Mol Gen Genet (1990) 224:91–100

MGG

Light-harvesting proteins of diatoms: Their relationship to the chlorophyll *a/b* binding proteins of higher plants and their mode of transport into plastids

Arthur Grossman[1], Annamaria Manodori[2], and Donna Snyder[1]

[1] Carnegie Institute of Washington, 290 Panama Street, Stanford, CA 94305, USA
[2] Molecular Biology 151-S, Veterans Medical Center, 4150 Clement, San Francisco, CA 94121, USA

Received May 8, 1990

Summary. We have cloned and characterized members of a gene family encoding polypeptide constituents of the fucoxanthin, chlorophyll *a/c* protein complex, a light-harvesting complex associated with photosystem II of diatoms and brown algae. Three cDNA clones encoding proteins associated with this complex in the diatom *Phaeodactylum tricornutum* have been isolated. As deduced from the nucleotide sequences, these light-harvesting proteins show homology to the chlorophyll *a/b* binding polypeptides of higher plants. Specifically, the N-terminal regions of the fucoxanthin, chlorophyll *a/c*-binding proteins are homologous to the chlorophyll *a/b* binding proteins in both the third membrane-spanning domain and the stroma-exposed region between membrane-spanning domains 2 and 3. Like the chlorophyll *a/b*-binding proteins, the mature fucoxanthin, chlorophyll *a/c* polypeptides have three hydrophobic α-helical domains which could span the membrane bilayer. The similarities between the two light-harvesting proteins might reflect the fact that both bind chlorophyll molecules and/or might be important for maintaining certain structural features of the complex. There is little similarity between the N-terminal sequences of the primary translation products of the fucoxanthin, chlorophyll *a/c* proteins and any transit sequences that have been characterized. Instead, the N-terminal sequences have features resembling those of signal sequences. Thus either transit peptides used in *P. tricornutum* show little resemblance to those of higher plants and green algae or the nuclear-encoded plastid proteins enter the organelle via a mechanism different from that used in higher plants.

Key words: Light-harvesting – Fucoxanthin, chlorophyll *a/c* proteins – Protein transport – Chromophytic algae

Introduction

Different groups of photosynthetic organisms use different antennae pigment complexes to capture light energy used for photosynthesis. In cyanobacteria and red algae the major light-harvesting complex is a peripheral membrane complex called the phycobilisome (Gantt 1981; Glazer 1982, 1985; Grossman et al. 1988) while higher plants and green algae use chlorophyll *a/b* binding (CAB) proteins. Many other algal groups such as the diatoms, chrysophytes, and dinoflagellates have light-harvesting complexes in which xanthophylls are very prominent pigments.

In both higher plants and green algae a variety of related chlorophyll *a/b* binding proteins harvest light energy in complexes integral to the thylakoid membranes (Chitnis and Thornber 1988; Green 1988; Hoffman et al. 1987; Pichersky et al. 1987). Although all of these polypeptides (gene designation *cab*) exhibit sequence similarities, some serve to harvest light energy for photosystem II while others gather light energy for photosystem I (Hoffman et al. 1987). The CAB polypeptides are encoded in the nuclear genome and are synthesized as higher molecular weight precursors in the cytoplasm of the cell. The presequence, or transit peptide, is important for transport of the newly synthesized polypeptides across the double membrane of the chloroplast envelope (Schmidt and Mishkind 1986; Mishkind and Scioli 1988; Keegstra 1989). A considerable body of work has focused on the biosynthesis of the light-harvesting chlorophyll *a/b* protein complex, the determination of CAB protein sequences (as deduced from gene characterizations) and the light-modulated expression of the different *cab* genes (Chitnis and Thornber 1988; Murphy 1986; Tobin and Silverthorne 1985).

A number of pigment-protein complexes have been isolated from brown algae and diatoms (Alberte et al. 1981; Barrett and Anderson 1977, 1980; Berkaloff et al. 1990; Caron and Brown 1987; Caron et al. 1988; Kirk 1977; Peyriere et al. 1984). In the diatoms much of the light-harvesting is achieved by a fucoxanthin, chlorophyll *a/c* complex (FCPC) (Fawley and Grossman 1986; Friedman and Alberte 1984). As determined by fluorescence emission spectra, both chlorophyll *c* and fucoxanthin present in the complex can transfer light energy to chlorophyll *a* (Fawley and Grossman 1986; Friedman and Alberte 1984; Guggliemelli et al. 1981; Owens and

Offprint requests to: A. Grossman

Wold 1986; Owens 1986). The ratio of fucoxanthin to chlorophyll in this complex is approximately 2:1 (Friedman and Alberte 1984; Owens and Wold 1986) and the characteristic brown color of diatoms is a consequence of this high level of fucoxanthin in the cells. Similar pigmented systems are found in the brown algae (Barrett and Anderson 1977; Goedheer 1970, 1973), and xanthophyll-dominated light-harvesting complexes are present in a number of the other algal groups (Boczar and Prézelin 1986; Hiller et al. 1988; Prézelin and Alberte 1978; Prézelin and Haxo 1976).

In contrast to the considerable body of information on CAB polypeptides, little is known about the fucoxanthin, chlorophyll proteins (FCPs) and the biosynthesis of the FCPC. The polypeptides of the complex are integral to the thylakoid membranes and the association of fucoxanthin with the apoproteins of the complex is disrupted during SDS-polyacrylamide gel electrophoresis (PAGE). Like CAB polypeptides, FCPs are encoded in the nuclear genome and synthesized as higher molecular weight precursors in the cytoplasm of the cell (Fawley and Grossman 1986; Friedman and Alberte 1986). The presequence is probably involved in the transport of the newly synthesized polypeptide into the chloroplast. We have demonstrated that at least 3 polypeptides are constituents of the FCPC; these range in apparent molecular mass from 18 to 19.5 kDa (Fawley and Grossman 1986). Others have reported that this complex is composed of two polypeptides with molecular masses of approximately 17.5 and 18 kDa (Friedman and Alberte 1984). Differences in the apparent molecular masses and subunit composition are probably a consequence of different SDS-PAGE systems used in analyzing the polypeptides.

Antibodies raised against FCPs have been used to examine the relationship between the diatom light-harvesting complex and complexes of similar function in other chromophytic algae on the one hand (we will apply the term chromophyte to algae that have xanthophyll, chlorophyll *a*/*c* complexes), and the light-harvesting chlorophyll *a*/*b* complex of both chlorophytes and higher plants on the other. Friedman and Alberte (1987) reported that polyclonal antibodies raised against the polypeptides of the FCPC cross-reacted with light-harvesting polypeptides of other diatoms but did not cross-react with membrane polypeptides isolated from *Chrysophyta*, *Cryptophyta* and *Pyrophyta*. Cross-reactivity with the CAB polypeptides of *Chlorophyta* was not observed. Fawley et al. (1987) found that antibodies raised to the light-harvesting proteins of the diatom *Phaeodactylum tricornutum* cross-reacted with light-harvesting proteins of other diatoms, the prymnesiophyte *Pavlova gyrans* and brown algae, but did not cross-react with thylakoid membrane polypeptides from several Chrysophytes and other Prymnesiophytes. In contrast, Manodori and Grossman (1990) observed that antibodies raised against the light-harvesting proteins of *P. tricornutum* cross-reacted with thylakoid proteins from a range of different organisms including those of higher plants. Plumley and Schmidt (1984) observed that antibodies raised against a CAB polypeptide from *Chlamydomonas reinhardtii* cross-reacted with light-harvesting polypeptides of the diatom *Cylindrotheca fusiformis*. The results of Manodori and Grossman (1990) and Plumley and Schmidt (1984) suggest that some epitopes of the CAB light-harvesting proteins and the FCPs are similar.

To help establish the relationship between the CAB polypeptides and FCPs, we cloned and characterized cDNAs encoding the FCPs. The deduced protein sequences demonstrate that the FCPs have a region similar in structure to a region of the CAB polypeptides and, like CAB polypeptides, may be anchored in the thylakoid membranes by three membrane-spanning domains. The data also suggests that FCPs might enter the plastid by a mechanism that is different from that used by higher plants.

