RNA-binding properties of HCF152, an *Arabidopsis* PPR protein involved in the processing of chloroplast RNA

Takahiro Nakamura[1], Karin Meierhoff[2], Peter Westhoff[2] and Gadi Schuster[1]

[1]*Department of Biology, Technion – Israel Institute of Technology, Haifa, Israel;* [2]*Heinrich-Heine-Universitat, Institut für Entwicklungs und Molekularbiologie der Pflanzen, Universitatstrasse 1, Düsseldorf, Germany*

The nonphotosynthetic mutant of *Arabidopsis hcf152* is impaired in the processing of the chloroplast polycistronic transcript, *psbB-psbT-psbH-petB-petD*, resulting in non-production of the essential photosynthetic cytochrome b_6f complex. The nucleus-encoded *HCF152* gene was identified to encode a pentatricopeptide repeat (PPR) protein composed primarily of 12 PPR motifs, similar to other proteins of this family that were identified in mutants defected in chloroplast gene expression. To understand the molecular mechanism of how HCF152 modulates chloroplast gene expression, the molecular and biochemical properties should be revealed. To this end, HCF152 and several truncated versions were produced in bacteria and analyzed for RNA-binding and protein–protein interaction. It was found that two HCF152 polypeptides bind to form a homodimer, and that this binding is impaired by a single amino acid substitute near the carboxyl terminus, replacing leucine with proline. Recombinant HCF152 bound with higher affinity RNA molecules, resembling the *petB* exon–intron junctions, as well as several other molecules. The highest affinity was found to RNA composed of the poly(A) sequence. When truncated proteins composed of different numbers of PPR motifs were analyzed for RNA-binding, it was found that two PPR motifs were required for RNA-binding, but had very low affinity. The affinity to RNA increased significantly when proteins composed of more PPR motifs were analyzed, displaying the highest affinity with the full-length protein composed of 12 PPR motifs. Together, our data characterized the nuclear-encoded HCF152 to be a chloroplast RNA-binding protein that may be involved in the processing or stabilization of the *petB* transcript by binding to the exon–intron junctions.

Keywords: RNA processing; nucleus-encoded factor; pentatricopeptide; PPR-motif; *Arabidopsis*.

Chloroplast genes are often transcribed in polycistronic units. Following transcription, the precursor RNA undergoes a variety of maturation events, including *cis-* and *trans-*splicing, cleavage, processing of 5′- and 3′-end termini, and editing. In response to development, light stimuli and environmental cues, the modulation of gene expression is controlled in multiple steps including transcription, splicing, RNA stability and translation [1–5]. Nuclear-encoded (but chloroplast located) proteins possibly involved in chloroplast RNA processing and translation were identified while analyzing mutants having impaired expression of certain genes required for photosynthesis [2,4,6–12]. Such mutants were identified mainly in *Chlamydomonas*, maize and *Arabidopsis*. These studies revealed a complex regulation of gene expression, that is coordinated and involves a large number of proteins [2,13]. For example, about 14 nuclear-encoded loci were identified as being involved in the *trans*-splicing of the *psaA* transcript in the *Chlamydomonas* chloroplast [14].

The nuclear-encoded proteins identified so far that are involved in chloroplast gene expression can be divided into two groups: the first group includes proteins displaying amino acid sequence homology to enzymes involved in RNA maturation processes (such as peptidyl-tRNA hydrolase, pyridoxamine 5′-phosphate oxidase and pseudo-uridine synthase [7,8,15]); the second group is characterized by two similar repeated motifs of several dozen amino acids. The first motif is the tetratricopeptide (TPR) motif composed of about 34 nucleotides present 1–19 times in the proteins identified so far [9,16–18]. The second motif is the pentatricopeptide (PPR) motif that is similar yet distinguished from the TPR motif, and has been defined using a bioinformatics approach [19,20]. Proteins of the PPR motifs were identified upon analyzing RNA- and DNA-binding proteins, proteins that are involved in male sterility, and mutants impaired in RNA maturation [20–29]. The *Arabidopsis* genome contains more than 400 PPR proteins of this family in comparison to those of yeast and humans that contain only a few [30]. Indeed, most of the PPR proteins in *Arabidopsis* are believed to be imported into the chloroplast and mitochondria, taking part in the gene expression processes of these organelles. Recently, while this manuscript was under review, an additional group of chloroplast group II intron splicing factors has being reported [31]. These proteins were characterized by their similar repeated

domains, termed CRM (chloroplast RNA splicing and ribosome maturation), that were proposed to be derived from an ancient RNA-binding module [31].

The *Arabidopsis* high chlorophyll fluorescent mutant, *hcf152,* is nonphotosynthetic, characterized as being impaired in the processing and accumulation of the *psbB-psbT-psbH-petB-petD* cotranscriptional unit that encodes subunits of the photosystem II and cytochrome b_6f complexes. A more detailed analysis revealed that the processing of the *petB* transcript or the stabilization of the spliced transcript is impaired in the absence of the HCF152 protein [32]. The nucleus-encoded *HCF152* gene encodes a chloroplast located protein composed primarily of 12 PPR motifs [32]. In addition to the *hcf152-1* mutant, in which the gene was not expressed, an ethyl methylsulfonate (EMS)-induced mutant (*hcf152-2*), in which a single amino acid substitution was observed, showed a similar but less pronounced phenotype [32]. In previous work, we showed that HCF152 is not associated in a high molecular mass complex and that it is an RNA-binding protein displaying high binding affinities to synthetic RNA molecules representing the *petB* intron–exon junctions [32]. Here, in order to better characterize the RNA-binding properties and the possible protein–protein interactions of HCF152, we produced the protein and several truncated versions in bacteria, and analyzed its protein–protein interactions and RNA-binding properties. HCF152 was found to form a homodimer that is impaired in the *hcf152-2* mutant in which one amino-acid at the C-terminus was substituted. The affinity of the protein to RNA is significantly dependent on the number and nature of the PPR motifs.

Materials and methods

Production of recombinant HCF152 and its truncated versions

The expression and purification of the mature full-length protein in bacteria was performed as described previously [32]. The truncated HCF152 proteins were prepared according to the same procedure using the primers indicated in Table 1 and in the supplementary material. In the case of T152-P2, P1a and P1b, the protein was further purified using a Mono Q column.

Table 1. The RNA probes and HCF152 truncated proteins used in this work. R, reverse; F, forward.

Name	From – to[a]	Length (mer)	Region	Primer (F)	Primer (R)
RNA probes					
BDa	72 431–72 728	297	*psbB* coding	BD001F	BD002R
BDb	74 106–74 373	267	*psbT* coding & UTR	BD003F	BD004R
BDc	74 357–74 700	343	UTR & *psbH* coding	BD004F	BD005R
BDd	74 683–75 110	427	*psbH* coding, UTR *petB* 5′exon & intron	BD005F	BD006R
BDe	75 093–75 478	385	*petB* intron	BD006F	BD007R
BDf	75 461–75 808	347	*petB* intron & 3′exon	BD007F	BD008R
BDg	75 289–76 286	497	*petB* exon	BD008F	BD009R
BDh	77 325–77 628	303	*petD* 3′exon	BD010F	BD011R
BDi	76 266–76 668	402	*petB* coding, UTR *petD* 5′exon & intron	BD009F	BD012R
BDj	76 650–77 008	358	*petD* intron	BD012F	BD013R
BDk	76 986–77 325	340	*petD* intron & 3′exon	BD013F	BD010R
BA	961–1323	389	*psbA* coding	BA001F	BA002R
16S	101 094–101 521	427	16S rRNA	16S001F	16S002R
Dd350	74 683–75 032	350	*psbH* coding, UTR *petB* 5′exon & intron	BD005F	BDd350R
Dd225	74 683–74 907	225	*psbH* coding, UTR *petB* 5′exon & intron	BD005F	BDd225R
Dd120	74 683–74 802	120	*psbH* coding & UTR	BD005F	BDd120R
Df260	75 461–75 720	260	*petB* intron & 3′exon	BD007F	Df260-R
Df166	75 461–75 626	166	*petB* inton	BD007F	Df166-R
Df82	75 461–75 542	82	*petB* intron	BD007F	Df82-R
D1	74 703–74 802	100	UTR	da001F	BDd120R
D2	74 722–74 802	81	UTR	da002F	BDd120R
D3	74 742–74 802	61	UTR	da003F	BDd120R
D4	74 683–74 782	100	*psbH* coding & UTR	BD005F	da005R
D5	74 683–74 762	79	*psbH* coding & UTR	BD005F	da004R
D6	74 683–74 742	60	UTR	BD005F	da006R
D7	74 765–74 822	58	UTR	da008F	da009R
HCF152 truncated proteins					
TF152-F	713[b]	12[c]	–	EX-H	EX-R
T152-CH	465[b]	8[c]	–	EX-HG	EX-R
T152-NH	254[b]	4[c]	–	EX-H	H152-N(c)
T152-P2	70[b]	2[c]	–	152B-F(eco)2	152 A-R
T152-P1a	35[b]	1[c]	–	152 A-F1(eco)2	152D-R
T152-P1b	35[b]	1[c]	–	152C-F(eco)2	152 A-R

