Plant Molecular Biology **15**: 157–160, 1990.

Plant Molecular Biology Update

Sequence of two tomato nuclear genes encoding chlorophyll *a*/*b*-binding proteins of CP24, a PSII antenna component

Egbert Schwartz and Eran Pichersky*
*Biology Department, University of Michigan, Ann Arbor, MI 48109, USA (*author for correspondence)*

Received 27 February 1990; accepted 28 February 1990

We have isolated and characterized (Fig. 1) two tomato nuclear genes each encoding a protein of 256 residues with discernible sequence identity with other CAB polypeptides. The two genes, which we designate *Cab*10A and *Cab*10B (tomato genes designated *Cab*1 through *Cab*9 encoded other CAB polypeptides of PSI and PSII [9, 10]), are tandemly linked and are separated by approximately 5 kb. Each gene contains a single, short intron – 199 bp in *Cab*10A, 97 bp in *Cab*10B – whose position (same in both genes) was determined by comparisons with cDNA clones. An intron in an equivalent position is also found in the tomato PSI Type II and Type III CAB genes [9]. Within the coding region, *Cab*10A and *Cab*10B are 92.5% identical, but the introns and 5′ and 3′ flanking regions show little similarity. Comparison of the sequence of the proteins encoded by *Cab*10A and *Cab*10B with the N-terminal sequence of the homologous (see below) spinach protein (Fig. 2A) indicates that the transit peptide of the CAB10A protein consists of 48 residues, and the transit peptide of CAB10B consists of 46 residues. Thus, the mature CAB10A protein has 208 residues, with a calculated molecular weight of 22.6 kDa, and the mature CAB10B protein has 210 residues, with a calculated molecular weight of 22.8 kDa. CAB10A and CAB10B are 94.5% identical within the mature part of the protein, and the sequences of the transit peptides are less similar.

Three hydrophobic regions are found in all CAB polypeptides examined to date, and it has therefore been hypothesized that they represent alpha-helices which span the thylakoid membrane, and that the three-dimensional structure of all CAB polypeptide is therefore similar despite the high level of primary sequence divergence [3, 10]. The hydropathy plots of CAB10A and CAB10B also indicate 3 hydrophobic regions of sufficient length to traverse the thylakoid membrane (Fig. 2B). It should be noted, however, that algorithms to predict secondary structure (e.g. [4]) do not always indicate complete overlap between the hydrophobic regions in the CAB polypeptides and regions predicted to assume the alpha-helix conformation, and this is especially true with CAB10A and CAB10B, where no extensive alpha-helix conformation is predicted (not shown) anywhere between the first and third hydrophobic regions. (However, these programs can predict alpha-helical regions in only 50%–60% of the cases where they indeed occur, as determined with proteins whose structure has been experimentally determined [4]). In addition to possessing the three hydrophobic regions, the CAB10A and CAB10B proteins also share the two regions of sequence similarity (Fig. 2C) found in all CAB proteins [10]. The two conserved regions include the first and third hydrophobic regions and their immediate N-terminal regions, and in these regions CAB10A and CAB10B dis-

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers *Cab*10A–M32605; *Cab*10B–M32606.