Materials and methods

Materials. All chemicals were of reagent grade. Restriction enzymes were from Bethesda Research Laboratories, Boehringer Mannheim, United States Biochemical, and Pharmacia. The Klenow fragment of DNA polymerase I, T4 DNA ligase, and Sequenase (version II) were from United States Biochemical. The radioactive nucleotides [α-^{32}P]dCTP (>3000 Ci/mmol) and [α-^{35}S]dATP (>600 Ci/mmol) were from Amersham.

Cultures. Phaeodactylum tricornutum Bohlin (University of Texas Culture Collection, strain 646) was grown at 25° C with continuous illumination of 150 μE m^2 s. Cultures were bubbled with air containing 3% CO_2 and grown in ESAW artificial sea-water medium (Harrison et al. 1980) supplemented with ten times the normal levels of nitrate ($NaNO_3$) and phosphate (K_2HPO_4), and buffered at pH 7.7 with 10 mM TRIS-HCl. Vitamins and silicate were omitted from the medium. Recombinant plasmids or phage particles were grown in *Escherichia coli* strains BB4, XL-1 Blue, NM514, or JM101.

Preparation of genomic DNA. Cells were centrifuged for 10 min in a GS3 rotor at 6000 × *g* at 4° C. The supernatant was discarded and the cell pellet from a 6 l culture was resuspended in 15 ml of 50 mM TRIS-HCl, pH 8.0, 10 mM EDTA, 1.0% SDS, 10 mM dithiothreitol and incubated at 37° C for 15 min. During this period the cells lysed. The lysate was extracted with one vol. of phenol followed by extractions with one vol. of phenol:chloroform (1:1) and one vol. of chloroform. The aqueous phase was collected, CsCl was added to 1.2 g/ml and Hoechst dye (bisbenzimide 33258) to 0.4 mg/ml and the extract was centrifuged in a Beckman Vti62.5 rotor for 12–15 h. Both a plastid and nuclear DNA band were collected, the dye was removed with isopropanol saturated with 10 mM TRIS-HCl, pH 8.0, 1 mM EDTA (TE) and CsCl (five extractions) and the DNA solution dialyzed overnight against TE. The dialysate was made 150 mM in NaOAc and the DNA was precipitated overnight at −20° C after the addition of two vol. of ethanol. The DNA was collected by centrifugation at 10000 × *g*

for 10 min, dried in vacuo and resuspended in distilled H_2O to a final concentration of approximately 1 mg/ml. Restriction digests were performed according to conditions described either by Maniatis et al. (1982) or by the manufacturer of the restriction enzyme.

Preparation of expression library. Genomic DNA from *P. tricornutum* was digested with *Sau*3A in reactions that ranged from 10% to 90% completion. DNA fragments were separated in a 0.8% agarose gel and the region of the gel containing fragments 3–6 kb in length was excised. The DNA was electroeluted from the gel in 45 mM TRIS-borate, 1.25 mM EDTA (Maniatis et al. 1982), and extracted five times with phenol, once with chloroform:phenol (1:1 and twice with chloroform prior to precipitation at −20° C with two volumes of ethanol. λZAP DNA was digested to completion with *Xho*I. The first two bases of the *Sau*3A (insert DNA) and *Xho*I (vector DNA) sites were filled-in using the Klenow fragment of DNA polymerase I and the appropriate nucleotides. The filled-in products, which have compatible ends, were ligated with T4 DNA ligase, and packaged in vitro. The packaged ligation mixture was used to infect *E. coli* strain BB4 and the plaques were screened with antibodies (Fawley and Grossman 1986) raised against the FCPs. Two immunopositive clones were obtained. The internal Bluescript SK− plasmid containing inserts of between 3 and 6 kb were rescued from the λ clones in *E. coli* XL-1 Blue (Short et al. 1988). Restriction sites in the clones were mapped using the enzymes *Bam*HI, *Hin*dIII, *Eco*RI, *Pvu*II, *Pst*I, *Sal*I, and *Xho*I.

Preparation of RNA. For the preparation of RNA, all glassware was baked for 8 h and all solutions and plasticware were autoclaved for 1 h. RNA was isolated from *P. tricornutum* by a modification of the method of Coleman and Grossman (1984). Cells were collected by centrifugation at $10000 \times g$ for 10 min and resuspended in 50 mM TRIS-HCl, pH 9.0, 20 mM EDTA, 1% SDS. The lysate was extracted once with phenol, once with chloroform:phenol (1:1), and twice with chloroform. The aqueous phase was made to 0.2 M in NaCl and the nucleic acid was precipitated with two volumes of ethanol by gentle inversion. At this stage the DNA formed an aggregate that was easy to spool out of solution. The RNA was allowed to flocculate out of the solution for 2 h at −20° C before being collected by centrifugation for 10 min at $10000 \times g$. Two additional ethanol precipitations were performed before polyadenylated RNA was isolated according to the method of Cashmore et al. (1978) using polyU-Sepharose beads.

Hybridizations. Southern hybridizations were performed according to Conley et al. (1985). Regions homologous among the clones were mapped using specific restriction fragments from each of the clones as hybridization probes. A 1.2 kb *Bam*HI fragment from clone 6A (see Fig. 1) was used to screen a *P. tricornutum* cDNA library. For Northern analyses RNA was resolved on a 1.2% agarose gel containing 6% formaldehyde in 20 mM MOPS buffer, 5 mM NaOAc, 1 mM EDTA, pH 7.0, and transferred to nitrocellulose. The DNA fragments were labeled by primer extension using random hexamers (Feinberg and Vogelstein 1983) and hybridized to RNA as described by Conley et al. (1985). RNA size markers in the low molecular weight range (0.16–1.77 kb, Bethesda Research Laboratories) were used to determine transcript lengths.

Preparation of cDNA. cDNA was prepared and cloned using kits RPN 1256Y/Z and RPN 1257 from Amersham. First-strand synthesis was primed with oligo dT and carried out using reverse transcriptase. Ribonuclease H was used to generate nicks in the RNA of the hybrids and DNA polymerase I was used to synthesize the second strand of DNA. Small 3′ overhangs on the first strand were eliminated with T4 DNA polymerase and the cDNA produced was ligated to *Eco*RI adapters. The *Eco*RI ends of the adapted cDNA were phosphorylated prior to ligation to λgt10 vector arms (predigested with *Eco*RI and treated with alkaline phosphatase). The recombinant molecules were packaged in vitro and the cDNA library was amplified in the host NM514. Plating and screening of the phage library and isolation of the phage DNA were performed as described by Maniatis et al. (1982).

Sequencing. The cDNA inserts encoding the light-harvesting polypeptides were excised from phage DNA preparations using either *Bam*HI or *Kpn*I. Inserts were ligated into appropriate restriction sites in the polylinker of M13mp19. Single-stranded templates were isolated and sequencing was performed by the dideoxy chain-termination reactions (Sanger et al. 1977) using Sequenase (version II), according to the manufacturer's protocols (United States Biochemical Corporation). When required as primers for sequencing reactions, oligonucleotides were synthesized using the Biosearch 8600 oligonucleotide synthesizer with reagents from Milligene Biosearch. Each cDNA clone was sequenced in both directions. DNA sequence analysis was performed using the IBI sequence analysis program, Version 2.02. Analysis of protein structure and hydropathy was performed with both the IBI and the MSEQ (University of Michigan Software) programs.

Results

Approximately 100000 λZAP genomic clones were screened using the antibodies (Fawley and Grossman 1986) specific for the light-harvesting proteins of *P. tricornutum*. Restriction maps of the two immunologically positive clones that were isolated, clones 1A and 6A, are shown in Fig. 1. Regions of the inserts from these clones cross-hybridized (depicted as dark bars in Fig. 1). Preliminary sequence analyses of these and other similarly isolated clones (Manodori and Grossman 1990; and data not shown) suggested that FCPs show amino acid sequence similarities with the CAB proteins of higher plants (see below).