[a] Nucleotide range. The numbering as in [44]; [b] length (amino acids); [c] number of PPR motifs.

Size-exclusion chromatography

Size-exclusion chromatography was performed by applying the purified recombinant HCF152 onto a Superdex 200 column in buffer E (20 mM Hepes pH 7.9, 12.5 mM MgCl$_2$, 60 mM KCl, 0.1 mM EDTA, 2 mM dithiothreitol and 17% glycerol) at a flow rate of 0.5 mL·min^{-1}. Proteins were precipitated by cold acetone and analyzed by SDS/PAGE. For digestion of the RNA, the extract was incubated with RNase A (1 mg·mL^{-1}) and 3 U·μL^{-1} of RNase T1 at 37 °C for 1 h. The Superdex 200 column was calibrated with the following protein standards: thyroglobolin, 669 kDa; catalase, 232 kDa; aldolase, 158 kDa; bovine serum albumin, 67 kDa and casein, 30 kDa.

Analyzing the protein–protein interaction of HCF152

The HCF152, HCF152-2 and luciferase were synthesized *in vitro* and ^{35}S-labeled using the TNT T3 coupled reticulocyte lysate system with the plasmid constructs containing, HCF152 (pcAT152), HCF152-2 (pcAT152/119) and luciferase (luciferase control T3 DNA, Promega), respectively. The labeled protein was mixed with a His$_6$-fused recombinant protein [Trx (thioredoxin), T152-F, T152-NH or T152-CH] in a binding buffer (50 mM Tris/HCl, pH 8.0, 2 mM imidazole, 0.1% Tween 20, 2 mM dithiothreitol and 10 mM MgCl$_2$) containing 100 mM NaCl for 20 min at room temperature. Ni-nitrilotriacetic acid agarose resin (50 μL) was then added, and incubation continued for an additional 30 min. The resin was washed five times with a binding buffer containing 100 mM NaCl, and the ^{35}S-labeled proteins that bound the resin via the His$_6$-fused proteins were eluted with the binding buffer containing 500 mM NaCl. The bound proteins were then analyzed by SDS/PAGE and autoradiography.

Preparation of RNA probes

Certain fragments of *Arabidopsis* chloroplast DNA (Table 1) were PCR amplified using the appropriate primers. They were used as templates for the transcription of the corresponding RNA by the T7 RNA polymerase primed by the T7 promoter sequence (AATACGACTC ACTATAG) attached to the 5′-end of the forward primer. The PCR product was purified from gel using a QIAquick gel extraction kit (Qiagen), and the radiolabeled RNA probe was transcribed as described previously [33]. For the production of nonradioactive RNA, the transcription reaction mixture included 5 mM of each nucleotide.

UV-crosslinking

UV-crosslinking of the protein to radiolabeled RNA was carried out as described previously [34]. The proteins (1 pmol) were incubated with [^{32}P]RNA (25 fmol) in buffer containing 10 mM Hepes/NaOH (pH 7.9), 30 mM KCl, 6 mM MgCl$_2$, 0.05 mM EDTA, 2 mM dithiothreitol, 8% glycerol, 0.0067% of Triton X-100 and 67 μg·mL^{-1} of yeast tRNA (Sigma) for 15 min. The protein and RNA were crosslinked by 1.8 J of UV irradiation in a UV-crosslinker (Hoefer Inc.) following digestion of the RNA by 10 μg of RNaseA and 30 U of RNase T1 at 37 °C for 1 h,

fractionation by SDS/PAGE and analysis by autoradiography. For the competition assay, the protein was mixed with nonradioactive RNA for 5 min and the radiolabeled RNA was then added. When ribohomopolymers were used as competitors, an average length of 400 nucleotides was used to calculate the molar amount. When ssDNA and dsDNA were used as competitors, the PCR fragment of BDd, described above, was used when denatured (ssDNA; 90 °C for 5 min) or not (dsDNA).

Results

Preparation of recombinant HCF152 and the different fragmented proteins

The molecular analysis of the high chlorophyll fluorescent mutant 152 (*hcf152*) revealed that the processing of the chloroplast *petB* transcript is impaired [32]. The cloning and characterization of the *HCF152* locus revealed that the nucleus-encoded protein contains 12 repetitions of the PPR motifs (Fig. 1A). Another *hcf* mutant generated by chemical EMS treatment revealing a point mutation in the
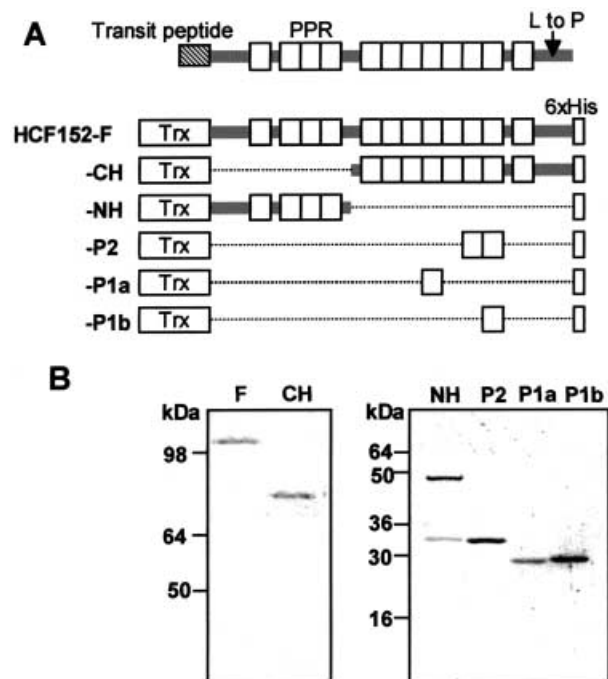

Fig. 1. Protein constructs used in this paper. (A) The HCF152 protein is presented schematically. Open boxes show the 12 PPR motifs and a hashed box the chloroplast transit peptide. An arrow indicates the position of the single amino acid substitution, leucine with proline, found in the mutant HCF152-2. The HCF152 protein, as well as its truncated versions, was expressed in *E. coli* fused to the thioredoxin (Trx) and His$_6$ tag as shown in the figure. (B) Stained gel profile of the recombinant proteins following expression in bacteria and purification by affinity chromatography. The gel on the left contains 8% polyacrylamide used to resolve the high molecular proteins, while that on the right contains 15% to resolve the shorter truncated proteins. The lower 33 kDa band at the HCF152-NH lane is a degradation product that is constituently copurified with the recombinant protein.