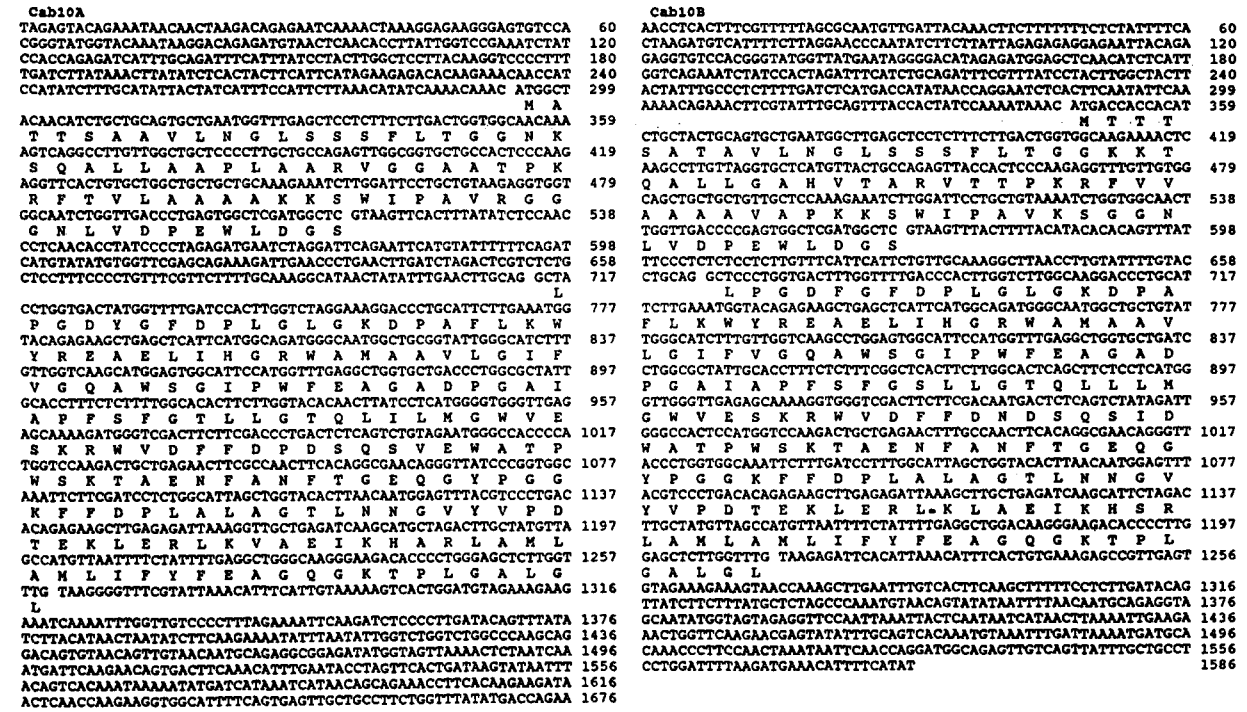

Fig. 1. Nucleotide sequence of *Cab*10A and *Cab*10B. The amino acid sequence of the encoded polypeptide is shown below the DNA sequence.

