AN ANALYSIS OF THE FUCOXANTHIN-CHLOROPHYLL PROTEINS AND THE GENES ENCODING THEM IN THE UNICELLULAR MARINE RAPHIDOPHYTE, Heterosigma carterae: CHARACTERIZATION AND EVOLUTION

by

DION GLENN DURNFORD

B.Sc. (Biology) Dalhousie University 1989

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF BOTANY

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
June 1995
©Dion Glenn Durnford, 1995

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of _BOTANY_

The University of British Columbia
Vancouver, Canada

Date _7-27-95_

ABSTRACT

The light-harvesting complexes (LHC) of the unicellular marine chromophyte,
Heterosigma carterae, were fractionated by sucrose-density gradient centrifugation, following
digitonin solubilization, and by non-denaturing SDS-PAGE. The sucrose gradient allowed for
the isolation of a major light-harvesting complex fraction, containing approximately 53% of
the total chlorophyll, the majority of the chlorophyll *c* and a single polypeptide of 19.5 kDa.
Up to 12 different polypeptides immunologically related to both the fucoxanthin-Chl *a/c*
complexes (FCPs) and to the chlorophyll *a* + *b*-binding proteins (CABs) were detected in
thylakoids and in the lower photosystem I (PS I) enriched fractions. Using a modification of
the non-denaturing gel system of Allen and Staehelin (1991 Anal. Biochem. 194, 214-222)
allowed the resolution of a number of large pigment-protein complexes which included several
PS I and PS II fractions along with a predominant LHC fraction, an improvement over
previously published methods.

A *Fcp* cDNA from *Heterosigma carterae* has been cloned and sequenced. It encodes a
210 amino acid polypeptide that has similarity to other FCPs and to the CABs of terrestrial
plants and green algae. Comparison of the FCP sequence to the recently determined 3-
dimensional structure of the pea LHC II complex indicates that many of the key amino acids
thought to participate in the binding of chlorophyll and in the formation of complex-stabilizing
ionic interactions between hydrophobic regions of the protein are well conserved. In addition,
the *Fcp* genes are part of a large multigene family with greater than 20 related members in
Heterosigma. Phylogenetic analyses of the LHC protein sequences shows that the FCPs form a
natural group separate from the *i*PCPs of the dinoflagellates. Though there are obvious
similarities between the FCPs and the CABs, the relationships are very distant.

Analyses of polypeptides in the red algae *Aglaothamnion neglectum* and *Porphyridium
cruentum*, in collaboration with Greg Wolfe and Beth Gantt, were the first to demonstrate that
polypeptides immunologically related to the CABs and the FCPs are present within the
Rhodophyceae. In addition, CAB/FCP-related LHCs have not been detected in a

cyanobacterium (*Nostoc*) and a prochlorophyte (*Prochlorothrix*). This suggests the CAB/FCP-related LHCs arose only once after the establishment of the chloroplast and provides some evidence that suggests chloroplasts evolved from a symbiotic cyanobacterium-like organism only once (monphyletic).

The organization of the antennae in *Heterosigma carterae* is equally as complex as that in the terrestrial plants. This is indicated by the detection of at least 12 LHC-related polypeptides and the presence of a large multigene family encoding the FCPs. In addition, the immunological relatedness and the sequence conservation of the FCPs with the CABs indicates that the structure of the LHCs has been conserved throughout evolution and that these different antennae complexes share a common ancestor.

TABLE OF CONTENTS

CHAPTER 1   General introduction

CHAPTER 2   *Characterization of the light-harvesting proteins from Heterosigma carterae*

CHAPTER 3   *An immunological characterization of LHC related-polypeptides in red algae*

CHAPTER 4 *Characterization of the* Fcp *cDNAs of* Heterosigma carterae

CHAPTER 5 *Phylogenetic analysis of the LHCs*

LIST OF TABLES

AP   Allophycocyanin

bp   base pairs

CAB   chlorophyll a + b-binding protein

Chl   chlorophyll

FCP   fucoxanthin-chlorophyll protein

kb   kilobase

kDa   kiloDalton

LHC   light-harvesting complex

MSR   membrane spanning region

PBS   phycobilisome

iPCP   *intrinisic* peridinin-chlorophyll-protein

sPCP   *soluble* peridinin-chlorophyll protein

PC   phycocyanin

PE   phycoerythrin

PS I   photosystem I

PS II   photosystem II

SDS-PAGE   sodium dodecyl sulfate-polyacrylamide gel electrophoresis

ACKNOWLEDGMENTS

To my parents

Gordon and Sarah Durnford

CHAPTER 1

Introduction

The goal of this study was primarily to characterize the light harvesting antennae from the Raphidophycean alga, *Heterosigma carterae*. This involved characterization of the fucoxanthin-chlorophyll protein (FCP) complexes, cloning of the genes encoding them, and comparing them to the chlorophyll *a* + *b*-binding (CAB) family of proteins. The eukaryotic algae are diverse and have been studied very little in regards to their photosynthetic structure and function. At the start of this study, the antennae from only three main groups of chromophytes had been examined in any detail: the brown algae, the diatoms and the dinoflagellates (discussed in section 1.5.2). The structural relatedness of the FCPs to the CABs had not been demonstrated until the first FCP gene from a diatom was sequenced (Grossman et al. 1990).

Heterosigma carterae was chosen for this study for several reasons. First, the LHCs from a member of this class of algae had not been previously examined. Given the diversity amongst the chromophytes (see section 1.3 and Table 1.1), I thought it would be useful to examine a different representative from one of the other major algal taxa in order to assess the evolutionary relationships amongst the light-harvesting antennae. Second, like other chromophytes, the presence of four membranes surrounding the chloroplast in *Heterosigma* (see section 1.3) indicated that nuclear encoded, chloroplast localized proteins probably have a different mechanism for the targeting and translocation of these proteins into the chloroplast. Since a chloroplast targeted-nuclear encoded gene had been determined from only a single chromophyte (the diatom FCP), I was interested in examining the transit sequence of the *Heterosigma* FCP sequence to get an idea as to possible mechanisms of chloroplast targeting. Third, the chloroplast genome from *Heterosigma carterae* (published under the name of *Olisthodiscus luteus*) had been

1

well characterized and a few chloroplast genes had been sequenced by Rose-Anne Cattolico's lab. In addition, easily maintainable, axenic cultures were available.

1.1   *Heterosigma carterae*: characteristics and taxonomy

Heterosigma carterae is a unicellular alga that is approximately 11-20 μm long, 9-12 μm wide, has anywhere from 9-25 yellowish-brown chloroplasts and possesses two flagella (see Fig. 1.1; Hulburt 1965; Hara and Chihara 1987). The chloroplasts have four surrounding membranes; two chloroplast envelope membranes and two membranes of the chloroplast E.R. (CER). The outer CER membrane is not continuous with the nuclear envelope, as occurs with some chromophytes (Gibbs 1981). The chloroplasts are also distributed along the periphery of the cell. *H. carterae* possesses Chl $a + c$ along with abundant amounts of fucoxanthin (74% of total carotenoid) followed by significant levels of diatoxanthin (14%) and β-carotene (12%) (Riley and Wilson 1967). *H. carterae* was originally grouped with the xanthophytes due to its yellowish colour but ultrastructure analyses and other data indicates that it is a member of the Raphidophyceae (Hara et al. 1985). *H. carterae* is widely distributed in coastal marine habitats. It is an important component in red tides and has been implicated in fish kills, causing the loss of millions of dollars within the aquaculture industry of British Columbia (Taylor 1993).