same gene resulted in the replacement of leucine with proline. The EMS mutant, termed *hcf152-2*, essentially displayed similar phenotype characteristics of the gene's inactivated *hcf152-1* mutant, albeit to a lesser extent. The single amino acid substitution was located near the C-terminus of the protein but in none of the 12 PPR motifs (Fig. 1A) [32]. As proteins characterized in mutants affected in chloroplast gene expression were found to belong to the PPR motif family of nucleus-encoded genes, and in order to explore the molecular mechanism in which these proteins affect gene expression post-transcriptionally, we decided to analyze the RNA-binding properties and the protein–protein interactions of a member of this family, HCF152. We therefore prepared the mature HCF152 recombinant protein, as well as several fragmented proteins containing different numbers of PPR motifs, while using the bacterial expression system (Fig. 1). As HCF152 is a nuclear-encoded and chloroplast located protein, it contains a transit peptide that is removed upon entering the chloroplast. This part of the protein could be defined using the CHLOROP program [35]. In this work, the recombinant proteins were produced without the predicted transit peptide in order to resemble the mature protein in the chloroplast. Most of the available bacterial expression systems that were examined produced insoluble proteins. Finally, the production of the protein at 16 °C using the pBAD/Thio-Topo expression system, in which the protein was fused to a 18-kDa thioredoxin and a His$_6$-tagged residue at the N- and C-terminus halves, respectively, was found to be the most efficient way to obtain a significant amount of soluble and active protein, as well as the different truncated forms. All experiments attempting to produce the HCF152-2 protein, that harbors a single mutation as a soluble protein, failed. In addition to the full-length protein, the N-terminus half (HCF152-NH) and the C-terminus half (HCF152-CH) of the protein containing four and eight PPR motifs, respectively, were also produced (Fig. 1). To characterize further the PPR motif, proteins containing both one and two PPR motifs were also produced (Fig. 1).

The recombinant proteins were purified and analyzed by SDS/PAGE in order to determine the correct molecular mass (Fig. 1B). In addition, all proteins were verified by immunoblotting using antibodies against the His$_6$ tag (not shown).

HCF152 forms a homodimer of about 180 kDa

Several regulatory proteins described previously to be involved in chloroplast gene expression were found to be associated in high molecular mass complexes of about 300–1700 kDa [8,15,17,18,20,36,37]. However, when chloroplast soluble proteins were fractionated through a size-exclusion column, and the presence of HCF152 was detected with specific antibodies, HCF152 was eluted in one peak at about 180 kDa and was not associated in a high molecular weight complex [32]. As the molecular mass of a mature HCF152 is 85 kDa, this result could be obtained by three possibilities. First, the protein is not associated in a complex and is fractionated at this molecular mass. Second, HCF152 is associated with other chloroplast proteins in a complex and third, HCF152 forms a homodimer. As it was found that
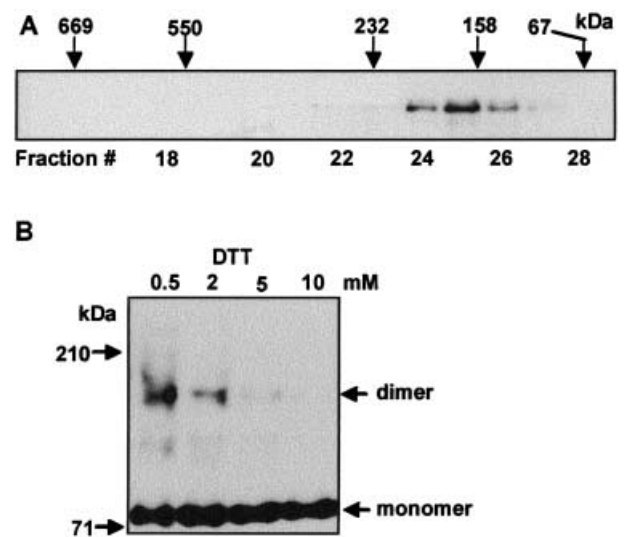

Fig. 2. HCF152 forms a homodimer. (A) The purified recombinant HCF152 was fractionated by a Superdex 200 size-exclusion column. The elution profile of several molecular mass markers is indicated on the top. Extensive treatment of the fractions loaded with ribonucleases did not change the elution profile. (B) Purified recombinant HCF152-F was incubated with the indicated amount of dithiothreitol (DTT) followed by fractionation on nondenaturing SDS/polyacrylamide gel. The migration of markers of known molecular masses is indicated on the left.

two molecules of HCF152 could interact together (see below), the homodimer option seemed feasible. In order to analyze this possibility, purified recombinant HCF152 (HCF152-F) was fractionated on the same column. Figure 2A shows that the purified recombinant protein eluted at a molecular mass of about 180 kDa. As no protein other than recombinant HCF152 was loaded on the column, we concluded that HCF152 either forms a homodimer of about 180 kDa or that it is a monomer eluting from the column at this position. In order to distinguish between these possibilities, the purified recombinant protein was fractionated by nondenaturing PAGE. It was found that part of the protein population migrated at 180–200 kDa, while increasing the dithiothreitol concentration from 0.5 to 10 mM resulted in migration of the entire HCF152 population at 80–90 kDa (Fig. 2B). Taken together, these results suggested that HCF152 forms a homodimer and is not associated with other proteins.

In order to further characterize the homodimer formation of the HCF152 protein, we analyzed the one amino acid substitution mutant HCF152-2, and the C- and N-terminus halves of the protein. We had to use the *in vitro* translation system as HCF152-2 could not be produced in bacteria in a soluble form. First, the HCF152, HCF152-2 and luciferase (as a negative control) were produced and ^{35}S-labeled by the *in vitro* transcription/translation system. Each protein was then incubated with a recombinant HCF152, HCF152-NH (N-terminus half) and HCF152-CH (C-terminus half), tagged with His$_6$, followed by the addition of Ni-nitrilotri-acetic acid/agarose and precipitation of the bound proteins. The results of this experiment are presented in Fig. 3. The ^{35}S-labeled HCF152 bound the His$_6$ HCF152, producing a
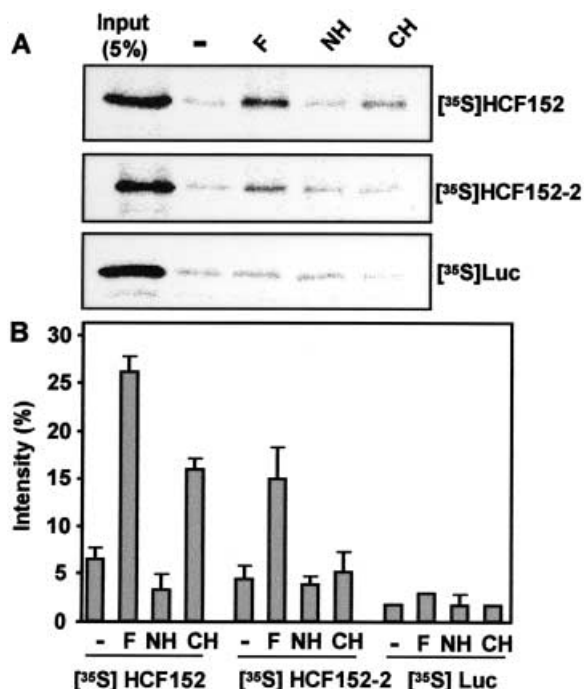
Fig. 3. Dimer formation is impaired in the HCF152-2 mutant. (A) The mature forms of HCF152, HCF152-2 and luciferase (Luc) (as a negative control) were produced and ^{35}S-labeled in an *in vitro* transcription/translation reaction. The ^{35}S-labeled proteins were each incubated with recombinant HCF152-F protein (F), or the N- or C-terminus halves (NH and CH, respectively) and the Ni-nitrilotriacetic acid resin. Following binding and extensive washing, the bound proteins were eluted by a high salt concentration and analyzed by SDS/PAGE and autoradiography. In the lane marked 'input', 5% of the corresponding ^{35}S-labeled protein was analyzed. In the lanes marked '-', the thioredoxin protein was incubated in the assay as a negative control. (B) Intensities of the ^{35}S-labeled protein signals were quantified using a phosphorimager. Radioactivity of the band for each input was designated as 100%. The assay was repeated three times. Bars indicate SEM.

signal five times greater than that of the background. The binding site was found to be located at the C-terminus of the protein although binding of the C-terminus half (CH) is only half of the full-length protein (Fig. 3). Interestingly, the binding efficiency of the full-length mutant HCF152-2 was also approximately half of the full-length HCF152, while the C-terminus half of HCF152 did not bind HCF152-2 at all (Fig. 3). No binding beyond the background level was observed for the luciferase protein that was used as a negative control (Fig. 3). In another control experiment, a His[6] fused thioredoxin did not bind any of the ^{35}S-labeled proteins beyond the background level (Fig. 3, lane marked '-'). These results confirmed the formation of a HCF152 dimer suggested by the size fractionation experiments. The C-terminus part of HCF152 is partially responsible for the intermolecular interaction. Furthermore, it is suggested that the one amino acid substitution in the EMS generated *hcf152-2* mutant produced a protein partially impaired in dimer formation, and therefore this phenomenon is important for the biological activity of HCF152.

