# An Arabidopsis cDNA encodes an apparent polyprotein of two non-identical thylakoid membrane proteins that are associated with photosystem II and homologous to algal ycf32 open reading frames

Alexandra Mant, Colin Robinson\*

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

Received 22 January 1998

Abstract We have characterised an Arabidopsis thaliana cDNA homologous to the ycf32 open reading frames present in the Synechocystis genome and the plastid genomes of several eukaryotic algae. The predicted protein is also homologous to a novel protein reported to be associated with photosystem II. The protein is synthesised as a 23 kDa precursor with an N-terminal presequence that appears to be bipartite in structure, and the protein is targeted into the thylakoid membrane of pea chloroplasts. Although the Ycf32 presequence contains an apparent signal peptide, we find that this protein is not imported by either of the standard Sec- or  $\Delta pH$ -dependent pathways. The mature protein is also unusual in two respects. First, there are two distinct, non-identical copies of typical single-span Ycf32 sequences in the Arabidopsis sequence, separated by an additional hydrophobic region. Secondly, the imported protein runs as a doublet of 6 kDa and 7 kDa polypeptides whereas the mature protein is predicted to be 14 kDa. We speculate that the protein undergoes further maturation once inserted into the thylakoid membrane to yield two separate Ycf32-like polypeptides.

 $© 1998 Federation of European Biochemical Societies.$ 

Key words: Chloroplast; Ycf32; Photosynthesis; Protein translocation; Thylakoid

### 1. Introduction

The targeting and assembly of chloroplast thylakoid proteins is a complex process involving the operation of an unexpected variety of pathways (reviewed in [1]). Most of these proteins are synthesised in the cytosol and imported into the chloroplast, after which they traverse the stromal phase and either insert into, or translocate across, the thylakoid membrane. The latter phases are accomplished using at least four distinct mechanisms, two of which have been identified for lumenal proteins and two for integral membrane proteins. Lumenal proteins are synthesised with bipartite presequences and, following removal of the envelope transit signal in the stroma, the second signal directs translocation by either a Sectype or  $\Delta pH$ -dependent protein translocase in the thylakoid membrane  $[2-5]$ . The targeting signals all resemble bacterial `signal' peptides in overall structure but are nevertheless capable of specifying translocation by only one of the two pathways [6,7]. A critical determinant in this process is the presence of a twin-arginine motif in transfer signals for the  $\Delta pH$ dependent system, shown to be essential for targeting by this route [8]. After translocation into the lumen, Sec- and  $\Delta pH-$ 

E-mail: CG@dna.bio.warwick.ac.uk

dependent proteins are cleaved to the mature sizes by a lumen-facing thylakoidal processing peptidase, TPP [9].

Integral thylakoid membrane proteins are inserted by two additional pathways. The multispanning light-harvesting chlorophyll-binding protein (LHCP) is synthesised with only an envelope transit signal and internal signals thus mediate insertion into the thylakoid membrane [10,11]. This process requires nucleoside triphosphates (NTPs) and a stromal homologue of bacterial signal recognition particles (SRPs) [12]. Recent evidence suggests that SRP plays an important role in the insertion of a range of bacterial plasma membrane proteins (reviewed in [13]), and the chloroplastic SRP pathway thus appears to fulfil a similar role. While this pathway was presumably inherited from the cyanobacterial-type progenitor of the chloroplast, a fourth pathway appears to be exclusive to chloroplasts. Three single-span membrane proteins  $(CF_0II,$ and the X and W subunits of photosystem II, termed PSII-X and PSII-W) are synthesised with bipartite presequences in which the second domain closely resembles Sec-type signal peptides. However, the insertion of  $CF_0II$  does not require NTPs or a  $\Delta pH$ , and insertion is furthermore unaffected by protease treatments that block the Sec-,  $\Delta pH$ - and SRP-dependent pathways [14,15]. PSII-X and PSII-W may well use a similar insertion mechanism [16,17]. These proteins have been proposed to insert spontaneously into the thylakoid membrane, possibly by a mechanism akin to that used by the M13 phage coat protein (reviewed in [18]).