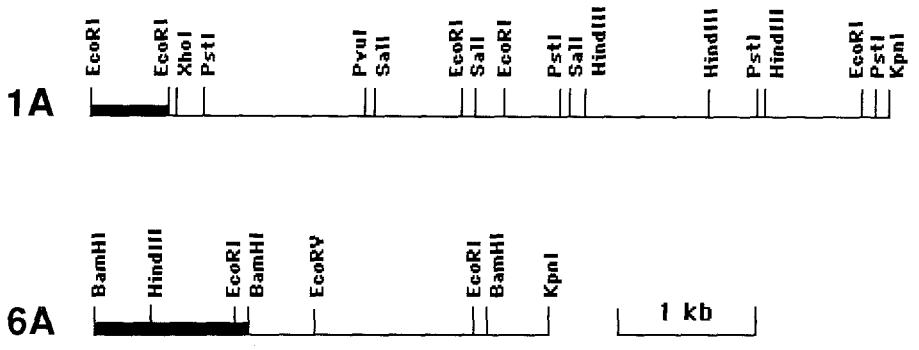

Fig. 1. Restriction map of genomic clones selected by immunologically screening a library of *Phaeodactylum tricornutum* genomic DNA in λZAP, with antibodies raised to the fucoxanthin chlorophyll proteins (FCPs). The *dark bars* above the insert DNA indicate the regions in which the two clones are homologous. The restriction sites at the borders of each of the clones (*Eco*RI and *Kpn*I for clone 1A and *Bam*HI and *Kpn*I for clone 6a) derive from the polylinker of the vector

Since genomic DNA was used to construct the expression library, it was possible that introns interrupted the coding regions of the genes selected. Therefore, we prepared a cDNA library from *P. tricornutum* polyA$^+$ RNA in the vector λgt10. The library, which represented over a million different cDNA clones, was amplified and screened with the homologous region on the immunologically selected clones (a 1.2 kb *Bam*HI fragment at the left border of clone 6A, see Fig. 1). DNA from the cDNA clones was purified and the insert excised with restriction endonucleases that cut in the *Eco*RI adapter that had been attached to the cDNA prior to ligation into λgt10. Inserts from two of the clones (λfcp1, λfcp2) were excised from the vector with *Kpn*I while for the third clone (λfcp3) the insert was removed with *Bam*HI. The three inserts were between 650 and 800 bp and hybridized strongly to the 1.2 kb *Bam*HI fragment that was initially used in the isolation of these clones.

Insert DNAs from λfcp1 and λfcp2 were hybridized to digests of the *P. tricornutum* genomic DNA, as shown in Fig. 2. The *fcp1* gene hybridized strongly to a 2.2 kb *Eco*RI fragment, a 7.4 kb *Sal*I fragment, a 7.8 kb *Pst*I fragment, and 3.1 kb and 0.9 kb *Bam*HI fragments. The *fcp2* gene hybridized strongly to a 1.55 kb *Eco*RI fragment, 3.1 kb and 1.7 kb *Bam*HI fragments, and *Pst*I and *Sal*I fragments of the same size as those to which *fcp1* hybridized. The *fcp1* and *fcp2* genes were likely to be different since they hybridized strongly to different *Eco*RI fragments. This was confirmed by sequence analyses (see below). However, since they hybridized to the same *Sal*I and *Pst*I fragments (and one common *Bam*HI fragment), they are linked on the *Phaeodactylum* genome. Based on a similar analysis, *fcp3* may also be linked to *fcp1* and *fcp2* (data not shown). Therefore, there are clusters of genes in the diatom genome encoding the light-harvesting polypeptides. In addition to the strongly hybridizing bands, *fcp1* and *fcp2* hybridized weakly to several other DNA fragments in each of the digests, demonstrating that these genes are part of a nuclear, multigene family.

The sizes of the cDNAs, as shown in the sequences of Fig. 3A, are 658, 733, and 676 bp for *fcp1*, *fcp2*, and *fcp3*, respectively. The polypeptides encoded by the cDNAs are shown below the nucleotide sequences. The *fcp2* and *fcp3* cDNAs end with a polyA tail which in both cases is between 30 and 50 nucleotides long. The *fcp1* cDNA either has a short 3′ nontranslated region

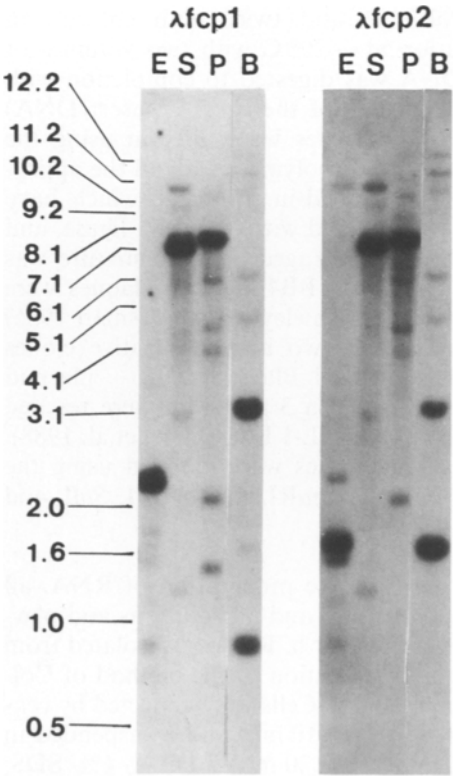

Fig. 2. Autoradiograms showing hybridization of *fcp1* and *fcp2* to restriction digests of genomic DNA. The genomic DNA was cleaved with *Eco*RI (E), *Sal*I (S), *Pst*I (P), and *Bam*HI (B). Molecular weight markers indicated were from Bethesda Research Laboratories (1 kb ladder, 5615SA)

(and a short polyA tail which is included in Fig. 3) or was truncated at the 3′ end during the cloning or subcloning step. The largest cDNA is *fcp2* which, if the polyA tail is included, is approximately 780 nucleotides long. This is approximately the size of the largest transcript to which this cDNA hybridizes (see below and Fig. 4). Thus, the *fcp2* cDNA must be nearly full-length and contains the complete coding region of the gene. The other 2 clones may be truncated either at the 5′ end (*fcp2*), or at both the 5′ and 3′ ends (perhaps *fcp1*), although based on comparisons with *fcp2* they too appear to encode complete FCP primary translation products (see below).

A

FCP1

```
CTCAATACCATGGGTTTCGAGATATCATTCAAGATGAAATTTGCCGTTTTTGCCTTCCTCCTCGCCTCTGCTGCCGCCTTTGCTCCGGCTCAGCAGTCGGCTCGTACTTCGGTAGCCACC  120
                                  M  K  F  A  V  F  A  F  L  L  A  S  A  A  A  F  A  P  A  Q  Q  S  A  R  T  S  V  A  T    29
AACATGGCTTTCGAGAACGAAATCGGTGCTCAGCAACCCCTCGGATATTGGGATCCCCTCGGTCTGGTCGCCGACGGTGACCAGGAGAAGTTCGACCGTCTCCGTTACGTTGAGATCAAG  240
 N  M  A  F  E  N  E  I  G  A  Q  Q  P  L  G  Y  W  D  P  L  G  L  V  A  D  G  D  Q  E  K  F  D  R  L  R  Y  V  E  I  K    69
CACGGGCGTATTTGTATGCTTGCCGTTGCCGGATACCTCACCCAAGAAGCCGGCATTCGTCTTCCGGGAGACATTGACTACTCAGGCACCAGCTTCGAATCGATTCCCAATGGATTCGCT  360
 H  G  R  I  C  M  L  A  V  A  G  Y  L  T  Q  E  A  G  I  R  L  P  G  D  I  D  Y  S  G  T  S  F  E  S  I  P  N  G  F  A   109
GCCTTGAGTGCTGTGCCTGGAGCCGGCATTGCCCAAATTATTGCTTTCATTGGCTTTCTCGAAATCGCCGTCATGAAGGACATTACTGGAGGAGAGTTTGTTGGCGACTTCCGCAACAAC  480
 A  L  S  A  V  P  G  A  G  I  A  Q  I  I  A  F  I  G  F  L  E  I  A  V  M  K  D  I  T  G  G  E  F  V  G  D  F  R  N  N   149
TACCTCGACTTTGGCTGGGACACTTTCAGTGAGGACAAGAAGCTTCAGAAGCGTGCTATCGAACTCAACCAGGGACGTGCGGCACAAATGGGCATCCTTGCTCTCATGGTGCACGAACAG  600
 Y  L  D  F  G  W  D  T  F  S  E  D  K  K  L  Q  K  R  A  I  E  L  N  Q  G  R  A  A  Q  M  G  I  L  A  L  M  V  H  E  Q   189
TTGGGAGTCTCTATCCTTCCTTAAAAATTTAATTTTCATTAGTTGCAGTCAAAAAAAA                                                               658
 L  G  V  S  I  L  P  *                                                                                                   196
```