RNA-binding characteristics of HCF152

Analyzing the chloroplast transcript pattern of the *hcf152-1* and *hcf152-2* mutants by RNA gel blot revealed differences in the *psbB-psbT-psbH-petB-petD* polycistronic transcriptional unit, and, more specifically in the accumulation of the *petB* intron and processing of the 3′-termini of *psbH* [9,32]. This observation led to the hypothesis that the *HCF152* gene product is required, either directly or indirectly, for the correct 3′-end processing of *psbH* and splicing of *petB* intron, or alternatively, stabilization of the splicing products [32]. As HCF152 was characterized to bind RNA with preference to the *psbH* 3′-end and *petB* intron–exon sequences [32], we first asked whether this protein binds with high affinity other RNA molecules.

An RNA-binding UV-crosslinking experiment was performed analyzing several RNA molecules, as in previous work [32], spanning the *psbB* multicistronic transcript, as well as several other chloroplast transcripts. These included three RNA molecules resembling the *psbA, rbcL* and ribosomal 16S transcripts that do not contain introns and additional molecules of the *petD* transcript. In order to prevent nonspecific binding of HCF152 to RNA, an extra amount of about 330-fold yeast tRNA was included in the reaction mixture. As described in previous work [32], RNA molecules corresponding to the 5′- and 3′-borders of the *petB* intron, and to the corresponding parts of the related exons, were found to bind the recombinant HCF152 (BDd, BDe and BDf in Fig. 4). The UV-crosslinking assay gave very low signals with RNAs corresponding to the sequences of *psbB* and *petD* (BDa, and BDk). In addition, a high UV-crosslinking signal was also obtained with RNA corresponding to the *psbA*, but not with the 16S ribosomal RNA and *rbcL* or an RNA derived from the Bluescript plasmid (Fig. 4B). However, as the sequence of nucleotides differed within the RNA molecules, the lack of a UV-crosslinking signal does not necessarily imply that no binding takes place. In order to verify the binding properties of HCF152, we analyzed the binding of these RNAs using the UV-crosslinking competition method. In this method, only a single RNA is radioactively labeled to provide the UV-crosslinking signal when binding the protein, while extra amounts of the tested RNA molecule are added to compete with the binding of the radioactive RNA. An RNA that efficiently competes for the binding binds HCF152 with high affinity. The IC[50] parameter was defined as the concentration of the competitor RNA that resulted in a 50% reduction in the radioactive UV-crosslinking signal (examples of competition curves are found below) [38]. The lowest IC[50] value is the value of for a specific competitor RNA; the highest is the affinity of this RNA to the protein.

The UV-crosslinking assay was repeated using [^{32}P]BDd RNA, the protein, about 330-fold of yeast tRNA, and the corresponding nonradioactive RNAs in molar excess as indicated in the figures. The results of this assay confirmed our previous result that the RNAs derived from the *psbB* intron–exon junctions bound HCF152 with a relatively high affinity while the other RNAs' multicistronic transcript displayed very low affinities (Fig. 4, Table 2, [32]). In addition, two RNA probes derived from the boundaries of the second intron of this polycistronic transcription unit, that of *petD* (BDi and BDk), displayed
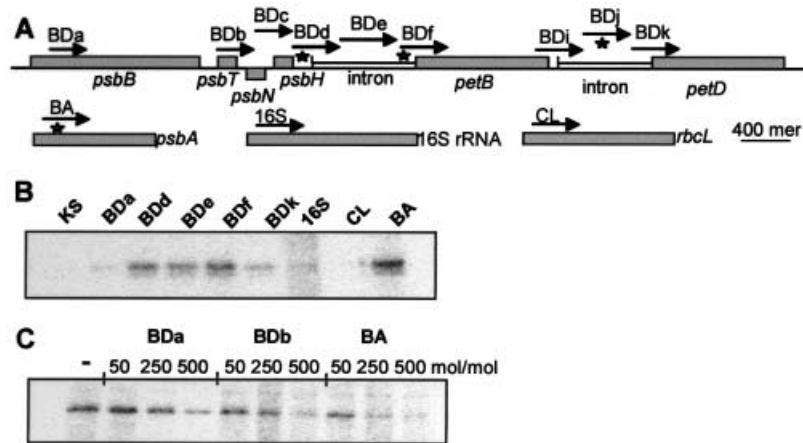
Fig. 4. RNA-binding of HCF152. (A) Schematic representation of the *Arabidopsis psbB-petD* operon, the *psbA*, the 16S rRNA and *rbcL*. The RNA probes for the binding assays are indicated with arrows and letters (Bda–k, BA and 16S). The length of the arrows indicates the length of the probes, and a scale bar for 400 nucleotides is shown. Stars indicate the high affinity binding sites for HCF152. (B) RNA-binding of HCF152-F to several RNAs derived from the *psbB-petD* operon was analyzed by the UV-crosslinking assay. The symbols are the same as in panel A. KS indicates RNA transcribed from the multicloning site of the plasmid pBluescript KS. (C) Competition UV-crosslinking assay. The RNAs indicated on top competed for binding to HCF152-F with the BDd RNA. The experiments were performed with radiolabeled BDd RNA and 50-, 250- or 500-fold molar excess of the nonradioactive RNAs.

Table 2. RNA-binding characteristics of HCF152. Competitive UV-crosslinking experiments were performed, as shown in Fig. 6, with radiolabeled BDd RNA and various *in vitro* synthesized competitor RNA probes (Bda–k, BA and 16S), as well as ribohomopolymers, *Synechocystis* total RNA, yeast tRNA and single- and double-stranded DNA. For each assay, the results were plotted as shown in Fig. 6. The concentration of the competitor that resulted in a 50% inhibition of the signal was defined as IC_{50} and is shown in the Table. Values represent at least three experiments.

RNA probe	IC_{50} (mol·mol^{-1})
BDa	> 250
BDb	163
BDc	210
BDd	38
BDe	175
BDf	50
BDi	95
BDj	44
BDk	86.7
ssDNA	45
dsDNA	> 100
Poly(A)	8
Poly(C)	> 300
Poly(G)	> 300
Poly(U)	53
BA	52
16S	142
Synechocystis RNA (ng)	(> 250)
Yeast tRNA (ng)	(> 250)

a lower affinity than BDd and BDf but a significantly higher one than the low affinity probes (Table 2). Moreover, RNA derived for the *petD* intron (BDj) showed high binding affinity. In addition, RNA derived from the *psbA* gene that does not contain an intron, displayed high binding affinity, while RNA derived from the ribosomal 16S RNA and *rbcL* displayed low binding affinity (Table 2). High binding affinity was obtained to ssDNA composed of the heat-denatured PCR fragment used to transcribe the BDd RNA. However, a very low binding affinity was observed with dsDNA composed of the same PCR fragment but not denatured (Table 2). The ssDNA binding phenomena is a characteristic of many RNA-binding proteins [39,40]. Upon analyzing ribohomopolymers, a high binding affinity to poly(A) was found, and to a lesser extent also to poly(U). Contrary to this, very low affinities were found for poly(C) and poly(G) (Fig. 5, Table 2). Total RNA of the photosynthetic cyanobacteria, believed to be related to the evolutionary ancestor of the chloroplast and yeast tRNA, displayed very low affinity to the recombinant HCF152 (Table 2).