There is confusion in the literature regarding the taxa referred to as *Olisthodiscus luteus* Carter, *Olisthodiscus carterae* Hulburt (Hulburt 1965) and *Heterosigma akashiwo* Hada. Morphological and ultrastructural studies on *Heterosigma akashiwo* Hada and on various culture collections identified as *Olisthodiscus luteus* Carter indicated that they were all very similar and could be combined under the genus, *Heterosigma* (Hara and Chihara 1987). However, these collections were different from the originally described *O. luteus* Carter culture (Hara et al. 1985) and appear to be the same as *O. carterae*, as described by Hulburt (1965). As this appears to be the case, it has been suggested that the genus *Heterosigma* be retained and the previous species name (*carterae*) be used (Taylor 1992). A number of studies have been published on the chloroplast genome of *H. carterae* (e.g. Reith and Cattolico 1986) and some physiological work

2

on this alga has been done. However, most of this work has been published under the name *O. luteus* due to a misidentification of the isolates in a number of culture collections (Taylor and Haigh 1993).
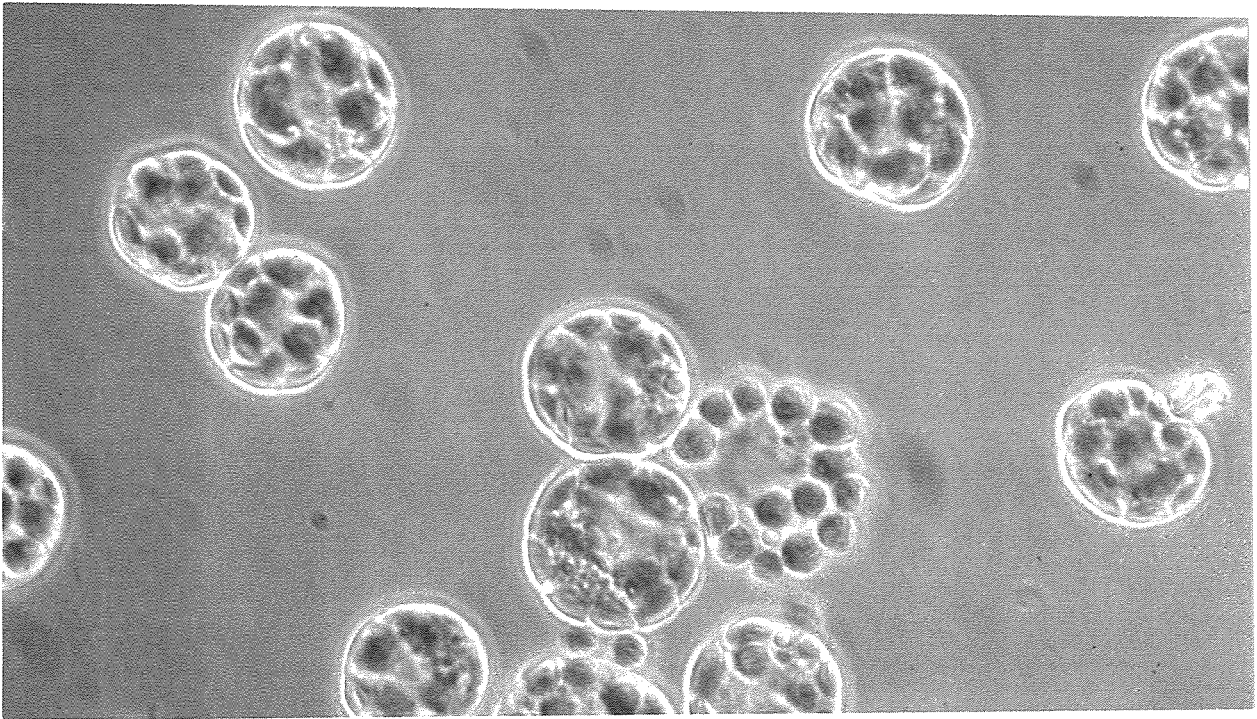
Figure 1.1

Photograph of *Heterosigma carterae* cells. In order to give a clearer view of the number of chloroplasts, the prepared slide was allowed to dry partially to cause the cells to flatten.

1.2   Photosynthesis—an overview

Cyanobacteria, algae and terrestrial plants utilize a chlorophyll a based system for the light reactions of photosynthesis. In these organisms there are two membrane integral photochemical reaction centers: photosystem I (PS I) and photosystem II (PS II). These complexes, along with the membrane soluble plastoquinones (PQH-PQH_2), the cytochrome b_6f complex and plastocyanin (Pc) are responsible for the non-cyclic electron transfer. This electron transport mechanism through the two photosystems is often referred to as the Z scheme and is illustrated in Figure 1.2 (Hill and Bendel 1960). The purpose of the process is for the production of the reducing agent NADPH and for the generation of an electrochemical gradient through the net transfer of protons from the stroma into the lumen. This gradient is then utilized for the generation of ATP via the thylakoid membrane bound ATP synthase (ATPase) (Fig. 1.2). NADPH and ATP are used for the fixation of carbon dioxide via the carbon reduction (Calvin) cycle, for the production of carbohydrate, and for many other cellular reactions. In addition, there is also a form of cyclic electron flow around PS I and the cytochrome b_6f complex mediated by the ferredoxin NADP-reductase (FNR), as indicated by the dashed line in Figure 1.2