In this report we have analysed the structure and import of a novel thylakoid protein that is homologous to the  $\gamma c f 32$ open reading frames found in cyanobacteria and the plastid genomes of several algae. We show that the Arabidopsis protein is highly unusual in containing two distinct copies of the Ycf32 sequence and we speculate that this precursor may be cleaved in the thylakoid membrane to yield two small polypeptides. One copy in particular shares strong homology with an N-terminal sequence determined for a novel small protein associated with PSII [19]. We also show that this protein is synthesised with a presequence that strongly resembles the bipartite presequences of some other thylakoid proteins, and targeted into the thylakoid membrane by a SecA- and  $\Delta pH$ independent process.

#### 2. Materials and methods

We searched the Expressed Sequence Tag (EST) database for clones homologous to putative single-span membrane proteins in the Porphyra purpurea plastid genome and identified several Arabidopsis cDNAs exhibiting significant homology to the  $ycf32$  open reading frame. One of these, clone number 132M13T7, was obtained from the Arabidopsis Biological Resource Center at Ohio State University. The clone was fully sequenced and the precursor protein synthesised

<sup>\*</sup>Corresponding author. Fax: (44) (1203) 523701.



#### 661

Fig. 1. Nucleotide and predicted amino acid sequence of an Arabidopsis thaliana cDNA encoding a ycf32 homologue. The start of the region of homology with other Ycf32 sequences is marked by an arrow (but note Fig. 2). The hydrophobic core region of the putative signal-type peptide and the predicted transmembrane helices in the mature protein are underlined.

by in vitro transcription-translation of the cDNA. The precursor protein was imported into pea chloroplasts or thylakoids as detailed in [14,16].

The EMBL accession number for the cDNA encoding Ycf32 is AJ223306.

# 3. Results and discussion

3.1. The Arabidopsis Ycf32 homologue contains an internal duplication, both of which are homologous to a novel PSII protein

An Arabidopsis cDNA encoding a Ycf32-related protein was identified by searching the database using the *Porphyra* purpurea sequence. Several homologous cDNAs were found, and sequencing of a full-length cDNA yielded a 189-residue predicted protein, the structure of which is shown in Fig. 1. Similarity with the *Porphyra* sequence begins after the first 79 residues, suggesting the presence of a presequence in the primary translation product. This is not surprising since all known nuclear-encoded chloroplast proteins are synthesised with envelope transit peptides, but there are indications that the Arabidopsis Ycf32 presequence may in fact be bipartite in structure. The first ca. 50 residues are enriched in hydroxylated residues and this region is also hydrophilic and positively charged; these are typical features of a simple chloroplast-targeting signal [20]. However, the following domain is very hydrophobic and the hydropathy plot of the presequence (Fig. 2) closely resembles those of the cleavable signal peptidetype domains of CF<sub>o</sub>II, PSII-W and PSII-X. These domains are removed by TPP, which is known to cleave following an Ala-Xaa-Ala consensus sequence [21]. However, a putative TPP cleavage site within pre-Ycf32 is difficult to identify because at least two such motifs are present in the region concerned (the PALA and LAAA sequences prior to the start of homology with the *Porphyra* sequence).

Both the primary structure and the hydropathy plot exhibit other unexpected features. ycf32 open reading frames have been identified in the plastid genomes of Porphyra purpurea, Odontella sinensis and Cyanophora paradoxa, and in the cyanobacterium Synechocystis sp. PCC 6803. The encoded proteins are all in the range of 36^39 residues and each is strongly predicted to contain a single transmembrane span (our un-



Fig. 2. Hydropathy analysis of pre-Ycf32. The pre-Ycf32 sequence was analysed using the TopPred II program [25] with a core window of 11 residues and a full window of 13 residues. Mean hydrophobicity is plotted against residue number; the hydrophobic core region of the signaltype peptide is denoted H and the predicted transmembrane spans labeled I-III. An arrow denotes the start of homology with algal Ycf32 sequences.