FCP2

```
AACAACACATATATACGTTTCGACAAAATGAAGACTGCTGTCATTGCCTCTCTCATCGCCGGCGCCGCCGCCTTTGCCCCTGCCAAGAACGCGGCCCGTACTTCGGTTGCCACCAACATG  120
                            M  K  T  A  V  I  A  S  L  I  A  G  A  A  A  F  A  P  A  K  N  A  A  R  T  S  V  A  T  N  M    31
GCATTCGAAGACGAGCTCGGTGCTCAGCCTCCGCTTGGATTCTTCGATCCCCTCGGCCTTGTCGCCGACGGTGACCAGGAGAAGTTCGACCGTCTCCGTTACGTTGAGATCAAGCATGGA  240
 A  F  E  D  E  L  G  A  Q  P  P  L  G  F  F  D  P  L  G  L  V  A  D  G  D  Q  E  K  F  D  R  L  R  Y  V  E  I  K  H  G    71
CGTATTTCCATGCTTGCTGTTGTTGGATACCTCGTCCAGGAAGCCGGCGTCCGTCTTCCAGGAACCATCGACTACTCCGGCAAGACCTTCGCTGAGATCCCCAACGTCGCCGCCTTCAAG  360
 R  I  S  M  L  A  V  V  G  Y  L  V  Q  E  A  G  V  R  L  P  G  T  I  D  Y  S  G  K  T  F  A  E  I  P  N  V  A  A  F  K   111
GAGATCCCCGCTGGTGGACTTGTACAGCTTCTTTTCTTCATTGGAGTCCTTGAATCCAGTGTCATGCGTGACTTGACCGGCGAAGCCGAGTTCGTTGGTGACTTCCGTAACGGAGCCATC  480
 E  I  P  A  G  G  L  V  Q  L  L  F  F  I  G  V  L  E  S  S  V  M  R  D  L  T  G  E  A  E  F  V  G  D  F  R  N  G  A  I   151
GACTTCGGCTGGGACACCTTTGACGAGGAGACCCAGTTCAAGAAGCGTGCCATTGAGCTCAACCAGGGCCGCGCTGCCCAGATGGGAATTCTTGCCCTTATGGTGCACGAGCAGTTGGGT  600
 D  F  G  W  D  T  F  D  E  E  T  Q  F  K  K  R  A  I  E  L  N  Q  G  R  A  A  Q  M  G  I  L  A  L  M  V  H  E  Q  L  G   191
GTCTCTCTCCTCCCCCAGTAATTTTGTTACATTTACTGACTTCAAGGAGTCGAGGAATCGATACTGCCGTCGTTTCCAGGATCCGAGGTTTCATAAACTCTGTTAACGTTATAGAAACAG  720
 V  S  L  L  P  Q  *                                                                                                     197
ACTTACCTCTCCT                                                                                                            733
```

FCP3

```
ATGGAAACTATTCAAGATGAAGTTCGCCGTTTTCGCCTCCCTACTTGCCTCCGCCGCCGCCTTCGCTCCGGCCCAGCAGTCCGCCCGTACCTCGGTCGCCACCAACATGGCCTTCGAGAA  120
                M  K  F  A  V  F  A  S  L  L  A  S  A  A  A  F  A  P  A  Q  Q  S  A  R  T  S  V  A  T  N  M  A  F  E  N    35
CGAGCTCGGTGCCCAGCCTCCGCTTGGATTTTTCGATCCCCTCGGCCTCGTCGCCGACGGTGACCAGGAAAAGTTCGACCGTCTTCGTTACGTTGAGATCAAGCACGGACGTATTTCCAT  240
 E  L  G  A  Q  P  P  L  G  F  F  D  P  L  G  L  V  A  D  G  D  Q  E  K  F  D  R  L  R  Y  V  E  I  K  H  G  R  I  S  M    75
GCTTGCCGTTGCTGGATACCTCGTCCAGGAGAACGGAATCCGACTTCCGGGAGACATCGACTACTCCGGAACCAGCTTCGAATCGATCCCTAATGGTTTCGCCGCTTTGACCACCATCTC  360
 L  A  V  A  G  Y  L  V  Q  E  N  G  I  R  L  P  G  D  I  D  Y  S  G  T  S  F  E  S  I  P  N  G  F  A  A  L  T  T  I  S   115
CGGTGCTGGTATTGCCCAGATCGTTGCTTTCATTGGATTCCTCGAGCTCGCCGTCATGAAGGATATCACTGGAGGAGAATTCGTAGGTGACTTCCGTAATGACTTTATTGATTTCGGCTG  480
 G  A  G  I  A  Q  I  V  A  F  I  G  F  L  E  L  A  V  M  K  D  I  T  G  G  E  F  V  G  D  F  R  N  D  F  I  D  F  G  W   155
GGACTCCTTTGATGAGGAAACCAAGATGCAGAAGCGTGCTATTGAGCTCAACCAGGGCCGTGCCGCTCAGATGGGTATCCTTGCCTTGATGGTCCACGAACAGCTAGGGGTCTCCCTCAT  600
 D  S  F  D  E  E  T  K  M  Q  K  R  A  I  E  L  N  Q  G  R  A  A  Q  M  G  I  L  A  L  M  V  H  E  Q  L  G  V  S  L  I   195
CCCCAACTAAGCTATGGGTTGATGTTCCTCCGGTTAGATCTAACAGGATTAGTGCAATTCGAGTTGAATCACTGGG                                            676
 P  N  *                                                                                                                  197
```

B

```
                10        20        30        40        50        60        70        80        90       100
                 .         .         .         .         .         .         .         .         .         .
FCP1           f                            I   q   Yw                           c        t  a
FCP3    MKFAVFASLLASAAAFAPAQQSARTSVATNMAFENELGAQPPLGFFDPLGLVADGDQEKFDRLRYVEIKHGRISMLAVAGYLVQENGIRLPGDIDYSGTS
FCP2      t  i   I g       kna            D                                           v      a V    t     kt

               110       120       130       140       150       160       170       180       190
                 .         .         .         .         .         .         .         .         .
FCP1               saVp       I       I        -         NYL     t s Dk L                             IL
FCP3    FESIPNGFAALTTISGAGIAQIVAFIGFLELAVMKDITG-GEFVGDFRNDFIDFGWDSFDEETKMQKRAIELNQGRAAQMGILALMVHEQLGVSLIPN
FCP2     ae   -v  Fke pag Lv LLf   v  ss  R L  Ea        ga      t     qfk                             L q
```

Fig. 3. A Nucleotide sequences of *fcp1*, *fcp2*, and *fcp3*. The amino acid sequences (single letter code) encoded by the cDNAs are given below the nucleotide sequences. Both the nucleotide and amino acid numbers are at the end of each line. **B** Homology among the FCPs. FCP1 and FCP2 are compared to FCP3. Blank spaces have been left for the amino acids of FCP1 and FCP2 which are identical to those of FCP3. A capital letter for an amino acid in FCP1 and FCP2 indicates that the amino acid represents a conserved or neutral substitution relative to FCP3 while a small letter indicates a nonconserved substitution. The amino acid number, in increments of ten, is located above the sequence

The designated initiator codons in all 3 clones were chosen for a number of reasons. Each is the first Met codon found in the 5′ region of the cDNA and initiates open reading frames (ORFs) of 196 amino acids for FCP1 and 197 amino acids for both FCP2 and FCP3. These sequences would have molecular weights of just over 21 kDa, which matches well the previously determined size for the primary translation products of the light-harvesting proteins of *P. tricornutum* (Fawley and Grossmann 1986; Friedman and Alberte 1986). Second-