Stroma

NADP + H$^+$ → NADPH   ADP + P   H$^+$ ATP

FNR

Fd

LHCII   H$^+$ LHCI   F$_A$/ F$_B$

Q$_A$ → Q$_B$ → PQH$_2$ F$_X$

Ph   PQ pool   A$_1$

A$_0$

D$_1$ D$_2$

P680   PQ   P700

Mn   Pc

OEC   H$^+$

2H$_2$O   4H$^+$ + O$_2$ H$^+$

Photosystem II   Cytochrome $b_6 f$ Photosystem I   ATPase

Lumen
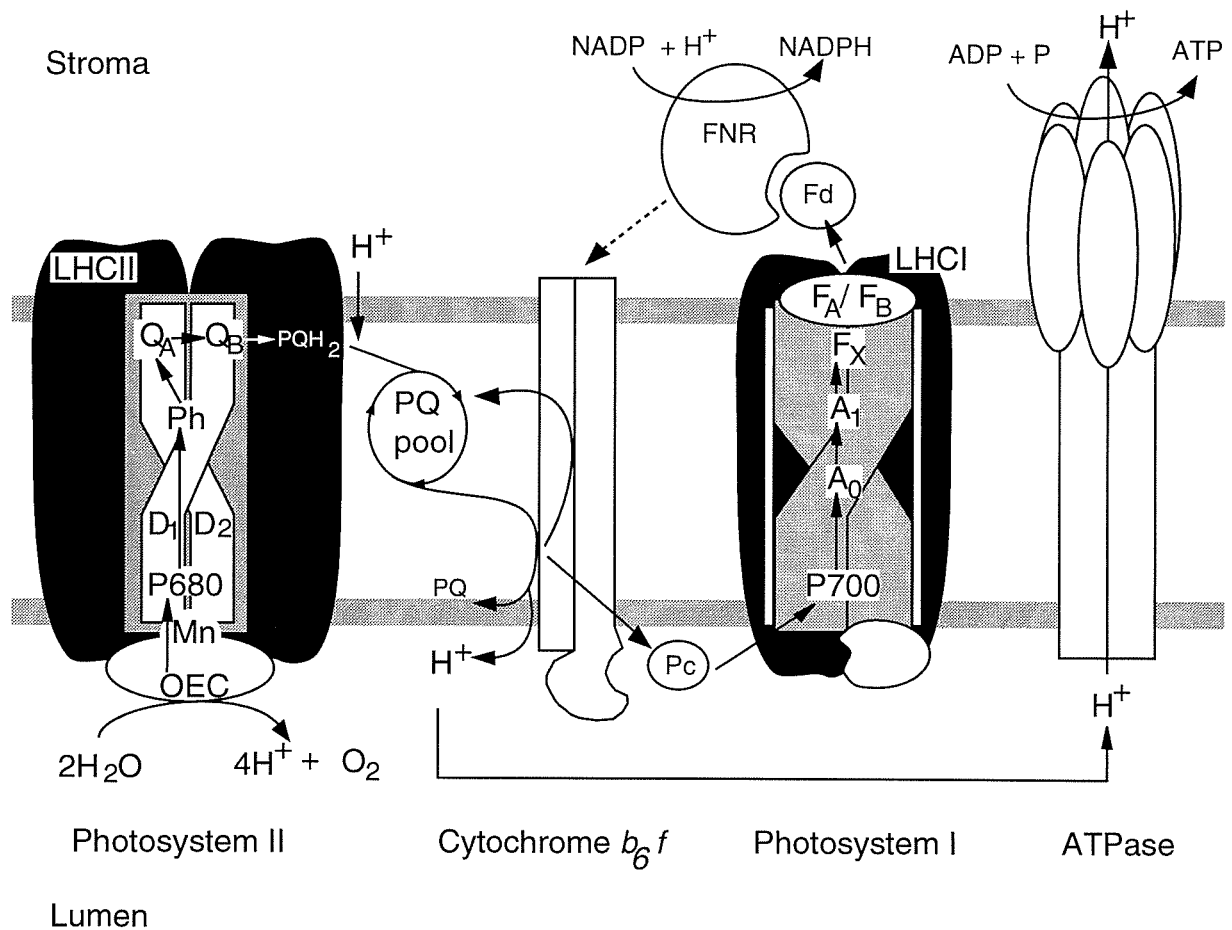
Figure 1.2

Thylakoid protein complexes involved in photosynthetic electron transport and ATP generation. The complexes include photosystem II, the cytochrome $b_6 f$ complex, photosystem I, and the ATP synthase complex. Black filled areas represent the peripheral antennae complexes of PS I and PS II. Grey filled areas represent the core antennae of the reaction center (PS II) or the chlorophyll binding reaction center complex of PS I.

5

1.3   Chloroplast characteristics and thylakoid ultrastructure of the major algal taxa

Many of the plastid and cytosolic features of the main algal groups are summarized in Table 1.1. This was included in order to emphasize the wide diversity amongst the algae and to provide the necessary background that will allow one to assess the relationships between them. In this section I will briefly compare and contrast some of the key traits of the different algal groups.

1.3.1   Terrestrial plants/green algae

The terrestrial plant and green algal chloroplast structure is well known. There are two membranes of the chloroplast envelope surrounding a network of thylakoid membranes that form a unit with a single connecting lumenal space. The two chloroplast envelope membranes are not equivalent. The outer membrane is more permeable to low molecular mass substances and the inner chloroplast envelope membrane contains 5-10 times more intramembrane particles (Staehelin 1986). The thylakoids form both appressed (stacked, grana lamellae) and nonappressed (unstacked, stroma lamellae) regions. The grana lamellae can consist of a few or many stacked thylakoids, depending on the conditions and the species. The external surfaces of nonappressed thylakoids are directly exposed to the stroma. The thylakoid components are nonrandomly distributed between the appressed and nonappressed regions (discussed in section 1.4). The green algae and terrestrial plants are unique in that starch (amylose, amylopectin—an α1-4 glucan) is stored inside the chloroplasts.

1.3.2   Red algae

Thylakoids of the red algae, like those of the cyanobacteria, have a parallel arrangement of unstacked thylakoids. However, in some taxa the thylakoids are concentrically arranged, form a close association with the inner envelope of the chloroplast, and/or have one or several girdle thylakoid bands surrounding the internal parallel thylakoids (Staehelin 1986; Mörschel and Rhiel 1987; Pueschel 1990). All red algae have the main antennae, phycobilisomes, attached to the

6

outer surfaces of the thylakoid membrane. Accessory chlorophylls, equivalent to Chl *b* or Chl *c*, are not present in the red algae. At one time, Chl *d* was thought to be an accessory chlorophyll in the red algae, though it had only been observed in *Gigartina* (Gigartenalies) extracts and had not been shown to exist *in vivo* (Holt 1966). If Chl *d* is not an artifact of preparation, it is in a very low concentration and will not make a significant contribution to the action spectrum (Gantt 1990). Red algae store various forms of starch (α1,4-glucan) outside the chloroplast, in the cytoplasm.

1.3.3   Glaucophytes

The glaucophytes are a group of unicellular algae that possess inclusions referred to as cyanelles, which are considered by some to be modified cyanobacteria functioning as plastids. However, the cyanelles have a reduced genome compared to free-living cyanobacteria and are dependent upon the 'host'. In this regard, they resemble plastids. Most, but not all, contain a rudimentary cell wall made up of peptidoglycan surrounding the cyanelle (for a review see Kies and Kremer, 1990). This group possesses phycobilisomes, like the red algae, and the thylakoids in the cyanelle are unstacked and usually concentrically arranged. The photosynthate reserve, starch, is stored in the cytoplasm.

1.3.4   Euglenophytes

Euglenophytes comprise a large group with both photosynthetic (approximately 1/3) and nonphotosynthetic representatives. The photosynthetic taxa contain Chl's *a* + *b* and have appressed thylakoid membrane regions in bands of three to many, though the thylakoids never form grana like those in terrestrial plants (Gibbs 1970). Euglenophytes have a chloroplast envelope with three membranes. The outer most membrane of the chloroplast envelope does not bind ribosomes (Gibbs 1978). Interestingly, the main xanthophylls in this group of algae—diadinoxanthin and diatoxanthin—are more typical of the chromophyte algae than the Chl *a* + *b*-containing algae and terrestrial plants. Euglenophytes store paramylon (a β1-3 linked glucan) as a reserve in crystalline granules outside the chloroplast.

1.3.5 Dinoflagellates

The dinoflagellates are considered chromophyte algae because of the presence of Chl c_2; however, there are also non-photosynthetic taxa. The primary xanthophyll, peridinin, gives these organisms the reddish-brown colour associated with red tides. Like the euglenophytes, some of the dinoflagellates have a total of three membranes surrounding the chloroplast, the outer (third) membrane lacking bound ribosomes on its cytoplasmic surface. The thylakoids are usually arranged in three appressed bands, resembling those of the euglenophytes . They differ from the thylakoids of other chromophytes by having a reduced diameter of the appressed regions and by the lack of a surrounding thylakoid band (girdle thylakoid) (Staehelin 1986). Dinoflagellates store an α1-4 glucan outside the chloroplast.

1.3.6 Heterokonts/Haptophytes

The brown algae, diatoms, chrysophytes, xanthophytes, other heterokonts, and the haptophytes possess a total of four membranes surrounding the chloroplast. The outer two membranes are collectively referred to as the chloroplast ER (CER) due to the presence of ribosomes bound to the cytoplasmic side of the outermost membrane. The outer membrane of the CER tends to be continuous with the nuclear envelope when the number of chloroplasts are low (1 or 2). Frequently, a girdle thylakoid of three appressed membranes surrounds the internal thylakoid membranes (Staehelin 1986), though this does not occur in the haptophytes (Gibbs 1970). The internal thylakoid membranes usually form appressed regions of three bands. The heterokonts and haptophytes typically store a β1-3 glucan outside the chloroplast, within the cytoplasm.

1.3.7 Cryptophytes

The cryptophytes are unique in that within the space between the CER and the chloroplast envelope there is a putative vestigial nucleus (the nucleomorph) . This is thought to be a remnant from a permanent endosymbiosis of a heterotrophic organism with a photosynthetic eukaryote (Ludwig and Gibbs 1987). The thylakoids are usually in pairs and appear thicker than other thylakoid membranes (Gibbs 1970). Interestingly, the cryptophytes contain the phycobiliproteins, phycoerytherin or phycocyanin, within the thylakoid lumen (Spear-Bernstein and Miller 1989). Cryptophytes store starch (an α1-4 glucan) in the area between the CER and chloroplast envelope.

