FEBS Letters 582 (2008) 4163–4168

A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity

Takahiro Nakamura[a,b], Mamoru Sugita[a,*]

[a] Center for Gene Research, Nagoya University, Chikusa, Aichi, Nagoya 464-8602, Japan
[b] Graduate School of Agriculture, Kyushu University, Fukuoka 812-8581, Japan

Received 9 September 2008; revised 31 October 2008; accepted 14 November 2008

Available online 28 November 2008

Edited by Michael R. Sussman

Abstract **Many plant pentatricopeptide repeat (PPR) proteins are known to contain a highly conserved C-terminal DYW domain whose function is unknown. Recently, the DYW domain has been proposed to play a role in RNA editing in plant organelles. To address this possibility, we prepared recombinant DYW proteins and tested their cytidine deaminase activity. However, we could not detect any activity in the assays we used. Instead, we found that the recombinant DYW domains possessed endoribonuclease activity and cleaved before adenosine residues in the RNA molecule. Some DYW-containing PPR proteins may catalyze site-specific cleavage of target RNA species.**

Keywords: PPR protein; DYW domain; RNA editing; Endoribonuclease; RNase

1. Introduction

An unusually large gene family encoding the pentatricopeptide repeat (PPR) proteins exists in land plants, compared with a small gene family in animals and yeast [1,2]. For instance, *Arabidopsis thaliana*, *Oryza sativa*, and the moss *Physcomitrella patens* contain 450, 477, and 103 genes, respectively encoding PPR proteins [3]. Most PPR proteins are predicted to localize in mitochondria or plastids, and have been implicated in the control of organelle gene expression [2] since several PPR proteins were demonstrated to be involved in site-specific cleavage, splicing, or RNA editing for targeted organellar transcripts [4–6]. PPR motifs probably act as sequence-specific RNA binding proteins, and may recruit some catalytic factors for RNA cleavage, splicing or RNA editing, etc.

The PPR proteins are structurally divided into P and PLS subfamilies. The P subfamily is composed of only canonical PPR (P) motifs and the PLS subfamily consists of repeated units of the classic PPR (P) motif, and longer (L) and shorter (S) non-canonical PPR motifs [2,3]. The P and PLS-type PPR proteins occupy approximately half of all PPR proteins in *A. thaliana* and *O. sativa*. The remaining ones are PLS proteins with a highly conserved C-terminal domain, 87 of which in *A. thaliana* contain the DYW domain. This domain was named because of its characteristic terminal tripeptide (Asp-Tyr-Trp) and has not been found in any other proteins or in any organisms apart from land plants. The DYW domain contains a conserved region composed of an αβα secondary structure, which includes invariant residues that match the active site of cytidine deaminases from bacteria, plant, animals, and yeast [7]. Cytidine deaminases are zinc-dependent enzymes containing a motif corresponding to the active site, C/HxExx...xPCxxC [8]. The DYW domain contains the CxxCH motif of the cytochrome *c* family heme-binding signature [9], which suggests that it may have a certain catalytic activity. However, the actual function of these motifs is completely unknown.

Neither RNA editing nor DYW domains could be identified in algae or the marchantiid liverworts. There is an intriguing correlation between the presence of nuclear DYW genes and organelle RNA editing among embryophytes. These observations provide a hypothesis that the DYW domains are responsible for RNA editing in plant organelles and catalyze RNA editing [7,10]. Therefore, we investigated the function of the DYW domain. Here, we show that this domain can act as an endoribonuclease.

2. Materials and methods

2.1. Transient expression of At2g02980-green fluorescent protein (GFP) fusion protein in tobacco protoplasts

A DNA fragment (1809 base pair (bp)) encoding the full-length At2g02980 protein was amplified from *A. thaliana* ecotype Columbia-0 genomic DNA by polymerase chain reaction (PCR) using primers 2g02980(sal)-F and 2g02980F(nco)-R (see Supplemental Table 1). The amplified DNA fragment was digested by *Sal*I-*Nco*I and ligated into *Sal*I-*Nco*I digested CaMV35S-sGFP (S65T)-nos3′ [11]. The resultant plasmid p02980-GFP was introduced into *Nicotiana tabacum* Bright Yellow 4 leaf cell protoplasts as described previously [12].