published observations). However, the `mature' Arabidopsis protein is about 110 residues in length (assuming that the mature protein starts as depicted in Fig. 1) and the hydrophobicity plot strongly predicts three transmembrane spans. Even more surprising is the observation that the Arabidopsis sequence contains two separate copies of the Ycf32 sequence (Fig. 3). Alignment of the Arabidopsis sequence with that of Porphyra Ycf32 shows the presence of two non-identical regions bearing very high similarity to the algal sequence, which correspond to the first and third predicted transmembrane spans in the `mature' protein (regions I and III in the hydropathy profile). This finding is consistent with two possibilities. One is that Arabidopsis Ycf32 may simply be far larger than the algal and cyanobacterial sequences, and contain three transmembrane spans of which two presumably arose by gene duplication. However, it is also possible that pre-Ycf32 is effectively a polyprotein, and that the second hydrophobic region in the mature protein is in fact a second signal-type peptide. In this case, cleavage by TPP after the second hydrophobic stretch would yield two proteins, although the first would also contain the second hydrophobic region (region II) unless this too is proteolytically removed. The region immediately after this hydrophobic stretch is very alanine-rich and several potential TPP cleavage sites can be discerned, but



Fig. 3. Alignment of pre-Ycf32 with the ycf32 open reading frame from the plastid genome of Porphyra purpurea, and with a novel PSII protein. The A. thaliana (A. thal.) and Porphyra (P. purp.) sequences were aligned using the CLUSTAL V program to generate the first block of homology. Note that the *same Porphyra* sequence was then aligned by eye against a more C-terminal region within the Arabidopsis Ycf32 sequence. The first 'mature' Ycf32 sequence is also aligned separately against the N-terminal sequence of a spinach protein associated with PSII (see text). Identical residues are denoted by asterisks, conserved residues by dots. Conserved hydrophobic regions are underlined.



Fig. 4. Import of pre-Ycf32 into pea chloroplasts and localisation of the processed product. Pre-Ycf32 was synthesised in vitro by transcription-translation and the radiolabeled translation product (lane Tr) was incubated with intact pea chloroplasts. After incubation, samples of the chloroplasts were analysed directly (lane C) or after protease treatment of the organelles (lane C+). Other samples were lysed and centrifuged to generate samples of stromal fraction (S) or thylakoids (lanes T). Right hand panel: pre-Ycf32 was imported into chloroplasts in the absence of any protease treatment, after which samples were obtained of enriched envelope (en) and thylakoid (T) membranes as detailed in [22].

further studies involving mutagenesis of potential TPP sites will be required to resolve these possibilities.

Fig. 3 also shows an alignment with the N-terminal sequence of a novel, small photosystem II protein that has been characterised in spinach and tobacco [19]. This protein was isolated from PSII complexes and shown to run as a 7 kDa protein, and evidence was presented that the protein, termed `L-arginine metabolising enzyme' may be involved in the water-splitting reaction. The N-terminal 20 residues of this protein were determined [19] and Fig. 2 shows that this sequence is virtually identical to the region of the Arabidopsis protein that is also predicted to be the N-terminus of the mature protein (or at least the first mature protein). This spinach protein is clearly homologous with algal Ycf32 proteins and we conclude that the Arabidopsis Ycf32 cDNA encodes at least one PSII protein.