Table 1.1   Characteristics of the major algal groups

Taxon	Plastid Characteristics					
	Chl's and PBP[1]	main xanth.[2]	carotene[2]	thylakoid lamellae	girdle lamellae	# chlp membr's
Cyanophyta	*a* pc,pe apc,pec	z	β	1	x	na
Prochlorophyta	*a, b, c*	z	β	3+	x	na
Glaucophyta [3]	*a, pc, apc*	z, c_2	β	1	x	2, *pep f*
Rhodophyta	*a* pc, pe, apc	z, n, l	α,β	1	x √	2
Chlorophyta a)Chlorophyceae b)Prasinophyceae	*a, b* *a, b* Mg 2,4D	n, z, l, v_1 n, z, l, a_1	α, β α, β	3+ 3+	x x	2 2
Dinophyta	*a, c_2*	p, d_3 (f)	β	3	x	3¤ (2)
Euglenophyta	*a, b*	d_1, d_2, n	α, β	3+	x	3¤
Chlorarachnida [4]	*a, b*	?	?	1-3	x	4¤ nm†
Cryptophyta	*a, c_2,(c_1)* pe*, pc*	a_2, c,	α, β, γ	2	x	4 nm†
Raphidophyta	*a,c_1,c_2*	I) d_1, d_3, n II) f, d_1, d_2	β, (γ)	3	√ (x)	4
Chrysophyta	*a,c_1,c_2* *(c_3)*	f, z, a (d_1, d_2)	β, (α, γ)	3	√	4
Synurophyta	*a,c_1*	f	β, (α, γ)	3	√	2 (4)
Bacillariophyta	*a,c_1,c_2,c_3*	f, d_1, d_2	β, (γ)	3	√	4
Xanthophyta	*a,c_1,c_2*	d_1, d_2, v_2	β	3	√	4
Phaeophyta	*a,c_1,c_2*	f, v_1 (z, d_1, d_2)	β, (γ)	3	√	4
Eustigmatophyta	*a*	v_1, v_2	β	3	x	4
Haptophyta	*a,c_1,c_2,c_3*	f, d_1, d_2	β, (α, γ)	3	x	4

10

Table 1.1 continued.

Taxon	mito. char. type of cristae	flagellar features number/ type	flagellar features features	storage products type/ location	alternate names
Cyanophyta	na	x	x	na	cyanobacteria blue green algae
Prochlorophyta	na	x	x	na	
Glaucophyta	F	2	mastig.	α/O	Glaucophyceae
Rhodophyta	F	x	x	α/O	red algae
Chlorophyta a)Chlorophyceae b)Prasinophyceae	F F	2 (4) (M) 2 (1,4, 8)	(hairs) scales	α/I α/I α/I	green algae micromonads
Dinophyta	T	2 A	hairs	α/O	dinoflagellates
Euglenophyta	D	2 (4) A	hairs	β/O	Euglenoids
Chlorarachnida	T	1	hairs	α/B	green amoeba
Cryptophyta	F T	2 A	mastig.	α/B	cryptomonads
Raphidophyta	T	2 H	mastig.	?/O	chloromonads golden-brown algae
Chrysophyta	T	2 H	mastig.	β/O	chrysomonads golden-brown algae
Synurophyta	T	2 H	mastig. scales	β/O	often grouped with Chrysophytes
Bacillariophyta	T	1	mastig.	β/O	diatoms
Xanthophyta	T	2 H	mastig.	β/O	yellow-green algae
Phaeophyta	T	2 H	mastig.	β/O	brown algae
Eustigmatophyta	T	1 (2) (H)	mastig.	?/O	
Haptophyta	T	2 Ha	smooth	β/O	prymnesiophytes

11

Key to abbreviations/explanations for Table 1.1

heading	description
plastid features:	
Chl's/PBP	-Chlorophyll's and phycobiliproteins present. a, Chl a; b, Chl b; c_{1-3}, Chl c_{1-3}; pc, phycocyanin; pe, phycoerytherin; apc, allophycocyanin; pec, phycoerythrocyanin
main xanth.	prominant xanthophylls present in order of relative abundance: a, anteroxanthin; a_2, alloxanthin; c_1, Crocoxanthin; c_2, β-cryptoxanthin; d_1, diadinoxanthin; d_2, diatoxanthin; d_3, dinoxanthin; f, fucoxanthin; l, lutein; p, peridinin; v_1, violaxanthin; n, neoxanthin; v_2, vaucheriaxanthin; z, zeaxanthin
carotene	-main carotenoid present: α, β,ε-carotene; β, β,β-carotene; γ, ε,ε-carotene
thylakoids	- number of appressed or loosely appressed thylakoid membranes
girdle lamellae	-presence ($\sqrt{}$) or absence (x) of a surrounding thylakoid membrane
# chlp membr	-number of membranes surrounding the chloroplast: pep, petidoglycan wall present
Mitochondrial Features:	
type of cristea	-F, flat cristae; T, tubular cristae; D, discoidal cristae
Flagellar Features:	
number/ type	-M, many; A, anisokont (flagella of unequal length); H, heterokont (organism with a hairy and smooth flagellum); Ha, haptonema
features	-hairs, fibrous flagellar hairs; mastig., mastigonemes or tripartite, tubular hairs.
Storage Products:	
type/ location	-Type of stored carbohydrate: α, α1-4 glucan; β, β1,-3 glucan: Location of stored reserve; I, inside chloroplast; O, outside chloroplast; B, between chloroplast envelope and chloroplast periplasmic membrane

Other symbols used and notes

()   occasional occurrence/reports
*   located in thylakoid lumen
†   nucleomorph (nm) present
¤   no ribosomes bound to outer membrane
f   Glaucophyta contain a cyanelle and not a "true" chloroplast
na   not applicable
?   unknown
x   absent
$\sqrt{}$ present

References

[1] (Jeffery 1989)
[2] (Bjorland and Liaaen-Jensen 1989)
[3] (Kies and Kremer 1990)
[4] (Hibberd and Norris 1984)

This table is a modification of tables found in (Sleigh 1989) and (Lee 1989).

1.4   Main photosynthetic components of the thylakoid membrane

The main complexes involved in the electron transport process are PS II, the cytochrome b_6f complex (Cyt b_6f), and PS I. An ATPase is also present within the thylakoid membrane and it produces ATP using the proton gradient generated by the light reactions. These complexes are not equally distributed throughout the thylakoid membrane of the Chl $a + b$ containing organisms. In many cases, the complexes are preferentially localized to either appressed or nonappressed thylakoid regions, referred to as lateral heterogeneity. In terrestrial plants, PS II-LHC II complexes are primarily located in the appressed regions of the thylakoid (PS II$_\alpha$), though some PS II complexes are found in the unstacked sections and have a smaller cross sectional absorbance (PS II$_\beta$). PS I is primarily found in the nonappressed (stroma) regions along with the ATP synthase complex. The Cyt b_6f complex, on the other hand, is equally distributed in either region (Staehelin 1986). LHC II is found mainly in the appressed regions of the thylakoid, though it can be found in the nonappressed domains in a phosphorylation dependent manner. This reversible association of LHC II with PS II (appressed region) is believed to mediate the distribution of excitation energy between the two photosystems and to be important in the adaptation of the organism to varying light conditions (Allen 1992).