Taken together, these results indicated that HCF152 is an RNA-binding protein binding certain RNA molecules with higher affinity than others. In addition to previously shown molecules resembling the *petB* intron–exon junctions, it also binds RNA molecules resembling the *petD* intron, *psbA* and poly(A). Therefore, in order to better define the RNA-binding site, we carried out a deletion analysis of the BDd RNA, the high affinity binding molecule.

Defining the binding site of HCF152 in the psbH-petB transcript

In order to further characterize the HCF152 binding site, we synthesized a series of deleted RNA probes. Each RNA probe was designed by a subsequent deletion of the BDd and BDf sequences to which the HCF152 was bound with high affinity. When these RNA molecules were analyzed in the competitive UV-crosslinking assay, all were found to bind RNA with high affinity (Fig. 6A,B). In order to define
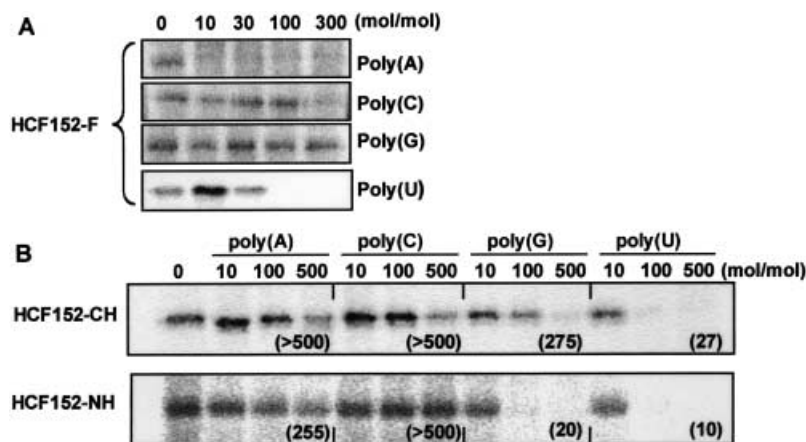
Fig. 5. Binding affinities of the full-length protein as well as the C- and N-terminus halves to ribohomopolymers. Competition of ribohomopolymers for RNA-binding of HCF152-F (full length; panel A), HCF152-CH (C-terminus half) and HCF152-NH (N-terminus half; panel B), each at 0.1 µg, was performed on the [^{32}P]BDd RNA in the UV-crosslinking competition assay. The numbers above the figure indicate the molar excess of the ribohomopolymer that were added to the [^{32}P]BDd RNA in the competition. The numbers in parentheses (B) show the IC$_{50}$ calculated from three independent experiments.

the target sequence better, the Dd120 sequence was subsequently deleted by 20 nucleotides from the 5′- or 3′-ends. When the resulting seven RNA molecules were analyzed in the UV-crosslinking competition assay, five (D1–D5) were found to display high binding affinity and two (D6 and D7) low affinity (Fig. 6C,D). Therefore, the 21 nucleotides that differed between D5 and D6 are the putative target sequence for HCF152 in the *psbH-petB* intergeneic region. In addition, it is possible that a secondary structure involving the interaction between the 21 nucleotides and neighboring sequences is involved in formation of the binding site.

Together, these experiments defined several target sequences for high affinity RNA-binding of HCF152. These included the 21 nucleotides of the UTR between *psbH* and *petB*, the 82 nucleotides of *petB* intron (Df82), part of the *psbA* transcript (BA) and the *petD* intron (indicated by stars in Fig. 4A). Analyzing the secondary structure of these molecules revealed the ability to form a stem-loop structure (though with very short stems) with a single-stranded region of an adenosine-rich sequence (not shown). As HCF152 displayed high binding affinity to poly(A) (Table 2, Fig. 5), the single-stranded region of adenosine stretch could be a putative binding site for HCF152.

Contribution of the multiple PPR motifs for the affinity of HCF152 to RNA

As the major characteristic of HCF152 is the 12 PPR motifs, our next question related to their contribution to the RNA-binding phenomenon. First, the protein was divided into the C- and the N-terminus halves, containing eight and four PPR motifs, respectively (Fig. 1). Each part was analyzed for binding affinities to ribohomopolymers. While the full-length HCF152 bound poly(A) with the highest affinity of all molecules examined in this study, this affinity was drastically reduced in the C- and N-terminus halves of the proteins (Table 2, Fig. 5). On the other hand, while the full-length protein did not bind poly(G), the N-terminus half of

the protein displayed affinity to this ribohomopolymer (Fig. 5B). The situation with poly(U) and poly(C) did not differ significantly between the full-length and parts of the protein. All bound poly(U) with a relatively high affinity and poly(C) with a very low affinity. Taken together, HCF152 as a full-length protein binds poly(A) with the highest affinity, but when divided into parts, each part displays a higher affinity to poly(U) than to poly(A). This experiment implies that the combinations of several PPR motifs, and probably the sequence of certain amino acids inside and perhaps outside the motifs, are responsible for the RNA-binding properties of the full-length protein.

To further characterize the RNA-binding properties of the proteins consisting of four, eight and 12 PPR motifs (HCF152-NH, -CH and -F, respectively), we performed a competitive UV-crosslinking assay using an *Arabidopsis* RNA sequence. As shown in Fig. 7, a 50-fold molar excess of BDd and BDf competitor RNAs, but not BDb, BDc and BDe RNAs, competed the RNA-binding of HCF152-F, the mature proteins containing 12 PPR motifs. However, binding of HCF152-CH, comprised of eight PPR motifs, to BDd RNA was competed efficiently by a 50-fold excess of BDe, BDd, and BDf (Fig. 7). These RNAs competed less efficiently with the binding of HCF152-NH, composed of four PPR motifs (Fig. 7). Therefore, the results of the experiments presented in Figs 5 and 7 showed that the number of PPR motifs in HCF152 is important for the specificity of RNA-binding. In addition, the results obtained so far strongly suggest that the combination of several PPR motifs determines the affinity and specificity of binding to RNA. However, as the PPR motifs differ in their sequence of amino acids, the specific amino acid sequence inside the PPR motifs might contribute significantly to their affinity and specificity. Furthermore, the number of PPR motifs seems to represent a critical parameter determining binding properties.