# 3.2. Evidence that pre-YCF32 is cleaved to two smaller transmembrane polypeptides

In order to examine in greater detail the biogenesis of this protein we used in vitro assays for the import of proteins into intact chloroplasts and isolated thylakoids. Transcriptiontranslation of the Ycf32 cDNA yields a precursor protein of apparent molecular weight (on SDS-polyacrylamide gels) of 18 kDa, reasonably consistent with the size of the protein predicted from the cDNA sequence (22.8 kDa). Interestingly, this precursor is imported into chloroplasts and processed to a doublet of proteins which migrate as 6.5 kDa and 7 kDa (Fig. 4A). The mature protein is found in the thylakoid fraction (lane T) and further tests were carried out to confirm that this protein is indeed located in the thylakoids, since in practice this fraction often contains contaminating envelope membranes. The two types of membrane can be enriched by differential centrifugation [22] and this method was used to prepare membrane fractions after an import reaction. We used the full precursor form of Ycf32 as an internal marker for the envelope membranes, since import assays usually generate precursor molecules that are bound to import receptors but not yet imported. This is clearly the case with Ycf32 because the pre-Ycf32 recovered with the chloroplasts in the import assays (lane C in Fig. 4A) is completely susceptible to proteolysis (lane C+) whereas the imported mature protein is resistant. The right hand panel of Fig. 4 shows that the `thylakoid' fraction does indeed contain essentially all of the maturesize Ycf32. Some precursor molecules are also found in this fraction, indicating the presence of contaminating envelope membranes. The envelope membranes can be prepared in a more highly purified form (our unpublished observations) and this fraction (lane En) contains a significant proportion of the pre-Ycf32 as expected. More importantly, the virtual absence of mature-size Ycf32 means that this protein can not reside in the envelope membrane.

It is notable in the import assays that Ycf32 undergoes an enormous shift in mobility when processed within the chloro-



Fig. 5. Ycf32 is an integral membrane protein. A: Pre-Ycf32 was imported into chloroplasts and the thylakoid fraction prepared as in Fig. 4 (lane T). This membrane fraction was then subjected to two washes in 20 mM tricine-NaOH, pH 8.0 containing either 6.8 M urea or 2 M KSCN as detailed above the lanes. The washes were carried out according to [23] and samples were analysed of the final membrane pellet (P) or the first supernatant (Sn). B: As in A except that the precursor to petunia LHCP (pre-LHCP) was used as import substrate. C: Stained gel of thylakoid membranes (T) and the pellet and supernatant fractions generated as described above. The polypeptides indicated are LHCP in the pellet fraction and the  $\alpha$ - and  $\beta$ -subunits of the CF<sub>o</sub>CF<sub>1</sub> ATPase (ATPase) together with the 23 kDa and 33 kDa extrinsic proteins of the PSII oxygen-evolving complex (23K, 33K).



Fig. 6. Pre-Ycf32 is imported by a SecA/ $\Delta pH$ -independent pathway. A: Pre-Ycf32 was imported into chloroplasts in the absence of inhibitors (control) or in the presence of 2  $\mu$ M nigericin or 10 mM Na azide as indicated. After incubation, the chloroplasts were fractionated and analysed as in Fig. 4; in addition, samples were analysed of the thylakoid fraction after treatment with 200 µg/ml thermolysin for 40 min on ice (lanes T+). DP, degradation product; other symbols as in Fig. 4. B: Pre-Ycf32 (lane Tr) was incubated with isolated pea thylakoids, after which samples were analysed directly (lanes T), after two washes of the thylakoids (lanes Tw) or after protease treatment as in A (lane +).

plasts, from an apparent molecular weight of 18 kDa to only 7 kDa. Given that the predicted N-terminal presequence accounts for only 7^8 kDa, this result is consistent with the suggestion outlined above that 'mature' Ycf32 undergoes further maturation to yield two single-span proteins with similar structures to the algal Ycf32 sequences. The spinach protein likewise runs as 7 kDa on polyacrylamide gels [19]. Although further studies are required to resolve this point, since it remains possible that mature Ycf32 simply runs aberrantly on SDS polyacrylamide gels, these data represent very strong evidence that pre-Ycf32 is indeed processed to separate proteins. At least one of these proteins is apparently functional in PSII and further studies will be required to determine the function and location of the putative second protein.