In the chromophyte algae, lateral heterogeneity of the thylakoid membrane complexes, similar to that found in the terrestrial plants, does not occur. In the brown algae (Lichtlé et al. 1992b), diatoms (Pyszniak and Gibbs 1992) and cryptomonads (Lichtlé et al. 1992) the Chl $a + c$-binding proteins were demonstrated to be homogeneously distributed throughout the thylakoid membrane, in both appressed and non-appressed regions, by immunolocalization. PS I complexes, however, were slightly enriched in the nonappressed regions of the chromophyte thylakoids but this distribution did not resemble the almost exclusive localization of PS I in the nonappressed regions of the terrestrial plant and green algal thylakoids. Moreover, the thylakoid membranes of the chromophytes are typically associated in groups of three loosely appressed thylakoids; grana stacks similar to those found in terrestrial plants are not present. The homogeneous localization of PS I and the antennae in chromophytes is in agreement with a study

13

in diatoms that shows energy absorbed by the main LHC complex is distributed equally to PS I and PS II (Owens 1986b). A role of light dependent phosphorylation in the regulation of excitation energy between the complexes has not been demonstrated. It is apparent that the mode of adaptation to light in the chromophytic algae is different from that in the terrestrial plants. A brief discussion of the characteristics of each major complex in the thylakoid membrane is given below.

1.4.1   *Photosystem II*

The electron transfer reactions are initiated when an electron from a special chlorophyll a (P680), which is probably a dimer, is excited to the singlet state by light. The donation of an electron from the excited P680 to pheophytin (Ph) results in a charge separation within the reaction center that is the driving force for the oxidation of water. This occurs via a manganese cluster (Mn) and is enhanced by the presence of an oxygen evolving complex (OEC). A series of oxidation state changes in the manganese cluster, referred to as the water oxidation clock, eventually result in the oxidation of H_2O (Rutherford 1989). A redox active tyrosine residue on D1 mediates the transfer of electrons from the Mn cluster to $P680^+$. This process results in the reduction of $P680^+$ and the liberation of 4 protons (H^+) and oxygen within the lumen. P680 is then able to undergo another round of light induced charge separation in the reaction center. From pheophytin, the electron is donated to the first bound plastoquinone, Q_A (on D2), then to the second, Q_B (on D1). An associated non-heme iron molecule may influence the stability of the quinones.

The PS II complex of terrestrial plants consists of a reaction center and a core antennae surrounded by more peripheral antennal complexes. The reaction center is a heterodimer consisting of the D1 and D2 polypeptides, cytochrome b_{559} (Cyt b_{559}) and the psbI protein. Both D1 and D2 are predicted to have five transmembrane helicies with a molecular mass of 32 and 34 kDa, respectively. The D1 and D2 polypeptides are related to the M and L subunits of the purple bacterial reaction center and, by analogy, are thought to associate in a similar fashion (Michel and Deisenhofer 1986; Trebst 1987). Since the determination of the 3-D structure of the

bacterial reaction center (Deisenhofer et al. 1985), it has been accepted that the core polypeptides (D1 + D2) bind the primary reactants involved in the initial light-induced charge separation, as discussed above. Evidence that D1 and D2 form the photosynthetic reaction center was provided through the isolation of a photochemically active core complex including these polypeptides and two hydrophobic polypeptides (4 and 9 kDa) of Cyt b_{559} (Nanba and Satoh 1987). The subunits of the Cyt b_{559} complex probably assemble in a heterodimeric fashion and cooperatively bind the heme molecule via a conserved His residue in each (Tae and Cramer 1994). The function of Cyt b_{559} is not known at the present time and its presence is a major difference between PS II and the bacterial reaction center. In addition, there are a number of other low molecular mass polypeptides associated with the PS II core reaction center (Hansson and Wydrzynski 1990).

Also associated with the PS II complex are the core antennae CP43 and CP47. These antennae bind only Chl a and have fluorescence emission peaks at 685 and 695 nm, respectively. Though not in general agreement, it has recently been determined that CP43 binds 20 Chl a and 5 β-carotene while CP47 binds 21-22 Chl a and 4 β-carotene molecules (Alfonso et al. 1994). These complexes are responsible for the coupling of the energy transfer from the peripheral antennae to the reaction center. In addition, lumen exposed regions of CP43 and CP47 may be important for the water splitting process as determined through inactivation and mutagenesis studies (see Vermaas et al., 1993).

In terrestrial plants and green algae there are three major extrinisic polypeptides associated with PS II on the lumenal side of the thylakoid membrane which influence the properties of O_2 evolution. They have apparent molecular masses of 33 (OEC1, *psb*O), 23 (OEC2, *psb*P) and 17 kDa (OEC3, *psb*Q). These polypeptides are not directly involved in oxygen evolution though they probably have a structural or regulatory function. In cyanobacteria, OEC1 is not essential for O_2 evolution though OEC1 deletion mutants are more sensitive to photoinhibition (Mayes et al. 1991). In contrast, *Chlamydomonas* mutants with low expression of OEC1 lack O_2 evolution capabilities and have low PS II stability (Mayfield et al. 1987). Despite these differences, which may be related to the organization of the complexes in the different organisms, it is generally agreed that OEC1 may help to stabilize the Mn cluster

(Ghanotakis and Yocum 1990; Vermaas et al. 1993). OEC2 and OEC3, in combination with OEC1, may be involved in concentrating specific ions near the catalytic site for enhanced water oxidation properties. Interestingly, homologues of the OEC2 and OEC3 complexes are not found in the cyanobacteria (Stewart et al. 1985). However, OEC1 is present and is immunologically related to the OEC1 of terrestrial plants.

1.4.2 Cytochrome $b_6 f$ complex and plastocyanin

As the Q_B site on D1 becomes reduced twice ($PQ \rightarrow PQH_2$), the plastoquinone diffuses from its binding site and becomes part of the plastoquinone "pool". An oxidized plastoquinone replaces the reduced molecule in order to continue the electron transfer process. The transfer of the electrons from PS II to PS I is mediated by the thylakoid membrane intrinsic cytochrome $b_6 f$ complex (Cyt $b_6 f$) and by the soluble plastocyanin (or cytochrome c_{553}) protein. The Cyt $b_6 f$ complex oxidizes plastoquinone and subsequently reduces plastocyanin in the non-cyclic electron transfer pathway. It is also involved in the process of cyclic-electron flow around PS I. There is a net transfer of two protons from the chloroplast stroma to the thylakoid lumen for every electron donated to the Cyt $b_6 f$ complex. The mechanism of plastoquinone oxidation is not fully understood but is thought to occur via a 'Q-cycle' where oxidation of the plastoquinone occurs in a two step process (Malkin 1992). One electron is transferred to the Rieske iron-sulfur center → cytochrome f → plastocyanin pathway. The second electron is passed to the two b cytochromes in succession which, after two turns of the cycle, reduces a quinone. The net result after two cycles is the transfer of four protons into the lumen, the transfer of two electrons to plastocyanin, the oxidation of two plastoquinones and the reduction of a single plastoquinone (Malkin 1992). The purpose of the electron transfer process is in the generation of a proton gradient that is used in the formation of ATP via a thylakoid membrane ATPase (Fig. 1.2).

The Cyt $b_6 f$ complex is composed of four main polypeptides in both spinach and cyanobacteria. These include cytochrome f (34 kDa), the 23 kDa polypeptide of cytochrome $b6$, the 20 kDa polypeptide of the Rieske FeS-protein and a 17 kDa subunit (subunit IV) (Hauska 1986). Cytochrome f (Cyt f) has a membrane spanning region, anchoring the subunit to the

16

thylakoid membrane, and a lumen exposed portion that provides a heme binding site. It is a basic domain on Cyt f that is thought to interact with a acidic domain on plastocyanin to allow electron transfer to occur between the two components (Cramer et al. 1994). As well, electron transfer to plastocyanin is thought to be mediated by a Tyr residue near the heme molecule of Cyt f. Plastocyanin is a small copper-binding protein (≈ 10 kDa) that is located in the thylakoid lumen and functions as an electron carrier between the Cyt $b_6 f$ complex and PS I, where it reduces $P700^+$. Plastocyanin has been found in all terrestrial plants and is present in many algae. However, in some algae, cytochrome c_{553} can replace plastocyanin.