In order to obtain further details about this question, RNA-binding assays were performed using truncated proteins composed of one or two PPR motifs. As the
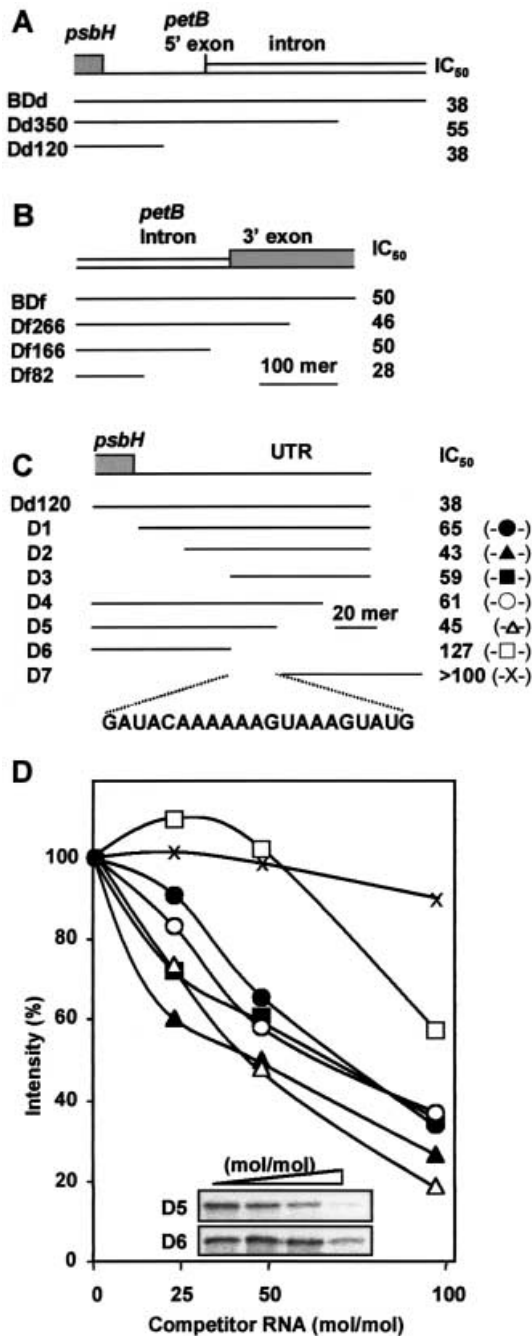
Fig. 6. Defining the HCF152 high affinity binding site. The BDd (A) and BDf (B) RNAs, as well as the truncated molecules of these RNAs that are schematically shown, were analyzed in the UV-crosslinking competitive assays, as shown in D. The IC$_{50}$ of each RNA was determined by plotting the data (D). The location of the high affinity binding site in the Dd120 RNA defined in panel A was further analyzed for binding to the HCF152 by constructing an additional seven truncated versions, as schematically presented (C). These RNAs were assayed in the UV-competitive test (inset of D) and the binding affinities were determined and plotted (D). The competitor RNA concentrations of the competition assay (D, inset) were 0, 25-, 50- and 100-fold excess. ●, D1; ▲, D2; ■, D3; ○, D4; △, D5; □, D6; -×-, D7. The IC$_{50}$ of each RNA is indicated and the nucleotide sequence shows the location of the binding site.
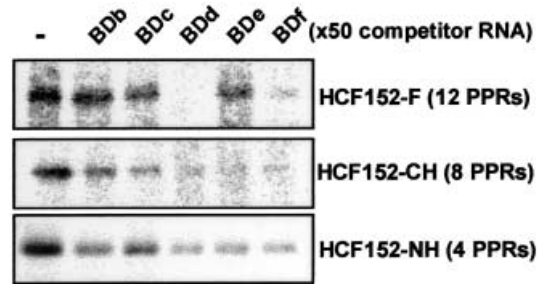

Fig. 7. Binding of the C- and N-terminus halves to different RNA molecules. The affinities of the full-length (HCF152-F), the C-terminus half (HCF152-CH) and the N-terminus half (HCF152-NH), each at 0.1 µg, to different RNAs were defined by a UV-crosslinking competition assay in which 50-fold molar excess of the corresponding RNA was competed with [^{32}P]BDd RNA. In the lane marked '-', no competitor RNA was added. The number of PPR motifs in each protein is indicated.

UV-crosslinking signals for these proteins were very faint due to the low affinity of RNA-binding, the amount of proteins in the reaction mixture was significantly increased (Fig. 8). The results of this experiment showed that the affinity of the truncated protein composed of four PPRs was significantly reduced compared to the full-length protein. Reducing the number of PPR motifs to two resulted in an additional over 10-fold decrease (Fig. 8). Moreover, a UV-crosslinking signal could be obtained with these proteins only when the yeast-tRNA was omitted from the binding assay, indicating a lesser specific binding to RNA. Finally, the two truncated proteins containing a single PPR domain did not show any binding to RNA (Fig. 8). Together, these experiments demonstrated that for HCF152, the PPR motif is indeed an RNA-binding domain, but for the particular domains tested here the binding activity requires at least two PPR motifs and is drastically increased by increasing the number of PPRs to four and 12, respectively. As each PPR is unique and distinct in sequence, it is possible that other PPR domains of this protein, as well as sequences between the PPR motives, display higher affinity than the two tested here. Indeed, recent analysis of LRP130, a human PPR protein located mainly in the mitochondria, revealed RNA-binding activity of truncated proteins composed of only two or even one PPRs [41].

Discussion

HCF152 is a specific RNA-binding protein

The results of this and the previous study [32] clearly show that HCF152 is an RNA-binding protein whose affinity and specificity are dependent upon the number and possibly the amino-acid sequence of the PPR domains. One of the four high affinity RNA-binding targets identified has been narrowed down to 21 nucleotides of the untranslated region between *psbH* and *petB*. The high affinity sequences are characterized by an adenosine stretch placed between sequences potentially forming short double-stranded regions. Indeed, the highest binding affinity of HCF152 to RNA was observed for poly(A). However, the adenosine
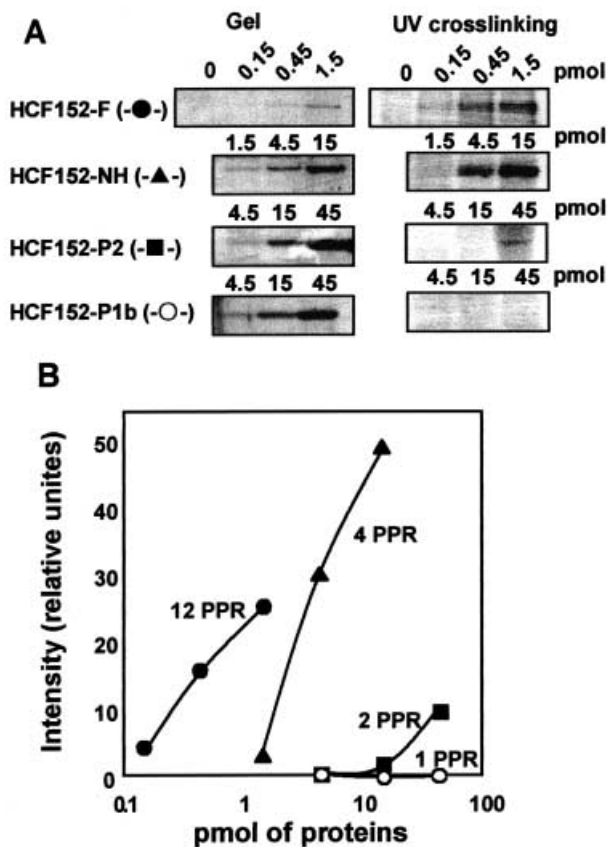
Fig. 8. RNA affinities of truncated proteins containing different numbers of PPR motifs. (A) RNA-binding of the HCF152 and truncated parts containing different amounts of PPR motifs were analyzed in a UV-crosslinking assay to [^{32}P]BDd RNA. Increasing amounts of proteins, as indicated in the figure, were UV-crosslinked to the RNA. A stained polyacrylamide gel is shown on the left and the UV-crosslinking radioactive signal of the same gel on the right. The number of PPR motifs per protein, as illustrated also in Fig. 1, is indicated. (B) The intensities (in relative units) of the UV-crosslinking signals, as shown in (A), were plotted against the amounts of corresponding proteins.

stretch could not solely serve as the target sequence as poly(A) stretches were spread throughout the chloroplast genome and were found easily in most of the chloroplast transcripts as well as in the RNA probes used in this study. In addition, as the results showed that there is no simple nucleotide sequence forming the matrix for the high-affinity binding, it may be suggested, as for most of the specific RNA-binding proteins, that the combination of structural and sequence properties defines the binding site for the HCF152 in the RNA molecule. Similar observations were reported for other PPR proteins, the p67 [22] and the LRP130 [24]. However, a detailed analysis of LRP130 (harboring nine PPR domains) published while this manuscript was in the reviewing process, revealed that unlike HCF152, it displayed high binding affinity to poly(G) and poly(U) but not poly(A) [41]. Additional major difference between HCF152 and LRP130 is that in LRP130 RNA-binding properties similar to the full length protein could be obtained with a truncated part composed only of two PPR

motives [41]. Therefore, because of the differences between the two proteins, the analysis of more proteins and PPR motives is required to define the specificity and affinity of RNA-binding and the interaction with proteins. The target region identified in this study for HCF152 is located downstream ($+36$ to $+56$) of the *psbH* stop codon and upstream (-79 to -99) of the *petB* translation start codon, suggesting that HCF152 is not involved in the translation regulation of the *petB* gene. However, a PPR protein of maize containing 14 PPR motifs that clustered in a very similar manner to HCF152, CRP1, has been proposed to be involved in the translation of the *petD* mRNA in addition to RNA processing [20].