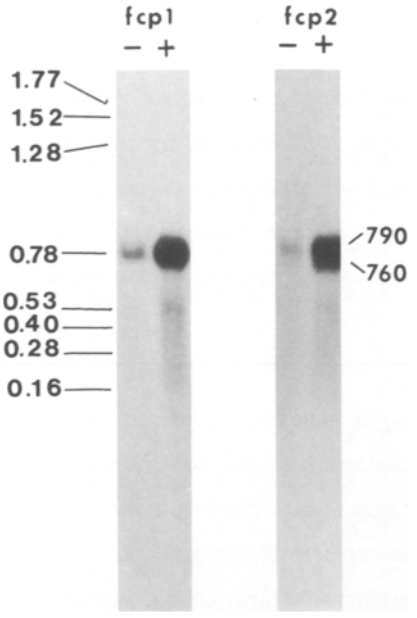

Fig. 4. Hybridization of *fcp1* and *fcp2* to polyA$^+$ and polyA$^-$ RNA. The inserts from *λfcp1* and *λfcp2* were purified from agarose gels and used as hybridization probes. The sizes of the transcripts were determined using low molecular weight markers (0.16–1.77 kb, 5623SA) from Bethesda Research Laboratories

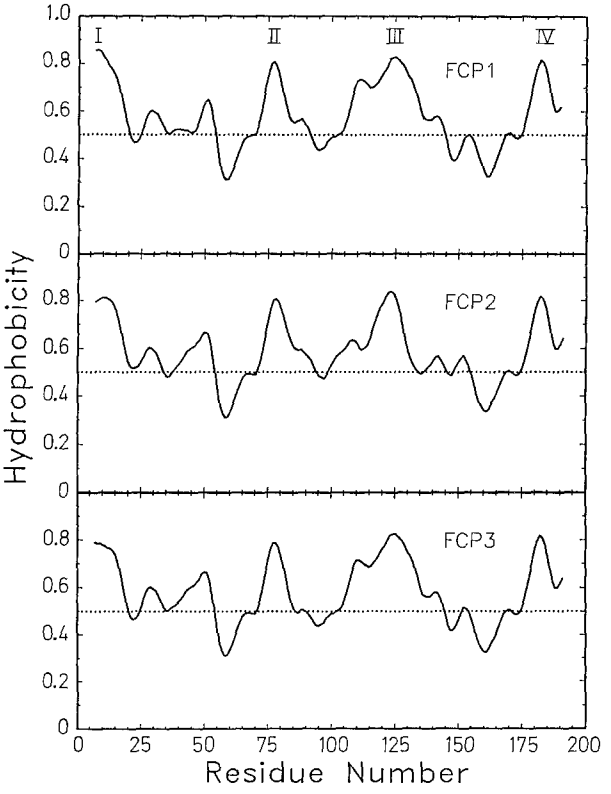

Fig. 5. Hydropathy plots of the FCPs. The range of the scale on the *y*-axis is from 0 to 1, with 0 being the most polar and 1 the most hydrophobic. The hydrophobicity parameters used were those of Argos et al. (1982) with a scanning factor of 13. The residue number is on the *x*-axis. The peaks labeled I, II, III, and IV are the most hydrophobic regions of the protein and have the potential to form membrane-spanning domains

ly, there is no homology between sequences upstream of this initiator Met among the different clones; the homology among the clones begins at, or very near to, this Met. Thirdly, the ORFs have homology to CAB polypeptides of green algae and higher plants (discussed below, see Fig. 6). Fourthly, in analyzing some genomic clones for different members of the gene family (which appear to be colinear with the cDNAs), stop codons were found just 5′ to this Met (data not shown).

The *fcp1* and *fcp2* genes hybridized strongly to very abundant transcripts, as shown in Fig. 4. While *fcp1* hybridized to a transcript of approximately 790 nucleotides, *fcp2* hybridized to two transcripts, one of approximately 750 nucleotides and the other of approximately 790 nucleotides. These transcripts, as expected for genes encoded in the nuclear genome, are enriched in the polyadenylated fraction of the RNA. The considerable homology among the *fcp* genes that have been characterized suggests that many will hybridize to transcripts derived from other members of the gene family. This, plus the variability in length of the polyA tail, may explain why the bands to which the cDNAs hybridized are broad.

A comparison of the three proteins encoded by the *fcp* genes is presented in Fig. 3B. The proteins are very similar; of 197 amino acids, 174 residues are identical between FCP3 and FCP1. Most of the differences represent conservative substitutions. FCP2 also exhibits a high degree of sequence similarity to FCP3 and FCP1. The strong similarity among all three proteins is reflected in the hydropathy plots in Fig. 5. There are four regions [labeled I (amino acids 1–20), II (amino acids 73–94), III (amino acids 110–133), IV (amino acids 174–196)] of the putative primary translation products which exhibit a pronounced hydrophobic character. These same regions have a tendency to form α-helical configurations that are punctuated by β turns, as determined using the parameters of Garnier et al. (1978). These α-helical, hydrophobic domains of the FCPs may be membrane-spanning domains. Regions II, III, and IV may span the membranes in the mature FCP and would be analogous to the three membrane-spanning domains of CAB polypeptides of higher plants. The first hydrophobic domain is at the N-terminus of the protein and is probably part of the presequence involved in targetting these proteins to plastids. However, this domain shows little resemblance to N-terminal extensions on nascent proteins targetted to higher plant chloroplasts. The latter sequences often have a tripartate conserved framework (Karlin-Neuman and Tobin 1986) which is not present at the N-terminus of the FCPs. The N-termini of the FCPs do, however, have some features reminiscent of signal sequences which initiate cotranslational transport across the endoplasmic reticulum. This intriguing finding is more thoroughly explored in the discussion.

Finally, a comparison of the FCPs with other CAB polypeptides suggests both a functional and structural relationship between light-harvesting polypeptides of the diatom and higher plants. The sequence from amino