1.4.3   Photosystem I

A second light induced charge separation occurs at PS I and starts with the photo-oxidation of P700, which is probably a chlorophyll a dimer (Golbeck and Bryant 1991). The electron released in this process follows a linear array of acceptors/donors starting with A_O, which is a Chl a molecule. The electron is then donated to A_1 (vitamin K_1) and to F_X (a 4Fe-4S cluster). The exact order of electron transfer from F_X to F_A (the second 4Fe-4S cluster) or to F_B (the third 4Fe-4S cluster) (reviewed by Golbeck and Bryant 1991) has not been determined conclusively. The presence of three 4Fe-4S clusters arranged in a triangular fashion has been confirmed by the 3-D structure of PS I at a resolution of 6Å (Krauss et al. 1993). The terminal electron acceptor is the soluble iron-sulfur protein, ferredoxin (Fd, Fig. 1.2). This protein is present in the stroma of the chloroplast and interacts with the soluble Ferredoxin-$NADP^+$ Reductase (FNR) that reduces $NADP^+$ to NADPH.

The primary reactants P700, A_O, A_1 and F_X are located on the core complex of PS I (CP1), which is composed of a heterodimer with polypeptides of approximately 83 (psaA) and 82 kDa (psaB) (Golbeck and Bryant 1991; Krauss et al. 1993). However, CP I usually migrates at 60-70 kDa on SDS-polyacrylamide gels, probably due to the hydrophobic nature of the complex (Green 1988).

The two iron-sulfur centers, F_A and F_B, are bound to PS I subunit VII (psaC) which is closely associated with CP1 on the stromal side of the membrane. This 9 kDa protein is highly

17

conserved between cyanobacteria and terrestrial plants. There are many other smaller, non-pigmented polypeptides, in the size range of 4-22 kDa (PsaD→N in plants), associated with PS I. The functions of many of these subunits have not been elucidated. However, the stroma located subunits, PsaD (22 kDa) and PsaE (10 kDa), have been shown to be closely associated with each other and with PsaC by cross-linking and reconstitution studies (Golbeck 1992). They are thought to be important in the 'docking' of ferredoxin and possibly for the binding and orientation of the PsaC subunit. Furthermore, PsaE has been suggested to be important for cyclic electron flow around PS I in cyanobacteria (Yu et al. 1993). PsaF, located in the thylakoid lumen, has been crosslinked to plastocyanin in plants and is thought to be a plastocyanin docking protein (see Globeck 1992). However, inactivation of the *psa*F gene in a cyanobacterium didn't alter the ability to grow photoautotrophically and is apparently dispensable (Chitnis et al. 1991).

The total molecular mass of PS I is estimated to be 340 kDa. It is thought to bind around 100 chlorophylls. Almost half of the chlorophylls have been located around the transmembrane regions of the cyanobacterial PsaA and PsaB polypeptides in the 3-D structure (Krauss et al. 1993). In terrestrial plants, the PS I core and LHC I together bind approximately 200 chlorophyll molecules.

Overall, there is remarkable conservation of the core reaction center complexes and other thylakoid complexes directly involved in the electron transfer process. In fact, all of the major complexes discussed above are quite highly conserved amongst the plants, eukaryotic algae and the cyanobacteria. This is in contrast to a considerable amount of variation in the antenna complexes which includes differences in size, chlorophyll and carotenoid content, and in their structural organization around the reaction centers.

1.5   Light-harvesting antenna systems

1.5.1   Function of antenna complexes

The light-harvesting antennae, located on the periphery of the reaction centers, increase the absorptive cross-section of the photosystems thereby increasing the probability of absorbing

the available solar radiation. The antennae are protein complexes that specifically bind pigments (chlorophylls and carotenoids) in a manner that determines the position and orientation of the chromophores; this allows for efficient capture and transfer of the excitation energy. The protein environment greatly influences the absorptive properties of the non-covalently bound chromophores.

Plants and algae typically use solar radiation in the visible range (350-700 nm) and are able to absorb these low energy photons due to the conjugated bond system of the chlorophyll molecule. The delocalization of electrons throughout this conjugated system lowers the energy difference between the ground and excited state allowing for the absorption of the photons in the visible range. Chlorophyll a is present in all oxygenic photosynthetic organisms and is the only chlorophyll type within the core complex of either PS I or PS II. Though Chl a is associated with all integral membrane (intrinisic) antennae, there is significant variation in the type of accessory chlorophylls and/or carotenoids that are also bound to the complex.

The two main accessory chlorophylls in oxygenic organisms are chlorophyll b, (in terrestrial plants, green algae, euglenophytes and the prochlorophytes—see Table 1.1) and chlorophyll c (c_1-c_3) (in the chromophytic algae). However, Chl c-like pigments have been found in the ancient green alga, *Mantoniella* and in some prochlorophytes (Table 1.1). Structures of the different chlorophylls are shown in Figure 1.3. The accessory chlorophylls increase the light capturing ability of the organism by broadening the absorption profile of the antennae. Chlorophylls absorb primarily in the blue (Soret band) and red (α-band) regions with low absorption in the green. Chl b, with respect to Chl a, is red shifted in the Soret band and blue shifted in the α-band, illustrating how the combination of the two chlorophylls results in a broadened absorption spectrum. Chl c absorption is similar to that of Chl b except absorption in the α-band is lower (relative to the Soret band) and blue shifted.

Carotenoids are important for both photoprotection and for light harvesting. In terrestrial plants the photoprotective role is of primary importance as the carotenoids do not make a significant contribution to the absorption spectrum. Carotenoids act as photoprotectors by quenching chlorophyll triplet states, that can result in the production of highly reactive singlet

oxygen, or they can quench singlet oxygen states directly (Rau 1988). In addition, oxygenated carotenoids (xanthophylls) may also take part in the xanthophyll cycle which is involved in dissipation of excess light energy.

In the chromophytes carotenoids are important for light-harvesting, in addition to their photoprotective functions. In these cases, the carotenoids are abundant and make significant contributions to the absorption properties of the cells. Carotenoids that play a dominant role in light capture typically absorb in the 480-560 nm range, significantly broadening the absorption capabilities in a region where chlorophylls have poor absorption. The structures of some of the main carotenoids are shown in Figure 1.3 and will be discussed in the sections that follow.

Increasing the antennal size surrounding the reaction centre necessitates an efficient process of excitation energy transfer inward towards the reaction center. When a chlorophyll absorbs a photon, an electron in the chlorophyll is knocked from a ground state orbital to a higher energy, excited state orbital. It is possible that the excitation may not be localized to the orbitals of a single pigment but may be delocalized over several pigments (a delocalized exciton) which may contribute to the migration of excitation energy (Sauer 1986). Energy transfer from excited chlorophyll complexes may also occur by inductive transfer (Förster transfer) which is effective over longer distances. In this mechanism, an excited donor pigment relaxes to the ground state after transferring the excitation energy to a neighboring acceptor, which is then excited (Sauer 1986). In either energy transfer process, a separation of a positive and negative charge between donor and acceptor pigments (electron transfer) does not occur. This only occurs at the reaction center chlorophyll of PS I and PS II.