2.2. Production of recombinant proteins

The DNA sequence encoding a 109 amino acid (aa) C-terminal DYW domain of At2g02980 was PCR-amplified from *A. thaliana* genomic DNA using primers 02980DYW-F(eco) and 02980-R(sal). The sequence encoding tobacco cp28 protein was PCR-amplified from plasmid pNS28 [13] using primers 28N-F(eco) and 28N-R(sal). The DNA sequence encoding a 108 aa C-terminal region containing the DYW domain of Os05g30710 was amplified from *O. sativa* genomic DNA using primers os05g30710-F and os05g30710-R. The DNA fragment for T-28DYW was prepared by combining by PCR (with primers 28N-F(eco) and 02980-R(sal)), the two partial fragments obtained by amplification using either primers (28N-F(eco) and 28DYW-R) and pNS28 [13], or primers (28DYW-F and 02980-R(sal)) and *A. thaliana*

*Corresponding author. Fax: +81 52 789 3080.
E-mail address: sugita@gene.nagoya-u.ac.jp (M. Sugita).

Abbreviations: aa, amino acid(s); bp, base pair(s); nt, nucleotide(s); GFP, green fluorescent protein; PCR, polymerase chain reaction; PPR, pentatricopeptide repeat; Trx, thioredoxin; RNase, ribonuclease

0014-5793/$34.00
doi:10.1016/j.febslet.2008.11.017

genomic DNA, as described previously [14]. The DNA fragment for T-DYW_M was generated likewise by two successive PCR reactions, using primers 02980DYW-F(eco), 02980M2-R, 02980M2-F and 02980-R(sal) and *A. thaliana* genomic DNA.

The PCR products were inserted in-frame into pBAD/Thio-TOPO (Invitrogen) and the plasmids pT-DYW, pT-osDYW, pT-28, and pT-28DYW were obtained. The recombinant proteins were expressed in *Escherichia coli* LMG194 as a fusion with thioredoxin at its N-terminus and six histidine residues at the C-terminus, and purified by binding to Probond N resin (Invitrogen). They were dialyzed against a solution (20 mM HEPES-NaOH, pH 7.9, 60 mM KCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 3 mM dithiothreitol (DTT), and 17% glycerol).

2.3. Preparation of RNA and DNA probes

A 538 bp 5′ translated region of the plastid *ndhB* gene was amplified from *A. thaliana* genomic DNA (10 ng) using primers ndhB-F and ndhB-R. ndhB-F contains a promoter sequence for T7 RNA polymerase and ndhB-R has a sequence forming a stem-loop sequence to prevent attack by $3' \rightarrow 5'$ exonucleases at the 3′ terminus of RNA. The PCR product (named NB500 DNA) was *in vitro* transcribed by T7 RNA polymerase to produce a 514 nucleotide (nt) of [α-^{32}P]-CTP or [α-^{32}P]-UTP labeled RNA (named NB500 RNA). The ^{32}P-labeled NB500 DNA was obtained by PCR amplification using [α-^{32}P]-dCTP and the same primer set. Single-stranded (ss) NB500 DNAs were obtained by heat denaturing ^{32}P-labeled double-stranded (ds) NB500 DNA at 90 °C for 3 min and subsequent cooling at 4 °C. ^{32}P-labeled RNA or DNA was gel-purified as described [15].

2.4. Assay of cytidine deaminase activity

[^{32}P-CTP]-labeled NB500 RNA (1 fmol, 0.05 nM, 10 000 cpm) was incubated for 30 min at 25 °C with the recombinant protein in 10 mM Tris–HCl, pH 7.9, 30 mM KCl, 6 mM $MgCl_2$, 2 mM ATP, 1 mM $ZnCl_2$ 2 mM DTT, and 8% glycerol [16]. Then, the RNA was extracted and digested at 37 °C for 3 h into 5′ mononucleotides by 1 μg of nuclease P1 and 120 units of S1 nuclease (Takara) in the presence of 50 mM ammonium acetate (pH 4.8). The resultant mononucleotides were separated on a cellulose thin-layer chromatography plate using isopropanol/hydrochloride/water (70:15:15). The separated ^{32}P-mononucleotides were visualized by autoradiography.

2.5. RNA cleavage assay

Internal [^{32}P-UTP]-labeled NB500 RNA (1 fmol, 0.05 nM, 10 000 cpm) was denatured by heating to 70 °C for 2 min, and transferred to 25 °C. Then, the ^{32}P-labeled RNA and the recombinant protein were incubated for 30 min at 25 °C in 20 μl of reaction mixture (10 mM Tris–HCl, pH 7.9, 30 mM KCl, 6 mM $MgCl_2$, 25 mM EDTA, 2 mM DTT, 8% glycerol). After incubation, the ^{32}P-RNA was extracted by phenol/chloroform followed by ethanol precipitation, and then analyzed on a 6% polyacrylamide gel containing 6 M urea.