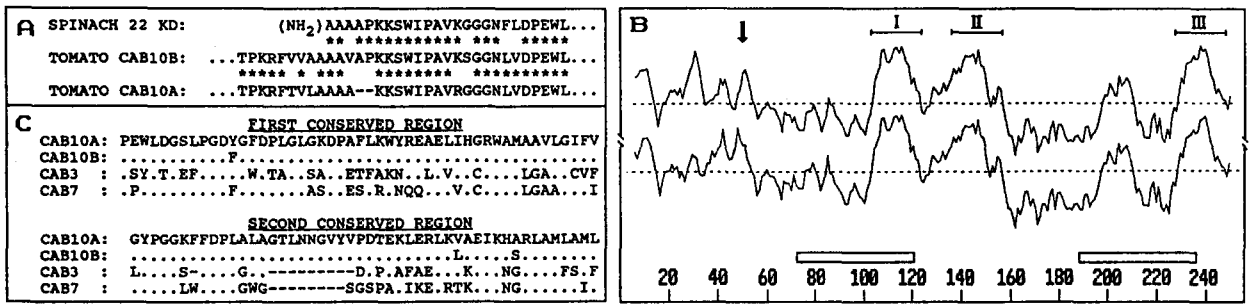

Fig. 2. A. Comparison of the N-terminal sequence of the spinach 22 kDa protein [6] with the corresponding regions in the CAB10A and CAB10B proteins. A dash represents a gap introduced to maximize identity. Dots represent additional sequence not determined (spinach sequence) or not shown (tomato sequence). Asterisks indicate same residues between two adjacent sequences. B. Hydropathy plots of CAB10A (upper plot) and CAB10B (lower plot). Calculations were carried out according to Kyte and Doolittle [5] with a window size of 15. Values above the dotted line are positive. Horizontal lines with Roman numerals above plots indicate the three hydrophobic regions of sufficient length (20 residues) to cross the membrane. Open bars below indicate the extent of sequence similarity of the CAB10A and CAB10B proteins with other CAB polypeptides. Arrows indicate the position of cleavage of the precursor. C. Comparisons of the two conserved regions of CAB polypeptides, corresponding to open bars in Fig. 2B. CAB3 is a PSII Type I CAB protein [7], CAB7 is a PSI Type II CAB [8]. Only the CAB10A sequence is shown in full. Dots in the other sequences represent residues identical to the amino acids in the equivalent positions in CAB10A. A dash represents a gap introduced to maximize identity.

play substantial sequence identity with other CAB proteins. Outside these two regions, CAB10A and CAB10B sequences display virtually no similarity with the corresponding regions of any other CAB proteins whose sequences have been determined. Especially noteworthy is the observation that a highly conserved region at the C-terminus of all other CAB proteins is completely missing from the CAB10A and CAB10B proteins, and the latter two proteins terminate with only a short extension past the third hydrophobic domain.

Murata *et al.* [6] have reported the N-terminal 26-residue sequence of a hydrophobic, 22 kDa PSII protein from spinach. Spangfort *et al.* [11, 12] have determined that this protein binds chlorophyll *a* and *b* and is localized in CP24, a minor antenna complex of PSII [2] (Spangfort *et al.* estimated the molecular weight of this protein as 20 kDa, but the N-terminal amino acid sequence they obtained is identical to that determined by Murata *et al.*, indicating that both groups were characterizing the same protein). Comparisons of the N-terminal sequence of the spinach protein with the corresponding regions in the CAB10A and CAB10B proteins (Fig. 2C) show 22/26 and 21/26 matches, respectively, indicating that these spinach and tomato proteins are homologous, and thus the CAB10A and CAB10B proteins are likely to be localized in the tomato CP24 complex. The high sequence identity at the mature N-terminus is especially indicative, since different types of CAB polypeptides show no discernible similarity among them in this region [10] (which is the reason why the terminal 26-residue sequence of the spinach protein did not reveal the relatedness of this protein to other CAB polypeptides). The conclusion of homology is further strengthened by the observation that no additional genes with high sequence identity to either a spinach probe encoding the 22 kDa protein or a probe derived from the cloned tomato genes were found in the tomato genome by Southern blot (data not shown) and by the comparison of the complete sequence of the cDNA clone encoding the spinach protein (N. Wedel *et al.*, in preparation) with the sequence of *Cab*10A and *Cab*10B.

A PSI CAB polypeptide designated PSI Type II CAB protein [8, 13] has been reported to be present in the LHCI-680 component of LHCI. Stayton *et al.* [13] concluded that this CAB protein was found also in CP24, but only because they made the *a priori* assumption that the LHCI-680 and CP24 were one and the same; they did not fractionate the thylakoid membranes into PSI and PSII fractions, and thus did not separate the two complexes, which migrated together on their mildly denaturing 'green' gel [1, 13]. We have determined that the PSI Type II CAB polypeptide is not present in PSII (Hoffman and Pichersky, unpubl.), and the spinach homolog of CAB10A and CAB10B was found in PSII but not in PSI [6, 11, 12]. Sequence comparisons indicate that the CAB10A and CAB10B proteins are greater than 60% divergent from all other CAB polypeptides, including the PSI Type II CAB protein. This strongly suggests that despite the observation that the CAB proteins of CP24 of PSII and LHCI-680 of PSI appeared to be immunologically and spectrally identical [1], the two complexes do contain distinct CAB polypeptides.