Function of HCF152 in *petB* RNA maturation

The *hcf152* strain phenotype suggests that HCF152 functions in the processing of *petB* by possibly stabilizing the 3′ *psbH* terminus and the splicing products [32]. In the chemically induced EMS mutant *hcf152-2*, in which one amino acid not located in a PPR motif was substituted, a similar yet less significant phenotype was observed. Unlike the *hcf152-1* mutant in which the HCF152 protein is not produced, the protein in the *hcf152-2* seems to be produced and accumulated, albeit with one amino acid changed. Our protein–protein binding experiment suggests that this single amino acid substitution has weakened the dimer formation in comparison to the HCF152 (Fig. 3). This observation suggests that the dimer formation is important for the function of HCF152 in RNA processing, and the inability to form the dimer results in a loss of function. Interestingly, the dimer formation was found to be located at the C-terminus half of the protein, and the single amino acid substitution next to the C-terminus of the protein but not in a PPR motif. Therefore, the question still arises as to whether or not the PPR motif, of which most of the HCF152 is composed, functions in the dimer formation.

The *petB* intron is classified as a group II intron that may be self-spliced *in vitro*. However, the group II introns of higher plant chloroplasts have lost their self-splicing ability when incubated *in vitro*, and auxiliary factors are therefore required for correct splicing. Several auxiliary factors from several organisms that assist group II intron splicing have been identified, and the molecular mechanisms regarding the way these proteins work are now under extensive study [31,42,43]. For example, the maize CRS1 and CRS2 proteins facilitate group II introns in the chloroplast; CRS1 is required for only one, the *atpF* intron, while CRS2 is involved in the splicing of nine of the 10 chloroplast group IIB introns [6,15,37]. The expression of CRS2 in *E. coli* together with the corresponding RNA did not promote splicing, indicating that other protein(s) are also required [15]. Indeed, while this manuscript was under review, the discovery of new group II splicing factors that bound CRS2 and harbor a new characterized repeated domain, CRM, was reported [31]. Both maize CRS2 and *Arabidopsis* HCF152 participate in the splicing of the *petB* intron. However, these two components are not engaged in the same protein complex. It will be interesting to explore whether CRS2 and HCF152 can interact with each other and/or
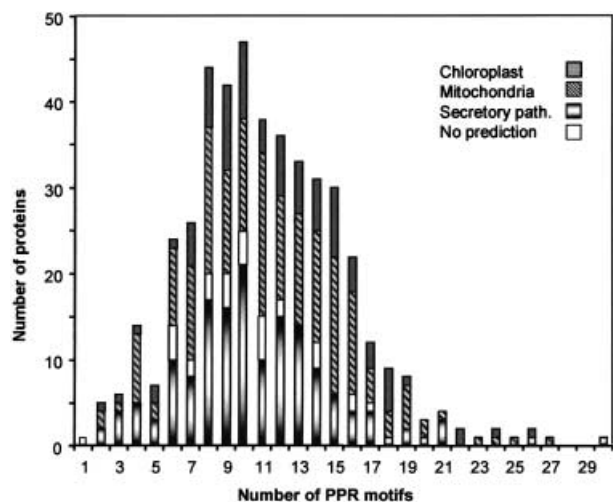
Fig. 9. Distribution of the number of PPR motifs in the *Arabidopsis* PPR proteins. The 5072 PPR motifs found in the 452 *Arabidopsis* proteins identified in the proteome database (*http://www.ebi.ac.uk/proteome/index.html*) were analyzed. The distribution of these proteins between the different organelles as predicted by TargetP (35) is indicated.

promote the splicing. A possible model of how HCF152 is involved in the stabilization of the splicing products of the *petB* intron could be that the protein binds to the UTR region between *psbH* and *petB*, and to domain IV of the *petB* intron. The binding of the homodimer of HCF152 in this region somehow stabilizes the splicing products, possibly by folding the RNA into the correct splicing structure.

PPR motif is a polynucleotide-binding domain

The PPR motif was first described by Small and Peeters as a special structural motif whereby six repeats create a tunnel that fits the size of one single strand of RNA [19]. So far, several PPR proteins, including HCF152, each containing a number of PPR motifs have been characterized as proteins involved in RNA and DNA metabolisms [20–27]. Two are characterized as DNA-binding factors [24,25], and therefore it appears that the PPR motif could be involved in both DNA- and RNA-binding. So far, proteins of the PPR family have not been identified in the prokaryote and in Archea, including cyanobacteria, which is believed to be closely related to the chloroplast ancestor (EMBL-EBI proteome database). Nevertheless, PPR proteins are very abundant in higher plants whereas other eukaryotic organisms contain no more than five PPR proteins. This observation suggests that this nucleus-encoded protein family has evolved into the 'tools' in which factors required for organelle gene expression are encoded and controlled by the nucleus gene expression machinery. Indeed, when the 452 'members' of the PPR family in *Arabidopsis* were analyzed for their location in the cell, 189 were predicted to be located in the mitochondria and 96 in the chloroplast (35; EMBL-EBI proteome database; Fig. 9).

In this study, we showed that the PPR motif is an RNA-binding domain. Yet high affinity binding could not be obtained with one motif only but was possible with several repetitions of the motif. Repetition of the motif seems to determine the specificity of binding to the target RNA sequence as well. Indeed, analyzing the PPR proteins of the *Arabidopsis* genome disclosed an average of 11 repetitions of this motif and often 7–16 repetitions were found (Fig. 9; EMBL-EBI proteome database). Accordingly, the predicted computerized structure of PPR proteins implies that six PPR motifs form a tunnel that fits the size of one single-stranded RNA [19]. In addition, the particular amino-acid sequence in each PPR motif is variable and probably contributes to the RNA binding properties. Indeed, as described above, the recent analysis of another PPR protein located mainly in the human mitochondria, LRP130, revealed a very limited contribution of the PPR motifs to the RNA-binding properties as the deletion of seven out of nine did not change the RNA-binding properties [41]. Defining the exact structure of the HCF152 homodimer together with the *petB* precursor (unspliced) RNA will uncover the molecular mechanism of how this protein specifically facilitates the processing of this transcript.

References

1. Sugita, M. & Sugiura, M. (1996) Regulation of gene expression in chloroplasts of higher plants. *Plant Mol. Biol.* **32**, 315–326.
2. Barkan, A. & Goldschmidt-Clermont, M. (2000) Participation of nuclear genes in chloroplast gene expression. *Biochimie* **82**, 559–572.
3. Monde, R.A., Schuster, G. & Stern, D.B. (2000) Processing and degradation of chloroplast mRNA. *Biochimie* **82**, 573–582.
4. Rochaix, J.D. (2001) Posttranscriptional control of chloroplast gene expression. from RNA to photosynthetic complex. *Plant Physiol.* **125**, 142–144.
5. Mullet, J.E. (1993) Dynamic regulation of chloroplast transcription. *Plant Physiol.* **103**, 309–313.
6. Jenkins, B.D., Kulhanek, D.J. & Barkan, A. (1997) Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. *Plant Cell* **9**, 283–296.
7. Perron, K., Goldschmidt-Clermont, M. & Rochaix, J.D. (1999) A factor related to pseudouridine synthases is required for chloroplast group II intron trans-splicing in *Chlamydomonas reinhardtii*. *EMBO J.* **18**, 6481–6490.
8. Rivier, C., Goldschmidt-Clermont, M. & Rochaix, J.D. (2001) Identification of an RNA-protein complex involved in chloroplast group II intron trans-splicing in *Chlamydomonas reinhardtii*. *EMBO J.* **20**, 1765–1773.

9. Felder, S., Meierhoff, K., Sane, A.P., Meurer, J., Driemel, C., Plucken, H., Klaff, P., Stein, B., Bechtold, N. & Westhoff, P. (2001) The nucleus-encoded *HCF107* gene of *Arabidopsis* provides a link between intercistronic RNA processing and the accumulation of translation-competent *psbH* transcripts in chloroplasts. *Plant Cell* **13**, 2127–2141.