The sequence data presented above strongly suggest that the mature Ycf32 protein(s) are membrane proteins. However, studies on the spinach L-AME protein [19] suggested that the purified protein may have enzymatic activity and the methodology used in this study did not rule out the possibility that this protein is soluble in the lumen. We therefore considered it important to resolve this point directly and we used diagnostic washes to determine whether imported mature Ycf32 is in the form of an integral membrane protein. Breyton et al. [23] have shown that urea and KSCN are highly effective in removing extrinsic proteins from thylakoid membranes, and we applied the same procedures to imported, mature-size Ycf32. Fig. 5A shows that the vast majority of the Ycf32 protein(s) remain membrane-associated after these treatments. A small proportion of protein is found in the supernatant fraction (Sn) after urea washing but this probably reflects failure to pellet some of the membranes since a similar phenomenon is observed in tests using imported light-harvesting chlorophyll-binding protein (LHCP, a multi-spanning membrane protein) as a marker (panel B). Neither Ycf32 nor LHCP are removed by KSCN treatment. The effectiveness of the urea washing is illustrated in the stained gel in panel C, which shows that LHCP remains in the membrane fraction whereas several extrinsic proteins, including the  $\alpha$ - and  $\beta$ -subunits of the ATPase and the PSII 23 kDa and 33 kDa proteins, are quantitatively removed. These data confirm that Ycf32 is an integral membrane protein, as predicted from the sequence data.

# 3.3. Ycf32 is imported by a SecA- and  $\Delta pH$ -independent pathway

Because Ycf32 is synthesised with an apparently bipartite

presequence (see below) we sought to analyse the import pathway used by this protein. Most proteins bearing cleavable signal peptides are imported by either a  $\Delta pH$ - or Sec-dependent route [1] although  $CF<sub>o</sub>II$  inserts by an apparently spontaneous mechanism and there are indications that PSII-X and PSII-W may use a similar mechanism [16,17]. Import assays were conducted in the presence of nigericin (a protonophore), or azide (a SecA inhibitor) to test whether pre-Ycf32 is imported by the typical  $\Delta pH$ - or Sec-dependent pathways (Fig. 6A). Neither compound has a significant effect on the import pro¢le, whereas in control tests the thylakoid-translocation of  $\Delta$ pH-dependent lumenal proteins was totally blocked and that of Sec-dependent proteins was drastically inhibited as found previously [7,24]. In this experiment samples of the thylakoids were treated with thermolysin, and the partial cleavage of the mature Ycf32 band(s) is further evidence that these are integral membrane proteins exposed on the stromal face of the thylakoid membrane. We conclude that pre-Ycf32 is imported by a different route, and we are now testing the possibility that this protein uses a  $CF<sub>o</sub>II$ -type insertion mechanism.

A different in vitro assay was used to test more directly the assumption that Ycf32 is synthesised with a bipartite presequence, in which we determined the location of the enzyme responsible for the maturation of pre-Ycf32. Simple `envelope transit' peptides are removed by the stromal processing peptidase whereas signal-type peptides are cleaved by a lumenfacing thylakoidal processing peptidase [21]. We have failed to observe any cleavage of pre-Ycf32 in assays for stromal processing activity (data not shown) but we have been able to reconstitute the insertion of this precursor into isolated thylakoids using the type of assay developed for  $CF_0II$  [14]. Incubation of pre-Ycf32 with isolated pea thylakoids leads to efficient maturation (Fig. 6B) and, importantly, all of the mature Ycf32 is found in the thylakoid fraction. We have not observed any mature Ycf32 in the supernatant from this type of assay and the mature-size protein is resistant to urea washing of the thylakoids, indicating that insertion has occurred (data not shown). All other membrane proteins with stromal processing peptidase cleavage sites are either cleaved in vitro by stromal processing peptidase in solution or are simply not cleaved in this type of in vitro assay; none has been found to be quantitatively processed once bound to the thylakoid membrane. This result represents strong evidence that pre-Ycf32 is matured by TPP after insertion into the thylakoid membrane, and we therefore propose that Ycf32 is indeed synthesised

with a bipartite presequence but targeted by a SecA- and  $\Delta p$ H-independent pathway.