The antennae are closely associated with the reaction center complexes in the thylakoid membrane. The following discussion has separated the different LHCs into two main categories depending on whether they are integral membrane complexes (intrinisic) or whether they are externally bound to the thylakoid membrane (extrinisic). I will concentrate on the LHCs from the oxygen evolving organisms, both prokaryotic and eukaryotic. I will not cover the LHCs of the anoxygenic photosynthetic bacteria.
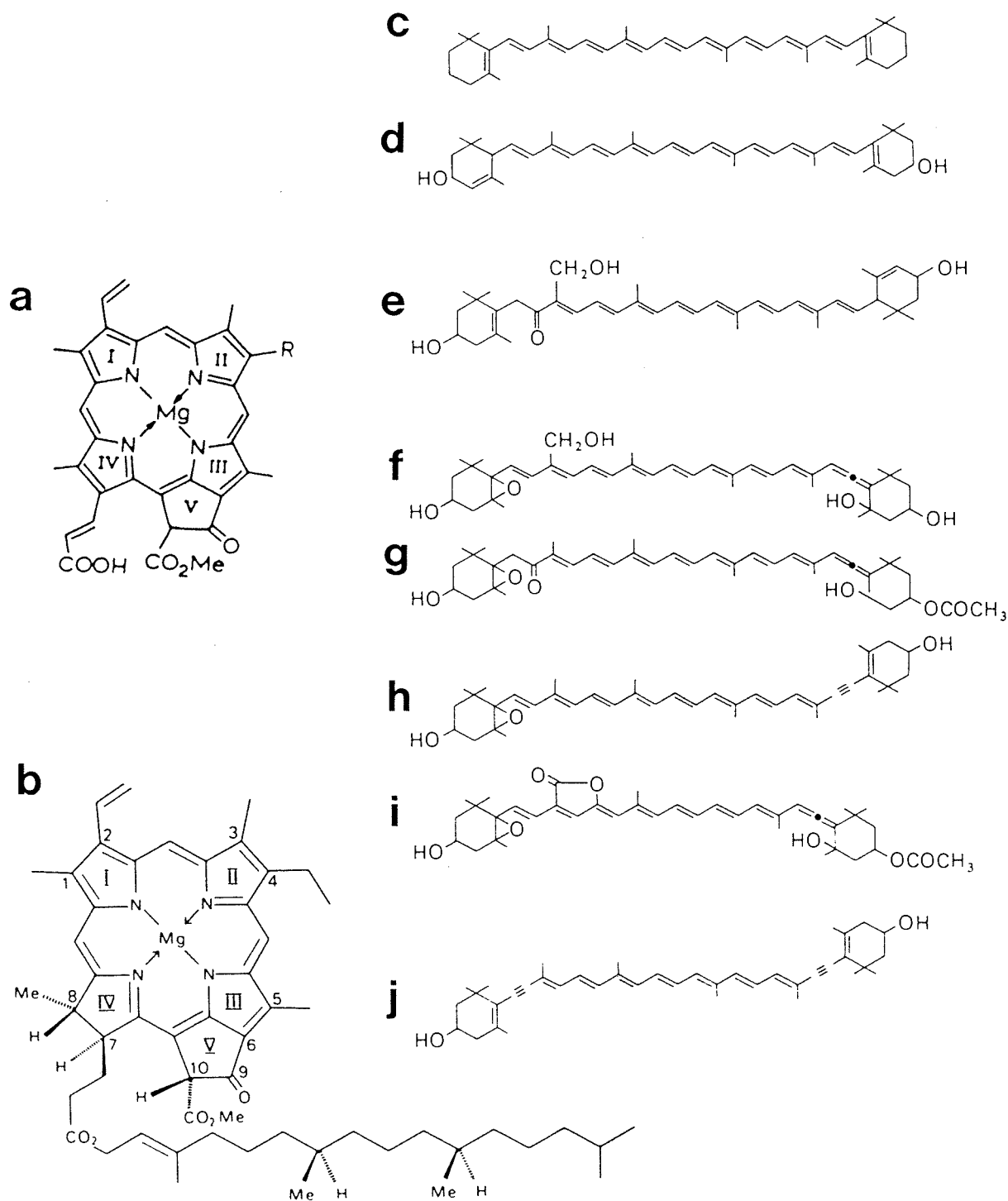
Figure 1.3

Structure of the main chlorophylls (a,b) and carotenoids (c-j) in the algae. (a)
Chlorophylls $c_1 + c_2$. In c_1 R is —C_2H_5; in c_2 R is —CH=CH$_2$. (b) Chlorophylls $a + b$. In
chlorophyll b the —CH3 on ring II is replaced by —CHO. Carotenoids are as follows: (c)
β-carotene, (d) lutein, (e) siphonaxanthin, (f) vaucheriaxanthin, (g) fucoxanthin, (h)
diadinoxanthin, (i) peridinin, and (j) alloxanthin.

21

1.5.2   Membrane-intrinsic light-harvesting antennae

1.5.2.1   Prokaryotic light harvesting complexes

The prochlorophytes are oxygen evolving prokaryotes that lack phycobisilomes and contain Chl $a + b$. This similarity to the terrestrial plant chloroplasts lead to suggestions that the prochlorophytes were the ancestors of chloroplasts containing Chl's $a + b$ (Lewin, 1975). The first identified prochlorophyte, *Prochloron* was not free-living and was found endosymbiotically associated with didemnid ascidians. The Chl $a + b$ antenna from *Prochloron* is 34 kDa, has a Chl a/b ratio of around 2.4, and can be phosphorylated (Schuster et al. 1984; Hiller and Larkum 1985). Interestingly, *Prochloron* has recently been shown to contain significant amounts of a Chl c-like pigment, in addition to Chl's a and b (Larkum et al. 1994). The first discovered free-living prochlorophyte, *Prochlorothrix hollandica*, was a fresh water dwelling species. In *Prochlorothrix*, there are several Chl $a + b$ antennae containing proteins in the 32 to 38 kDa range having a Chl a/b ratio of 4 (Bullerjahn et al. 1987; van der Staay 1992; van der Staay and Staehelin 1994). The LHC polypeptides from the above two prochlorophytes are immunologically related to each other (Bullerjahn et al. 1990) but neither shows relatedness to the CAB proteins of terrestrial plants, determined by a lack of cross reactivity with CAB directed antibodies (Hiller and Larkum 1985; Bullerjahn et al. 1990).

Interestingly, the *Prochlorothrix* LHC reacts with an antibody directed against a Chl a-binding, iron-stress induced protein from cyanobacteria (*isi*A) (Bullerjahn et al. 1987). Preliminary protein sequence of the LHCs from *Prochlorococcus* (LaRoche and Partensky, unpubl.), *Prochloron* (Hiller and Larkum, unpubl.) and *Prochlorothrix* (van der Staay and Green, unpubl.) indicates that the prochlorophyte LHC genes are related to the *isi*A gene product and to the Chl a core antenna, CP43. This confirms a lack of relatedness to LHC II of the terrestrial plants and green algae.

1.5.2.2   Eukaryotic light harvesting complexes (LHC)

The main intrinisic eukaryotic LHCs are a family of functionally analogous complexes that are evolutionarily related; discussed more thoroughly in Chapter 5. As the different pigment compositions of these complexes probably reflect significant evolutionary divergence, I will consider each major complex separately. The main intrinisic LHCs are the Chlorophyll *a* + *b*-binding proteins (CABs), the fucoxanthin-chlorophyll proteins (FCPs), and the *intrinisic* peridinin-chlorophyll *a* + *c*-proteins (*i*PCPs) of dinoflagellates. As well, the antennae of other chromophytes containing abundant xanthophylls other than fucoxanthin, will be reviewed.

The chlorophyll a + b-binding proteins (CABs)

The chlorophyll *a* + *b*-binding proteins are found in all terrestrial plants (vascular plants, ferns, mosses). They are also found in the Chlorophyta (Green algae) and in the photosynthetic members of the Euglenophyta. The green amoeba, *Chlorarachnion*, also contains Chl *a* + *b* (Hibberd and Norris 1984), but nothing is known about the LHCs of this organism. The best characterized plants in terms of chlorophyll protein complexes are the angiosperms, particularly tomato, spinach and barley. Some work has also been done on the green alga *Chlamydomonas*. The chlorophyll protein complexes and the different mildly-denaturing gel systems used to isolate them have been extensively reviewed (Green 1988; Thornber et al. 1991; Jansson 1994). The literature has been confused with a number of different labeling systems though there has been a recent revision of the gene nomenclature (Jansson et al. 1992). I will adhere to this system when referring to the *Cab* genes and will use the system of Green et al. (1991) when referring to the protein complex. One exception is in the designation of CP29 type I and II which I will refer to as CP26 and CP29, respectively (Bassi et al. 1990). This will be done due to the lack of close relatedness between the two proteins and to avoid confusion. A table reviewing the CAB nomenclature is shown in Table 1.2.