10. Meurer, J., Meierhoff, K. & Westhoff, P. (1996) Isolation of high-chlorophyll-fluorescence mutants of *Arabidopsis thaliana* and their characterisation by spectroscopy, immunoblotting and Northern hybridisation. *Planta* **198**, 385–396.

11. Meurer, J., Berger, A. & Westhoff, P. (1996) A nuclear mutant of *Arabidopsis* with impaired stability on distinct transcripts of the plastid *psbB*, *psbD/C*, *ndhH*, and *ndhC* operons. *Plant Cell* **8**, 1193–1207.

12. Barkan, A., Voelker, R., Mendel-Hartvig, J., Johnson, D. & Walker, M. (1995) Genetic analysis of chloroplast biogenesis in higher plants. *Physiol. Plant.* **93**, 163–170.

13. Rochaix, J.-D. (1996) Post-transcriptional regulation of chloroplast gene expression in *Chlamydomonas reinhardtii*. *Plant Mol. Biol.* **32**, 327–341.

14. Goldschmidt-Clermont, M., Girard-Bascou, J., Choquet, Y. & Rochaix, J.D. (1990) *Trans.-splicing mutants of* Chlamydomonas reinhardtii. *Mol. General Genet.* **223**, 417–425.

15. Jenkins, B.D. & Barkan, A. (2001) Recruitment of a peptidyl-tRNA hydrolase as a facilitator of group II intron splicing in chloroplasts. *EMBO J.* **20**, 872–879.

16. Blatch, G.L. & Lassle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein–protein interactions. *Bioessays* **21**, 932–939.

17. Boudreau, E., Nickelsen, J., Lemaire, S.D., Ossenbuhl, F. & Rochaix, J.-D. (2000) The *Nac2* gene of *Chlamydomonas reinhardtii* encodes a chloroplast TPR protein involved in *psbD* mRNA stability, processing and/or translation. *EMBO J.* **19**, 3366–3376.

18. Vaistij, F.E., Boudreau, E., Lemaire, S.D., Goldschmidt-Clermont, M. & Rochaix, J.D. (2000) Characterization of Mbb1, a nucleus-encoded tetratricopeptide-like repeat protein required for expression of the chloroplast *psbB/psbT/psbH* gene cluster in *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA* **97**, 14813–14818.

19. Small, I.D. & Peeters, N. (2000) The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **25**, 46–47.

20. Fisk, D.G., Walker, M.B. & Barkan, A. (1999) Molecular cloning of the maize *crp1* gene reveals similarity between regulators of mitochondrial and chloroplast gene expression. *EMBO J.* **18**, 2621–2630.

21. Manthey, G.M. & McEwen, J.E. (1995) The product of the nuclear gene PET309 is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial COX1 locus of *Saccharomyces cerevisiae*. *EMBO J.* **14**, 4031–4043.

22. Lahmy, S., Barneche, F., Derancourt, J., Filipowicz, W., Delseny, M. & Echeverria, M. (2000) A chloroplastic RNA-binding protein is a new member of the PPR family. *FEBS Lett.* **480**, 255–260.

23. Mancebo, R., Zhou, X., Shillinglaw, W., Henzel, W. & Macdonald, P.M. (2001) BSF binds specifically to the *bicoid* mRNA-3′ untranslated region and contributes to stabilization of *bicoid* mRNA. *Mol. Cell. Biol.* **21**, 3462–3471.

24. Tsuchiya, N., Fukuda, H., Sugimura, T., Nagao, M. & Nakagama, H. (2002) LRP130, a protein containing nine pentatricopeptide repeat motifs, interacts with a single-stranded cytosine-rich sequence of mouse hypervariable minisatellite Pc-1. *Eur. J. Biochem.* **269**, 2927–2933.

25. Ikeda, T.M. & Gray, M.W. (1999) Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria. *Mol. Cell. Biol.* **19**, 8113–8122.

26. Bentolila, S., Alfonso, A.A. & Hanson, M.R. (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc. Natl Acad. Sci. USA* **99**, 10887–10892.

27. Coffin, J.W., Dhillon, R., Ritzel, R.G. & Nargang, F.E. (1997) The *Neurospora crassa cya-5* nuclear gene encodes a protein with a region of homology to the *Saccharomyces cerevisiae* PET309 protein and is required in a post-transcriptional step for the expression of the mitochondrially encoded COXI protein. *Curr. Genet.* **32**, 273–280.

28. Koizuka, N., Imai, R., Fujimoto, H., Hayakawa, T., Kimura, Y., Kohno-Murase, J., Sakai, T., Kawasaki, S. & Imamura, J. (2003) Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosena radish. *Plant J.* **34**, 407–415.

29. Desloire, S., Gherbi, H., Laloui, W., Marhadour, S., Clouet, V., Cattolico, L., Falentin, C., Giancola, S., Renard, M., Budar, F., Small, I., Caboche, M., Delourme, R. & Bendahmane, A. (2003) Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. *EMBO Report* **4**, 1–7.

30. Arabidopsis genome initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.

31. Ostheimer, G.J., Williams-Carrier, R., Belcher, S., Osborne, E., Gierke, J. & Barkan, A. (2003) Group II intron splicing factors derived by diversification of an ancient RNA-binding domain. *EMBO J.* **22**, 3919–3929.

32. Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N. & Schuster, G. (2003) HCF152, a PPR protein of *Arabidopsis* involved in processing of chloroplast *psbB-psbT-psbH-petB-petD* RNAs. *Plant Cell* **15**, 1480–1495.

33. Lisitsky, I., Klaff, P. & Schuster, G. (1996) Addition of poly (A)-rich sequences to endonucleolytic cleavage sites in the degradation of spinach chloroplast mRNA. *Proc. Natl Acad. Sci. USA* **93**, 13398–13403.

34. Schuster, G. & Gruissem, W. (1991) Chloroplast mRNA-3′ end processing requires a nuclear-encoded RNA-binding protein. *EMBO J.* **10**, 1493–1502.

35. Emanuelsson, O., Nielsen, H. & von Heijne, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* **8**, 978–984.

36. Auchincloss, A.H., Zerges, W., Perron, K., Girard-Bascou, J. & Rochaix, J.D. (2002) Characterization of Tbc2, a nucleus-encoded factor specifically required for translation of the chloroplast *psbC* mRNA in *Chlamydomonas reinhardtii*, *J. Cell Biol.* **157**, 953–962.

37. Till, B., Schmitz-Linneweber, C., Williams-Carrier, R. & Barkan, A. (2001) CRS1 is a novel group II intron splicing factor that was derived from a domain of ancient origin. *RNA* **7**, 1227–1238.

38. Lisitsky, I., Liveanu, V. & Schuster, G. (1995) RNA-binding characteristics of a ribonucleoprotein from spinach chloroplast. *Plant Physiol.* **107**, 933–941.

39. Dreyfuss, G., Matunis, M.J., Pinol-Roma, S. & Burd, C.G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62**, 289–321.

40. Graumann, P. & Marahiel, M.A. (1996) A case of convergent evolution of nucleic acid binding modules. *Bioessays* **18**, 309–315.

41. Mili, S. & Pinol-Roma, S. (2003) LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is

bound in vivo to mitochondrial and nuclear RNAs. *Mol Cell Biol.* **23**, 4972–4982.

42. Matsuura, M., Noah, J.W. & Lambowitz, A.M. (2001) Mechanism of maturase-promoted group II intron splicing. *EMBO J.* **20**, 7259–7270.

43. Gregan, J., Kolisek, M. & Schweyen, R.J. (2001) Mitochondrial Mg^{2+} homeostasis is critical for group II intron splicing in vivo. *Genes Dev.* **15**, 2229–2237.

44. Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E. & Tabata, S. (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana. DNA Res.* **6**, 283–290.

Supplementary material

The following material is available from http://blackwell
publishing.com/products/journals/suppmat/EJB/EJB3796/
EJB3796sm.htm
Table S1. The oligonucleotides used for the production of
RNA probes and HCF152 truncated proteins. Italicized
letters show promoter sequences for T7 RNA polymerase.
Underlined letters indicate the restriction enzyme site.