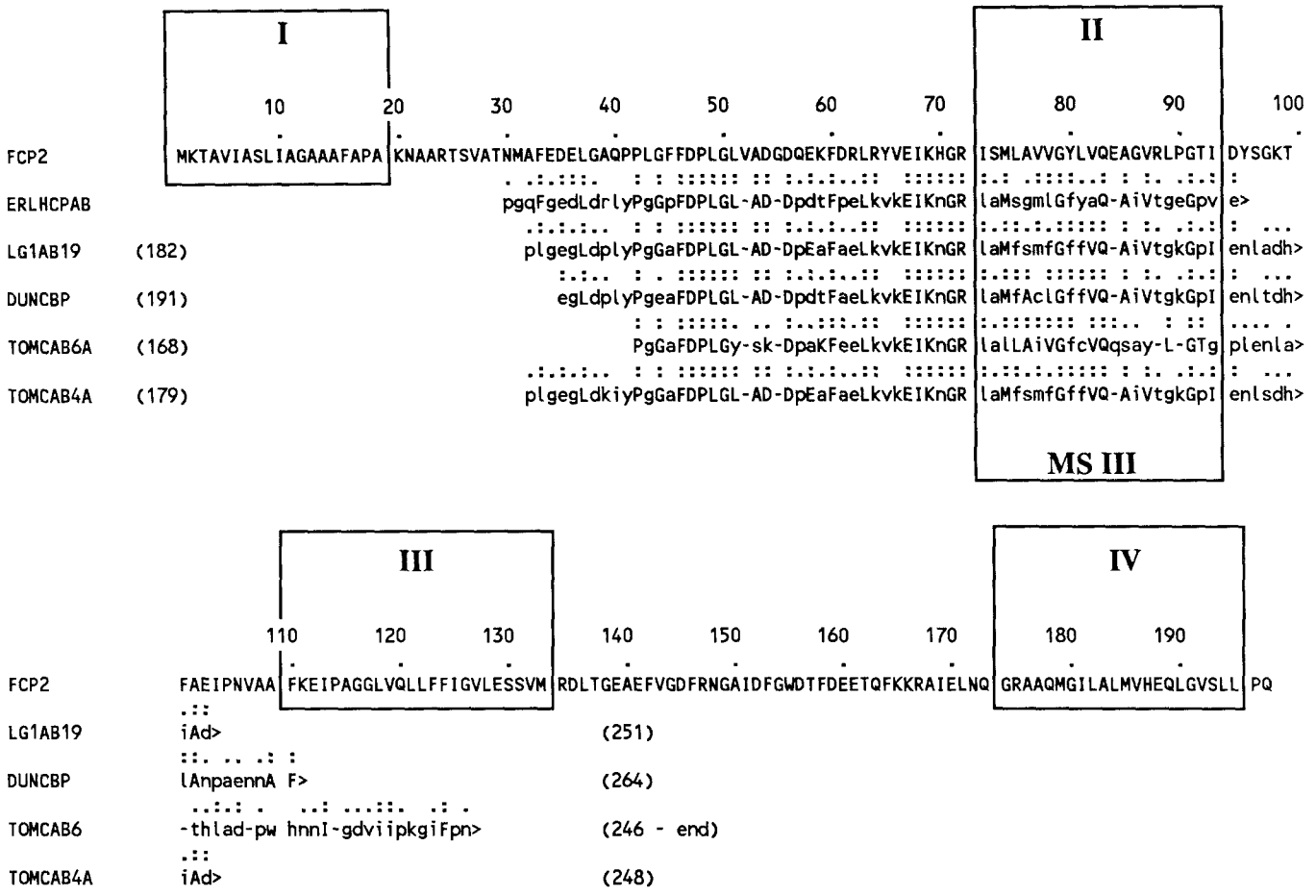

Fig. 6. Homology of FCP2 with the CAB polypeptides. Identical amino acids are capitalized, a colon indicates an identity or a conserved substitution, a dot a neutral substitution, and a space a lack of conservation. Hydrophobic region II of the FCPs (see Fig. 5) and membrane-spanning region III (MS III) of the CAB polypeptides are aligned and *boxed*. The other three hydrophobic regions (I, III, IV) of the primary translation product of FCP2 are also boxed. The amino acid residues of the CAB polypeptides at which the homology begins and ends (except for the *Euglena* sequence) are given in *parenthesis* to the left and right, respectively, of the amino acid sequences; only a partial sequence exists for the CAB polypeptide of *Euglena*. The numbering is from the initiator Met of the primary translation product. The different CAB polypeptides aligned are from *Euglena gracilis* (ERLHCPAB, Houlne and Schantz 1987), *Lemna gibba* (LG1AB19, Karlin-Neuman et al. 1985), *Dunaliella salina* (DUNCBP, Long et al. 1989), and *Lycopersicon esculentum* (TOMCAB4A and TOMCAB6A, Pichersky et al. 1987; Hoffman et al. 1987)

acids 30–100 of the putative primary translation products of the FCPs are homologous to the C-terminal regions of several CAB polypeptides. This explains the results which demonstrated an immunological similarity between FCP and CAB polypeptides. Figure 6 shows a comparison of FCP2 with CAB polypeptides of *Euglena*, *Lemna*, *Dunaliella*, and tomato. Based on models proposed for the integration of CAB polypeptides into the thylakoid membranes, the N-terminal regions of FCPs are homologous to CAB polypeptides in the regions of the third membrane-spanning domain and the stroma-exposed loop that connects membrane-spanning domains II and III.

Discussion

The characterization of genes encoding the fucoxanthin, chlorophyll *a*/*c* binding proteins, the first nuclear encoded genes that have been characterized in the diatoms, has expanded our knowledge of the components of the fucoxanthin-based light-harvesting complexes and provided some insight into the relationship between FCPs and CAB proteins of higher plants. First, the fucoxanthin, chlorophyll *a*/*c* polypeptides are encoded by a multigene family that resides in the nuclear genome. Based on both limited analyses of genomic clones and the number of different restriction fragments that hybridize to the cDNAs, there are a minimum of six distinct genes in this multigene family in *P. tricornutum*. Sequences of the three cDNA clones presented here, plus two additional genomic clones that have been sequenced (manuscript in preparation) have demonstrated that many of the genes in this family are similar. Second, based on both hydropathy and secondary structure analyses, the FCP primary translation products have four α-helical hydrophobic domains, only three of which are probably present in the mature protein. These regions have the potential to span the photosynthetic membrane in a manner similar to that of the CAB polypeptides (Karlin-

Neuman et al. 1985; Peter and Thornber 1988). Third, as suggested by previous information on the cross-reactivity of antibodies raised against the FCPs (Plumley and Schmidt 1984; Manodori and Grossman 1990), there are some sequence similarities between the CAB polypeptides and FCPs. The N-terminal regions of the FCPs contain sequences that are similar to the C-terminal regions of the CAB polypeptides. In CAB polypeptides this region encompasses the third membrane-spanning domain and the stroma-exposed loop located between membrane-spanning domains II and III. Finally, the putative presequence, which probably comprises 20–30 amino acids of the primary translation product (similarity with the C-terminus of the CAB polypeptides begins at approximately amino acid 30), does not resemble presequences, or transit peptides, found in nuclear-encoded proteins that are targetted to the plastids of higher plants. The properties of this presequence are more similar to those of signal sequences (see below), which are required for cotranslational transport of proteins into the endoplasmic reticulum (Gierasch 1989).

The region common to FCPs and CAB polypeptides probably reflects similar functions that have been maintained in these light-harvesting polypeptides. The conserved region could be important for the binding of chlorophyll molecules and/or the interactions of the light-harvesting proteins with other pigment protein molecules in the thylakoid membranes. Recently a model for the orientation of CAB polypeptides in the membrane has been presented. Each of the polypeptides is thought to bind between 6 and 13 chlorophyll molecules (Kuhlbrandt 1984; Ryrie et al. 1980; Suss 1983; Thornber 1975; Thornber et al. 1979). The three putative membrane-spanning regions are bordered by β turns and a cluster of charged residues that might anchor the hydrophobic α helices in the lipid bilayer. The FCPs also have charged amino acid residues and β turns that border the putative membrane-spanning helices. This is especially apparent for hydrophobic regions II and IV (see Fig. 6, amino acids 73–92 and 173–195). Some of the conserved charged residues may have more specific functions. Positively charged amino acids (His, Gln, Asn) in, or close to, the membrane-spanning regions might co-ordinate the chlorophyll molecules within the complex (Peter and Thornber 1988). In the purple and green photosynthetic bacteria, His residues in the reaction center polypeptides are important in orienting the bacteriochlorophyll molecules (Deisenhofer et al. 1985; Zuber 1985). The charged residues might also form salt bridges that stabilize the monomeric and/or oligomeric conformation of the light-harvesting proteins. Other functions of the regions conserved between FCP and CAB polypeptides may be related to the interactions of these proteins with other polypeptides of the photosynthetic membranes, or with ions and metabolites that might regulate both the efficiency of energy transfer to the reaction centers and/or the relative proportion of energy distributed between the two photosystems.

Fucoxanthin, an oxygenated carotenoid, is the major pigment associated with the FCPC. The CAB light-harvesting complexes in higher plants also contain oxygenated carotenoids, the major one being lutein. While the ratio of fucoxanthin to chlorophyll in FCPC in approximately 2:1 there is only one xanthophyll molecule for every 5–10 chlorophyll molecules (Kolubayev et al. 1985) in the CAB light-harvesting complexes. However, the xanthophyll in the latter complex is vital to its assembly (Plumley and Schmidt 1987) and can also function in harvesting light energy and transferring it to chlorophyll molecules (Siefermann-Harms and Ninnemann 1983). The lack of homology with CAB polypeptides at the C-termini of FCPs might reflect the differences in the pigments that bind to these light-harvesting proteins; the C-termini of the FCPs might be specialized to bind several molecules of fucoxanthin.