Acknowledgments

We thank Dr. R. Herrmann for communicating to us the unpublished sequence of a spinach CP24 cDNA clone (which facilitated the isolation of the tomato *Cab*10A and *Cab*10B genes) and Drs. B. Andersson and B. Green for communicating to us manuscripts prior to publication. This work was supported by grant 88-37262-3878 from US Department of Agriculture to E.P.

References

1. Bassi R, Hoyer-Hansen G, Barbato R, Giacometti GM, Simpson DJ: Chlorophyll-proteins of the photosystem II antenna system. J Biol Chem 262: 13333–13341 (1987).
2. Dunahay TG, Staehelin LA: Isolation and characterization of a new minor chlorophyll *a*/*b*-protein complex (CP24) from spinach. Plant Physiol 80: 429–434 (1986).
3. Karlin-Neumann GA, Kohorn BD, Thornber JP, Tobin

EM: A chlorophyll *a/b* protein encoded by a gene containing an intron with characteristics of a transposable element. J Mol Appl Genet 3: 45–61 (1985).
4. Garnier J, Osguthorpe DJ, Robson B: Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 120: 97–120 (1978).
5. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132 (1982).
6. Murata N, Kajiura H, Fujimura Y, Miyao M, Murata T, Watanabe A, Shinozaki K: Partial amino acid sequences of proteins of pea and spinach Photosystem II complex. In: Biggins J (ed) Progress in Photosynthesis Research, Vol. 1, pp. 701–704. Martinus Nijhoff, Dordrecht (1987).
7. Pichersky E, Bernatzky R, Tanksley SD, Breidenbach RB, Kausch AP, Cashmore AR: Molecular characterization and genetic mapping of two clusters of genes encoding chlorophyll *a/b*-binding proteins in *Lycopersicon esculentum* (tomato). Gene 40: 247–258 (1985).
8. Pichersky E, Tanksley SD, Piechulla B, Stayton MM, Dunsmuir P: Nucleotide sequence and chromosomal location of *Cab*-7, the tomato gene encoding the Type II chlorophyll *a/b*-binding polypeptide of Photosystem I. Plant Mol Biol 11: 69–71 (1988).
9. Pichersky E, Brock TG, Nguyen D, Hoffman NE, Piechulla B, Tanksley SD, Green B: A new nember of the CAB gene family: Structure, expression and chromosomal location of CAB8, the tomato gene encoding the Type III chlorophyll *a/b*-binding polypeptide of photosystem I. Plant Mol Biol 12: 257–270 (1989).
10. Pichersky E, Green BR: The extended family of chlorophyll *a/b*-binding proteins of PSI and PSII. In: Baltscheffsky M (ed) Current Research in Photosynthesis, Vol. III, pp. 553–556 (1990).
11. Spangfort M: Chlorophyll *a/b*-binding proteins of Photosystem II – isolation and characterization. Ph.D. Thesis, University of Lund, Sweden (1989).
12. Spangfort M, Larsson UK, Ljungberg U, Ryberg M, Andersson B, Klein R, Wedel N, Herrmann RG: The 20 kDa apo-polypeptide of the chlorophyll *a/b* protein complex of CP24. In: Barber J, Malkin R (eds) Techniques and New Developments in Photosynthesis Research. In press.
13. Stayton MM, Brosio P, Dunsmuir P: Characterization of a full-length petunia cDNA clone encoding a polypeptide of the light-harvesting complex associated with photosystem I. Plant Mol Biol 10: 127–137 (1987).

Note added in proof

The complete sequence of the spinach 22 kDa protein, deduced from the sequence of a cDNA clone [Spangfort *et al.*, in Current Research in Photosynthesis (M. Baltscheffsky, ed.), Vol. II, pp. 253–256 (1990)] is greater than 90% identical to CAB10A and CAB10B.