Acknowledgements: This work was supported by Biotechnology and Biological Sciences Research Council Grant C07900 to C.R. We are also very grateful to the Arabidopsis Biological Resource Center for their support in this work.

#### References

- [1] Robinson, C. and Mant, A. (1997) Trends Plant Sci. 2, 431-437.
- [2] Nakai, M., Goto, A., Nohara, T., Sugita, D. and Endo, T. (1994) J. Biol. Chem. 269, 31338-31341.
- [3] Mould, R.M. and Robinson, C. (1991) J. Biol. Chem. 266, 12189^12193.
- [4] Yuan, J., Henry, R., McCaffery, M. and Cline, K. (1994) Science 266, 796^798.
- [5] Cline, K., Ettinger, W.F. and Theg, S.M. (1992) J. Biol. Chem. 267, 2688-2696.
- [6] Robinson, C., Cai, D., Hulford, A., Brock, I.W., Michl, D., Hazell, L., Schmidt, I., Herrmann, R.G. and Klösgen, R.B. (1994) EMBO J. 13, 279^285.
- [7] Henry, R., Kapazoglou, A., McCaffery, M. and Cline, K. (1994) J. Biol. Chem. 269, 10189-10192.
- [8] Chaddock, A.M., Mant, A., Karnauchov, I., Brink, S., Herrmann, R.G., Klösgen, R.B. and Robinson, C. (1995) EMBO J. 14, 2715^2722.
- [9] Kirwin, P.M., Elderfield, P.E., Williams, R.S. and Robinson, C. (1988) J. Biol. Chem. 263, 18128^18132.
- [10] Viitanen, P.V., Doran, E.R. and Dunsmuir, P. (1988) J. Biol. Chem. 263, 15000-15007.
- [11] Lamppa, G.K. (1988) J. Biol. Chem. 263, 14996-14999.
- $[12]$  Li, X., Henry, R., Yuan, J., Cline, K. and Hoffman, N.E. (1995) Proc. Natl. Acad. Sci. USA 92, 3789-3793.
- [13] De Gier, J.-W.L., Valent, Q.A., von Heijne, G. and Luirink, J. (1997) FEBS Lett. 408, 1^4.
- [14] Michl, D., Robinson, C., Shackleton, J.B., Herrmann, R.G. and Klösgen, R.B. (1994) EMBO J. 13, 1310-1317.
- [15] Robinson, D., Karnauchov, I., Herrmann, R.G., Klösgen, R.B. and Robinson, C. (1996) Plant J. 10, 149^155.
- [16] Kim, S.J., Robinson, D. and Robinson, C. (1996) FEBS Lett. 390, 175^178.
- [17] Lorkovic, Z.J., Schröder, W.P., Pakrasi, H.B., Irrgang, K.-D., Herrmann, R.G. and Oelmüller, R. (1995) Proc. Natl. Acad. Sci. USA 92, 8930-8934.
- [18] Kuhn, A. (1995) FEMS Microbiol. Rev. 17, 185-190.
- [19] Gau, A.E., Thole, H.H. and Pistorius, E.K. (1995) Z. Naturforsch. 50c, 638^651.
- [20] von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1989) Eur. J. Biochem. 180, 535-545.
- [21] Shackleton, J.B. and Robinson, C. (1991) J. Biol. Chem. 266, 12152^12156.
- [22] Flügge, U.-I., Fischer, K., Gross, A., Sebald, W., Lottspeich, F. and Eckershorn, C. (1989) EMBO J. 8, 39^46.
- [23] Breyton, C., de Vitry, C. and Popot, J.-L. (1994) J. Biol. Chem. 269, 7597^7602.
- [24] Knott, T.G. and Robinson, C. (1994) J. Biol. Chem. 269, 7843-7846.
- [25] Claros, M.G. and von Heijne, G. (1994) CABIOS 10, 685-686.