Finally, the primary translation products of the FCPs do not appear to have presequences that resemble transit peptides, the sequences involved in the posttranslational transport of proteins into chloroplasts (Schmidt and Mishkind 1986; Mishkind and Scioli 1988). There is little similarity between the first 30 amino acids of the primary translation products of the FCPs (the size of the transit peptide varies but is often comprised of between 30 and 50 amino acids) and any transit peptide that has been characterized to date. An amino acid framework deduced for higher plant transit sequences has three specific domains (Karlin-Neuman and Tobin 1986) which are not present in the putative presequence of the FCPs. The N-termini of the FCPs are, however, similar to signal sequences. The N-terminal domain of a signal sequence has one, or a cluster of basic amino acids followed by a hydrophobic central region. As the cleavage site of the sequence is approached it becomes more hydrophilic. The FCPs initiate with a Met followed by the basic Lys residue (Lys is found at the second position of many signal sequences). Following the Lys-2 of the FCPs there is a sequence of 10–12 amino acids which is very hydrophobic. Similar to a signal sequence, this hydrophobic domain is followed by a region (amino acids 15–25) that becomes progressively more hydrophilic (see Fig. 6). In contrast, many transit peptides begin with the sequence Met–Ala–Ala/Ser/Thr (see Chitnis and Thornber 1988), and we are unaware of any that have a positively charged amino acid in the second position. For transit peptides the sequence Gly–Arg–Val often precedes the cleavage site. No such sequence is present in the N-terminal region of the primary translation products of the FCPs. However, based on the rules compiled by von Heijne (1985, 1986) for the prediction of signal sequence cleavage sites, we would predict such a site following the Thr residue at position 29 of FCP1, FCP2 and FCP3 (although there are other potential cleavage sites). The '−1, −3 rule' (amino acid position relative to cleavage site) predicts that the residue at position −1 would be small (Ala, Ser, Gly Cys, Thr, or Gln) and that the amino acid at −3 must neither be charged nor large and polar. For eukaryotic organisms the prevalent amino acids at −3 are Ala, Val, or Ser although the rules for predicting signal sequence cleavage sites are much less reliable than for prokaryotes. In prokaryotes only Ala, Gly, Ser, and Thr have been found at the −1 position and only Ala, Gly, Leu, Ser,

Thr, and Val at the −3 position. All of these requirements are satisfied for the three FCPs if cleavage occurs after the Thr 29 residue. The −3, −1 positions would be Val, Thr for all three of the FCPs.

The hypothesis that a signal sequence is required for the transport of proteins into plastids of chromophytic algae is consistent with the morphological characteristics of plastids in these organisms. These organisms contain a plastid that is completely encased by two layers of endoplasmic reticulum (Gibbs 1962, 1970, 1978, 1979). This network, thought to be a barrier to the flow of macromolecules into the plastid, may have attached ribosomes (Gibbs 1978). Furthermore, there are many vesicles between this ER system and the plastid (Gibbs 1978), suggesting vesicular traffic between the two membrane systems. From the examination of electron micrographs coupled with studies in which protein synthesis was inhibited on either 80S cytoplasmic or 70S plastid ribosomes, Gibbs has suggested that proteins enter the plastids of the Chrysophyte *Ochromonas danica* via the fusion of vesicles, derived from the plastid ER, with the plastid envelope (Gibbs 1978, 1979). Hence, the FCPs may pass into the plastid ER of *P. tricornutum* via a cotranslational process, and then become packaged in vesicles that fuse with the plastid envelope.

While the arguments derived from the sequence data here and from previously published morphological data (Gibbs 1970, 1978, 1979) compellingly suggest that the transport of proteins into the plastids of the chromophytic algae involves a cotranslational process, they are still only circumstantial. Recent data demonstrating that the primary translation products of the FCPs can enter dog pancreas microsomes and be processed to the size of the mature protein (Bhaya and Grossman, unpublished) provides direct support for the hypothesis. Biochemical analyses, including in vitro reconstitution of protein transport, will enable us definitively to characterize processes required for the transport of proteins into the plastids of the chromophytic algae.

Acknowledgements. The authors would like to thank Glenn Ford for his excellent technical assistance, Beverley Green for her help in analyzing the sequence information, and Neil Hoffman for valuable discussions. We are grateful to Jane Edwards for preparing the manuscript and both Kirk Apt and Devaki Bhaya for their critical appraisal of the data. This manuscript is CIW publication No. 1077.

References

Alberte RS, Friedman AL, Gustafson DL, Rudnick MS, Lyman H (1981) Light harvesting systems of brown algae and diatoms. Isolation and characterization of chlorophyll a/c and chlorophyll a/fucoxanthin pigment protein complexes. Biochim Biophys Acta 635:304–316

Argos P, Rao JKM, Hargrave PA (1982) Structural predictions of membrane-bound proteins. Eur J Biochem 128:565–575

Barrett J, Anderson JM (1977) Thylakoid membrane fragments with different chlorophyll *a*, chlorophyll *c* and fucoxanthin compositions isolated from the brown seaweed *Ecklonia radiata*. Plant Sci Lett 9:275–283

Barrett J, Anderson JM (1980) The P-700-chlorophyll *a*-protein complex of *Acrocarpia paniculata* and other brown seaweeds. Biochim Biophys Acta 590:309–323

Berkaloff C, Caron L, Rousseau B (1990) Subunit organization of PSI particles from brown algae and diatoms: polypeptide and pigment analysis. Photosyn Res 23:181–193

Boczar BA, Prézelin BB (1986) Light and $MgCl_2$-dependent characteristics of four chlorophyll-protein complexes isolated from the marine dinoflagellate, *Glenodinium* sp. Biochim Biophys Acta 850:300–309

Caron L, Brown J (1987) Chlorophyll-carotenoid protein complexes from the diatom *Phaeodactylum tricornutum*: spectrophotometric, pigment and polypeptide analysis. Plant Cell Physiol 28:775–785

Caron L, Remy R, Berkaloff C (1988) Polypeptide composition of light harvesting complexes from some brown algae and diatoms. FEBS Lett 229:11–15

Cashmore AR, Broadhurst MK, Gray RE (1978) Cell free synthesis of leaf protein: Identification of an apparent precursor of the small subunit of ribulose-1,5-bisphosphate carboxylase. Proc Natl Acad Sci USA 74:5463–5467

Chitnis PR, Thornber P (1988) The major light-harvesting complex of photosystem II: aspects of its molecular and cell biology. Photosyn Res 16:41–63

Coleman JR, Grossman AR (1984) The biosynthesis of carbonic anhydrase in *Chlamydomonas reinhardtii* during adaptation to low CO_2. Proc Natl Acad Sci USA 81:6049–6053

Conley PB, Lemaux PG, Grossman AR (1985) Cyanobacterial light-harvesting subunits are encoded in two red light-induced transcripts. Science 230:550–553

Deisenhofer J, Epp O, Miki K, Huber R, Michel H (1985) Structure of the protein subunits in the photosynthetic reaction center of *Rhodopseudomonas viridis* at 3Å resolution. Nature 318:618–624

Fawley MW, Grossman AR (1986) Polypeptides of light-harvesting complexes of the diatom *Phaeodactylum tricornutum* are synthesized in the cytoplasm of the cell as precursors. Plant Physiol 81:149–155

Fawley MW, Morton JS, Steward KD, Mattox KR (1987) Evidence for a common evolutionary origin of light harvesting fucoxanthin chlorophyll a/c-protein complexes of *Pavlova gyrans* (Prymnesiophyceae) and *Phaeodactylum tricornutum* (Bacillariophyceae). J Phycol 23:377–381

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Friedman AL, Alberte RS (1984) A diatom light harvesting pigment protein complex. Purification and characterization. Plant Physiol 76:483–489

Friedman AL, Alberte RS (1986) Biogenesis and light regulation of the major light harvesting chlorophyll-protein of diatoms. Plant Physiol 80:43–51

Friedman AL, Alberte RS (1987) Phylogenetic distribution of the major diatom light-harvesting pigment proteins determined by immunological methods. J Phycol 23:427–433

Gantt E (1981) Phycobilisomes. Annu Rev Plant Physiol 32:327–347

Garnier J, Osguthorpe DJ, Robson B (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 120:97–120