2.6. Primer extension analysis for identification of 5′ ends of RNA fragments

To determine the cleavage site by T-DYW protein, the cleavage reactions were carried out using 1 μg T-DYW protein and non-radiolabeled NB500 RNA (10 fmol) for 15 or 30 min at 25 °C. The resultant RNA was subjected to primer extension analysis [17], using the ^{32}P-labeled primer barrier-R (5′-CCCATAGGGATTTAGGTGACACTC-3′). The sequence ladder was obtained using NB500 DNA, the same primer, and the thermo-sequencing kit (GE healthcare).

3. Results

3.1. Characterization of DYW domain-containing PPR protein (At2g02980)

In this study we selected several DYW domain-containing PPR proteins for analysis. Among them, we chose the *Arabidopsis* PPR protein At2g02980, which consists of 603 aa residues with 10 PPR motifs and C-terminal E, E+ and DYW domains, because a transferred DNA (T-DNA) insertion mutant line Salk_008983 [18] displayed a severe dwarf phenotype (Supplemental Fig. 1). We next analyzed whether At2g02980 protein is localized in either plastids or mitochondria, or both. Transient expression assays in tobacco leaf cell protoplasts showed that At2g02980 is a mitochondrial protein (Supplemental Fig. 2). This suggested that disruption of the *At2g02980* gene impaired mitochondrial function (probably respiration), and resulted in retarded development of plant seedlings.

To identify the target transcripts for the At2g02980 PPR protein, we performed northern blot analysis of total cellular RNA from the wild-type and T-DNA tag line plants using 23 mitochondrial protein gene-specific probes. However, we could not detect any aberrant transcripts in the T-DNA tag line (data not shown).

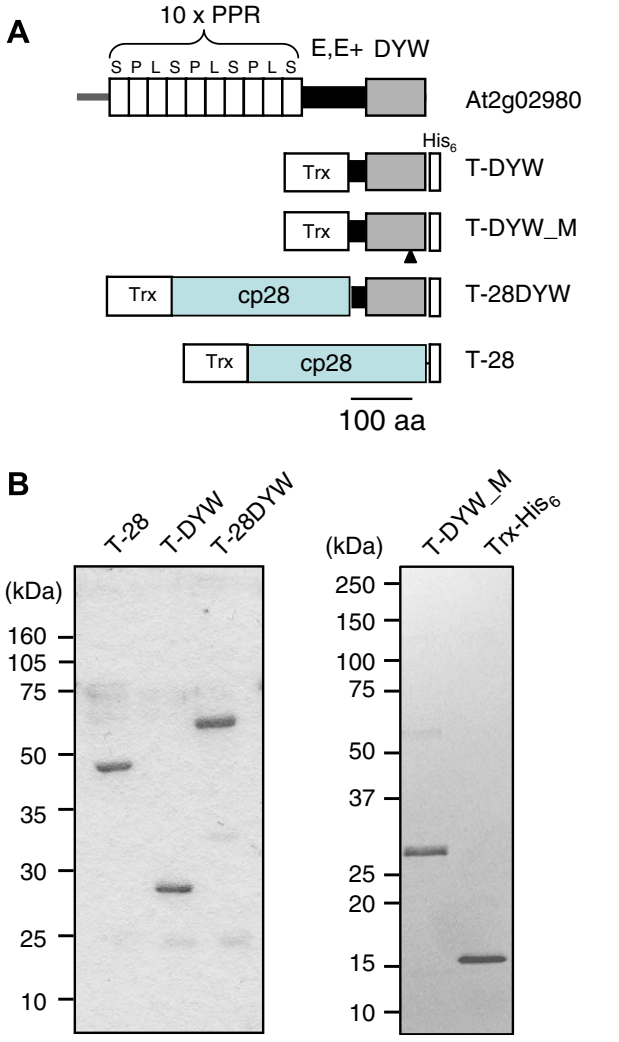

Fig. 1. Structure and expression of the recombinant proteins. (A) Predicted domain structure of At2g02980 protein and various recombinant proteins. PPR motifs are indicated as canonical PPR (P), PPR-like S, and L according to [2]. The C-terminal E, E+ motifs and DYW domain are indicated by a filled box and a gray box, respectively. The position of the mutated cysteine residues in putative heme-binding signature of DYW domain is shown by a filled triangle. The recombinant proteins were expressed as fused protein with thioredoxin (Trx) at the N-terminus and His_6 tag at the C-terminus. (B) The recombinant proteins were separated by SDS–PAGE and stained with Coomassie Brilliant Blue.