL5, a novel Ser-2-specific RNA polymerase II C-terminal domain phosphatase, positively regulates ABA and drought responses in Arabidopsis, N. Phytol. 190 (2011) 57–74.

C.E. Kang, S. Lee, D.H. Seo, W. Heo, S.H. Kwon, J. Kim, J. Lee, B.J. Ko, H. Koiwa, W. T. Kim, J.Y. Kim, Comparison of CD20 binding affinities of rituximab produced in

nicotiana benthamiana leaves and arabidopsis thaliana callus, Mol. Biotechnol. 63 (2021) 1016–1029.
C.H. Kang, Y. Feng, M. Vikram, I.S. Jeong, J.R. Lee, J.D. Bahk, D.J. Yun, S.Y. Lee, H. Koiwa, Arabidopsis thaliana PRP40s are RNA polymerase II C-terminal domain-associating proteins, Arch. Biochem. Biophys. 484 (2009) 30–38.
M.S. Kobor, J. Archambault, W. Lester, F.C.P. Holstege, O. Gileadi, D.B. Jansma, E. G. Jennings, F. Kouyoumdjian, A.R. Davidson, R.A. Young, J. Greenblatt, An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in *S. cerevisiae*, Mol. Cell 4 (1999) 55–62.
H. Koiwa, Salt stress and ion homeostasis, in: K. Shinozaki, K. Okada, A. Oka (Eds.), Crosstalk between plant environmental responses and development, Springer-Verlag, Tokyo, 2004, pp. 181–188.
H. Koiwa, Phosphorylation of RNA polymerase II C-terminal domain and plant osmotic-stress responses, in: T.T. Ashwani, K. Rai (Eds.), Abiotic Stress Tolerance In Plants-toward The Improvement Of Global Environment And Food, Springer, Dordrecht, The Netherland, 2006, pp. 47–57.
H. Koiwa, A.W. Barb, L. Xiong, F. Li, M.G. McCully, B.-h Lee, I. Sokolchik, J. Zhu, Z. Gong, M. Reddy, A. Sharkhuu, Y. Manabe, S. Yokoi, J.-K. Zhu, R.A. Bressan, P. M. Hasegawa, C-terminal domain phosphatase-like family members (AtCPLs) differentially regulate *Arabidopsis thaliana* abiotic stress signaling, growth, and development, Proc. Natl. Acad. Sci. USA 99 (2002) 10893–10898.
H. Koiwa, S. Hausmann, W.Y. Bang, A. Ueda, N. Kondo, A. Hiraguri, T. Fukuhara, J. D. Bahk, D.J. Yun, R.A. Bressan, P.M. Hasegawa, S. Shuman, Arabidopsis C-terminal domain phosphatase-like 1 and 2 are essential Ser-5-specific C-terminal domain phosphatases, Proc. Natl. Acad. Sci. USA 101 (2004) 14539–14544.
F. Li, C. Cheng, F. Cui, M.V.Vd Oliveira, X. Yu, X. Meng, A.C. Intorne, K. Babilonia, M. Li, B. Li, S. Chen, Xf Ma, S. Xiao, Y. Zheng, Z. Fei, R. Metz, C.D. Johnson, H. Koiwa, W. Sun, Z. Li, G.Ad.S. Filho, L. Shan, P. He, Modulation of RNA polymerase II phosphorylation downstream of pathogen perception orchestrates plant immunity, Cell Host Microbe 16 (2014) 748–758.
S.S. Mandal, H. Cho, S. Kim, K. Cabane, D. Reinberg, FCP1, a phosphatase specific for the heptapeptide repeat of the largest subunit of RNA polymerase II, stimulates transcription elongation, Mol. Cell Biol. 22 (2002) 7543–7552.
C. Nawrath, J. Schell, C. Koncz, Homologous domains of the largest subunit of eucaryotic RNA polymerase II are conserved in plants, Mol. Gen. Genet 223 (1990) 65–75.
Z. Ni, C. Xu, X. Guo, G.O. Hunter, O.V. Kuznetsova, W. Tempel, E. Marcon, G. Zhong, H. Guo, W.W. Kuo, J. Li, P. Young, J.B. Olsen, C. Wan, P. Loppnau, M. El Bakkouri, G.A. Senisterra, H. He, H. Huang, S.S. Sidhu, A. Emili, S. Murphy, A.L. Mosley, C. H. Arrowsmith, J. Min, J.F. Greenblatt, RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation, Nat. Struct. Mol. Biol. 21 (2014) 686–695.
V. Reiterer, K. Pawlowski, H. Farhan, STYX: a versatile pseudophosphatase, Biochem. Soc. Trans. 45 (2017) 449–456.
E. Rodriguez-Cazorla, S. Ortuno-Miquel, H. Candela, L.J. Bailey-Steinitz, M.F. Yanofsky, A. Martinez-Laborda, J.J. Ripoll, A. Vera, Ovule identity mediated by pre-mRNA processing in Arabidopsis, PLoS Genet 14 (2018), e1007182.
L. Shen, Y. Zhang, N. Sawettalake, A molecular switch for FLOWERING LOCUS C activation determines flowering time in Arabidopsis, Plant Cell 34 (2021) 818–833.
M.C. Thaller, S. Schippa, G.M. Rossolini, Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily, Protein Sci. 7 (1998) 1651–1656.
L.F. Thatcher, R. Foley, H.J. Casarotto, L.L. Gao, L.G. Kamphuis, S. Melser, K.B. Singh, The arabidopsis RNA polymerase II Carboxyl Terminal Domain (CTD) phosphatase-like1 (CPL1) is a biotic stress susceptibility gene, Sci. Rep. 8 (2018) 13454.
B.L. Timney, B. Raveh, R. Mironska, J.M. Trivedi, S.J. Kim, D. Russel, S.R. Wente, A. Sali, M.P. Rout, Simple rules for passive diffusion through the nuclear pore complex, J. Cell Biol. 215 (2016) 57–76.
J. Van Leene, E. Witters, D. Inze, G. De Jaeger, Boosting tandem affinity purification of plant protein complexes, Trends Plant Sci. 13 (2008) 517–520.
W. Wang, H.S. Cho, R. Kim, J. Jancarik, H. Yokota, H.H. Nguyen, I.V. Grigoriev, D. E. Wemmer, S.H. Kim, Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic "snapshots" of intermediate states, J. Mol. Biol. 319 (2002) 421–431.
J.W. Werner-Allen, C.J. Lee, P. Liu, N.I. Nicely, S. Wang, A.L. Greenleaf, P. Zhou, cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72, J. Biol. Chem. 286 (2011) 5717–5726.
W.A. Zehring, J.M. Lee, J.R. Weeks, R.S. Jokerst, A.L. Greenleaf, The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro, Proc. Natl. Acad. Sci. USA 85 (1988) 3698–3702.
Z.L. Zheng, Cyclin-dependent kinases and CTD phosphatases in cell cycle transcriptional control: conservation across eukaryotic kingdoms and uniqueness to plants, Cells 11 (2022), https://doi.org/10.3390/cells11020279.