Gibbs SP (1962) Nuclear envelope chloroplast relationships in algae. J Cell Biol 14:433–444

Gibbs SP (1970) The comparative ultrastructure of the algal chloroplast. Ann NY Acad Sci 175:454–473

Gibbs SP (1978) The chloroplast of *Euglena* has evolved from symbiotic green algae. Can J Bot 56:2883–2889

Gibbs SP (1979) The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER. J Cell Sci 35:253–266

Gierasch LM (1989) Signal sequences. Biochemistry 28:923–931

Glazer AN (1982) Phycobilisome structure and dynamics. Annu Rev Microbiol 36:173–198

Glazer AN (1985) Light harvesting by phycobilisomes. Annu Rev Biochem 14:47–77
Goedheer JC (1970) On the pigment systems of brown algae. Photosynthetica 4:97–103
Goedheer JC (1973) Chlorophyll *a* forms in *Phaeodactylum tricornutum*: a comparison with other diatoms and brown algae. Biochim Biophys Acta 314:191–201
Green B (1988) The chlorophyll protein complexes of higher plant photosynthetic membranes *or* Just what green band is that? Photosyn Res 15:3–32
Grossman AR, Lemaux PG, Conley PB, Bruns BU, Anderson LK (1988) Characterization of phycobiliprotein and linker polypeptide genes in *Fremyella diplosiphon* and their regulated expression during complementary chromatic adaptation. Photosyn Res 17:23–56
Guggliemelli LA, Dutton HJ, Jursinic PA, Siegelman HW (1981) Energy transfer in light-harvesting carotenoid-chlorophyll *c*-chlorophyll *a*-protein of *Phaeodactylum tricornutum*. Photochem Photobiol 33:903–907
Harrison PJ, Waters RE, Taylor FJR (1980) A broad spectrum artificial sea water medium for coastal and open ocean phytoplankton. J Phycol 16:28–35
Hiller G, Larkum AWD, Wrench P (1988) Chlorophyll proteins of the prymnesiophyte *Pavlova lutherii* (Droop): Identification of the major light harvesting complex. Biochim Biophys Acta 932:223–231
Hoffman NE, Pichersky E, Malik VS, Castresano C, Ko K, Darr SC, Cashmore AR (1987) A cDNA encoding a photosystem I protein with homology to photosystem II chlorophyll a/b-binding proteins. Proc Natl Acad Sci USA 84:8844–8848
Houlne G, Schantz R (1987) Molecular analysis of transcripts encoding the light harvesting chlorophyll a/b proteins in *Euglena gracilis*: unusual size of the mRNA. Curr Genet 12:611–616
Karlin-Neuman GA, Tobin EM (1986) Transit peptides of nuclear-encoded chloroplast proteins share a common amino acid framework. EMBO J 5:9–13
Karlin-Neumann GA, Kohorn BD, Thornber JP, Tobin EM (1985) A chlorophyll a/b-protein encoded by a gene containing an intron with the characteristics of a transposable element. J Mol Appl Genet 3:45–61
Keegstra K (1989) Transport and routing of proteins into chloroplasts. Cell 56:247–253
Kirk JTO (1977) Thermal dissociation of fucoxanthin protein binding in pigment complexes from chloroplasts of *Hormosira* (Phaeophyta). Plant Sci Lett 9:373–380
Kolubayev T, Geacintov NE, Paillotin G, Breton J (1985) Domain sizes in chloroplasts and chlorophyll-protein complexes probed by fluorescence yield quenching induced by singlet-triplet exciton annihilation. Biochim Biophys Acta 808:66–76
Kuhlbrandt W (1984) Three dimensional structure of the light-harvesting chlorophyll *a/b* protein complex from pea chloroplasts. Nature 307:478–480
Long Z, Wang S-Y Nelson N (1989) Cloning and nucleotide analysis of genes coding for the major chlorophyll binding protein of the moss *Physcomitrella patens* and the halotolerant alga *Dunaliella salina*. Gene 76:299–312
Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
Manodori A, Grossman AR (1990) Light harvesting genes of *Phaeodactylum tricornutum*. In: Baltscheffsky M (ed) Current research in photosynthesis, vol III. Kluwer Academic Publishers, Boston, pp 541–544
Mishkind ML, Scioli SE (1988) Recent developments in chloroplast protein synthesis. Photosyn Res 19:153–184
Murphy DJ (1986) The molecular organization of the photosynthetic membranes of higher plants. Biochim Biophys Acta 864:33–94
Owens TG (1986) Light harvesting function in the diatom *Phaeodactylum tricornutum* II. Distribution of excitation energy between the two photosystems. Plant Physiol 80:739–746
Owens TG, Wold ER (1986) Light harvesting function in the diatom *Phaeodactylum tricornutum*. Plant Physiol 80:732–738
Peter GF, Thornber JP (1988) The antenna components of photosystem II with emphasis on the major pigment protein, LHCIIb. In: Scheer H, Schneider S (eds) Photosynthetic light-harvesting systems. Walter de Gruyter & Co, Berlin, NY, pp 175–186
Peyriere M, Caron L, Jupin J (1984) Pigment complexes and energy transfer in brown algae. Photosynthetica 18:184–191
Pichersky E, Hoffman NE, Malik VS, Bernatzsky R, Tanksley SD, Szabo L, Cashmore AR (1987) The tomato Cab-4 and Cab-5 genes encode a second type of CAB polypeptide localized in photosystem II. Plant Mol Biol 9:109–120
Plumley FG, Schmidt GW (1984) Immunochemical characterization of a family of light-harvesting pigment protein complexes in several groups of algae. J Phycol Suppl 20:10 (Abstract No. 38)
Plumley FG, Schmidt GW (1987) Reconstitution of chlorophyll a/b light harvesting complexes: Xanthophyll-dependent assembly and energy transfer. Proc Natl Acad Sci USA 84:146–150
Prézelin BB, Alberte RS (1978) Photosynthetic characteristics and organization of chlorophyll in marine dinoflagellates. Proc Natl Acad Sci USA 75:1801–1805
Prézelin BB, Haxo FT (1976) Purification and characterization of peridinin-chlorophyll *a*-proteins from marine dinoflagellates *Glenodinium* sp. and *Gonyaulax polyhedra*. Plant 128:133–141
Ryrie IJ, Anderson JM, Goodchild DJ (1980) The role of light harvesting chlorophyll *a/b* protein complex in chloroplast membrane stacking. Eur J Biochem 107:345–354
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain termination inhibitors. Proc Natl Acad Sci USA 74:5463–5467
Schmidt GW, Mishkind ML (1986) The transport of proteins into chloroplasts. Annu Rev Biochem 55:879–912
Short JM, Fernandez JM, Sorge JA, Huse WD (1988) λZAP: a bacteriophage λ expression vector with in vivo excision properties. Nucleic Acids Res 16:7583–7600
Siefermann-Harms D, Ninneman N (1982) Pigment organization in the light-harvesting chlorophyll *a/b*-protein complex of lettuce chloroplast. Evidence obtained from protection of the chlorophylls against proton attach and from excitation energy transfer. Photochem Photobiol 35:719–731
Suss K-H (1983) A new isolation method and properties of the light-harvesting chlorophyll *a/b*-protein complex of higher plants. Photobiochem Photobiophys 5:317–324
Thornber JP (1975) Chlorophyll-proteins: light-harvesting and reaction center components of plants. Annu Rev Plant Physiol 26:127–158
Thornber JP, Markwell JP, Reinman S (1979) Plant chlorophyll-protein complexes: recent advances. Photochem Photobiol 29:1205–1216
Tobin E, Silverthorne J (1985) Light regulation of gene expression in higher plants. Annu Rev Plant Physiol 36:569–593
von Heijne G (1985) Signal sequences. The limits of variation. J Mol Biol 184:99–105
von Heijne G (1986) A new method for predicting signal sequence cleavage sites. Nucleic Acids Res 14:4683–4690
Zuber H (1985) Structure and function of light-harvesting complexes and their polypeptides. Photochem Photobiol 42:821–844

Communicated by E. Meyerowitz