23

Table 1.2   Summary of the tomato CAB proteins

gene	complex	size (kDa)	amino acids§	function location	Chl a/b ratio	gene copy #	# of introns	Chr. location
Lhcb1	LHC II type I	27	265			8	0	3 & 2
Lhcb2	LHC II type II	26	265	major PS II	1.3	2	1	7 &12
Lhcb3	LHC II type III	25	265	antenna		≥ 3	2	12
Lhcb4*	CP29 (CP29-I)	28	287	core PS II	≈2.6	?	?	?
Lhcb5	CP26 (CP29-II)	26	286	antenna	≈1.9	1	5	?
Lhcb6	CP24	24	210	minor PS II antenna		2	1	?
psbS	CP22 (22 kDa)	22	276	minor antenna?	3-4	1	3	?
Lhca1	LHC I type I	22	246	LHC I-730		2	3	5
Lhca2	LHC I type II	23	270	LHC I-680	2.2-3.5	1	4	10
Lhca3	LHC I type III	25	273	LHC I-680		1	2	10
Lhca4	LHC I type IV	21	250-251	LHC I-730		2	2	3 & 6

*Arabidopsis gene; § immature polypeptide; Chr.=chromosome
Table 1.1 is a modification of Green at al. (1991) and Jansson (1994).

The crystal structure of the pea LHC II complex has been determined at a 3.4 Å resolution. It contains three transmembrane α-helicies and the complex is stabilized by ionic bonds between buried residues within the first and third transmembrane α-helicies. The complex contains 12 chlorophylls (tentatively identified as 7 Chl a and 5 Chl b) and two lutein molecules at the center of the complex for photoprotection (Kühlbrandt et al. 1994). The other xanthophylls associated with the complex were not localized though there is expected to be a neoxanthin and a violaxanthin molecule also associated with LHC II (Peter and Thornber 1991).

LHC II is the most predominant antenna in the terrestrial plants and green algae and forms trimers that are specifically associated with PS II (Kühlbrandt and Wang 1991). In terrestrial plants LHC II consists of three polypeptides with an unequal stoichiometry. Of these, the 28 kDa LHC II type I protein (Lhcb1) is most abundant followed by the 27 kDa polypeptide (LHC II type II; Lhcb2). These are encoded by a multigene family consisting of anywhere from 3 to 16 members for Lhcb1 and 1 to 4 for members for Lhcb2, depending on the plant (Green et al. 1991;

24

Jansson 1994). The 25 kDa LHC II type III complex (*Lhcb3*) is the next most abundant LHC II component. There are reports of different subcomplexes of the main LHC II complex with the LHC II type I and type II (*Lhcb1* and *Lhcb2*) proteins forming a more peripheral, mobile LHC II antenna (Larsson et al. 1987; Larsson et al. 1987b; Peter and Thornber 1991; Jansson 1994). This would be distal to a LHC II antenna complex with primarily LHC II type I (Larsson et al. 1987; Larsson et al. 1987b) and possibly LHC II type III polypeptides (Peter and Thornber 1991). This second subcomplex would be more closely associated with the core complex of PS II. Presumably the excitation of PS II is partially controlled by the reversible detachment of the distal LHC II subcomplex (Larsson et al. 1987; Larsson et al. 1987b; Peter and Thornber 1991). Changes in light intensity, redox state of the plastoquinone pool or temperature can result in the phosphorylation of LHC II (Allen 1992). This leads to migration of the phosphorylated LHC II from PS II in the appressed regions to PS I in the nonappressed parts of the thylakoid membrane (Anderson and Andersson 1988).

Plants grown under intermittent-light conditions (limits Chl *b* production) preferentially accumulate LHC II type III (*Lhcb3*) over the other LHC II polypeptides indicating the production of these complexes is differentially regulated (White and Green 1988; Mawson et al. 1994). Because of this, it has also been suggested that LHC II type III is more proximally located to the reaction center and may function as a linker for the bulk LHC II antennae containing types I and II (Harrison and Melis 1992; Mawson et al. 1994). In agreement with this, regreening experiments with intermittent-light grown barley have shown that elevated levels of LHC II type III accumulate early in the continuous light phase, relative to type I and II (Dreyfuss and Thornber 1994).

The minor antennal complexes associated with PS II include CP29 (*Lhcb4*), CP26 (*Lhcb5*), CP24 (*Lhcb6*) and the 22 kDa polypeptide (*psbS*). They account for 6% of the total PS II associated Chl (Peter and Thornber 1991). CP29 (aka CP29 type II) and CP26 (aka CP29 type I) are often not recognized as distinct complexes since they comigrate in some gel systems. The Chl *a/b* ratio of CP29 and CP26 are different and estimated to be 2.6 and 1.9, respectively (Bassi et al. 1993). As these complexes (CP29 and CP26) were not as readily dissociated from

PS II with detergents as compared to LHC II, they were considered more tightly associated with the reaction center (Barbato et al. 1989; Camm and Green 1989; Peter and Thornber 1991). In addition, CP29/CP26 remain present in the thylakoids of a barley Chl *b*-deficient mutant that otherwise fails to accumulate LHC II (White and Green 1987b), indicating there is differential stability of some of the LHC II complexes in the absence of Chl *b*. Both CP29 and CP26 are related to LHC II and to each other (Pichersky et al. 1991; Morishige and Thornber 1992), though the relationship is distant (Chapter 5).

The third minor chlorophyll-protein complex (CP24, *Lhcb*6) has a low Chl *a/b* ratio ($\approx$ 1), a molecular mass of 24 kDa (Dunahay and Staehelin 1986) and is quite divergent from either LHC I or LHC II polypeptides (Jansson 1994). CP24 is not tightly associated with the PS II complex and has been proposed to connect the LHC II complex to the PS II reaction center (Barbato et al. 1989). The final minor complex, CP 22 (with a single 22 kDa apoprotein) is associated with the core complex of PS II and has recently been shown to bind chlorophyll (Funk et al. 1994). Sequence of the gene encoding this protein (*psbS*) revealed limited similarity to the CABs and showed four putative membrane spanning regions (Kim et al. 1992).

Evidence for an antennal complex specifically associated with PS I (LHC I) was first shown in pea PS I preparations (Mullet et al. 1980). Since then it has been isolated and more thoroughly investigated in a number of terrestrial plant taxa (Haworth et al. 1983; Lam et al. 1984; Lam et al. 1984b). The presence of a Chl *a* + *b* containing PS I associated LHC (CP O) has also been demonstrated in the green alga *Chlamydomonas* (Ish-Shalom and Ohad 1983). Two main subcomplexes of LHC I, termed LHC I 680 and LHC I 730, have been isolated and found to have Chl *a/b* ratios in the range of 2.2-3.5 (Lam et al. 1984; Bassi and Simpson 1987). They are so named due to their 77 K fluorescence emission at 680 and 730 nm, respectively. In barley, LHC I 680 has been shown to consist of two separate polypeptides encoded by the *Lhca*2 (LHC I type II) and *Lhca*3 (LHC I type III) genes which have apparent molecular masses of 23 and 25 kDa, respectively. LHC I 730 consists of two polypeptides, with apparent molecular masses of 22 and 21 kDa, and N-terminal sequencing has demonstrated that these polypeptides correspond to LHC I type I (*Lhca*1) and LHC I type IV (*Lhca*4), respectively (Knoetzel et al.

1992). The instability of PS I and the easy detachment of LHC I 730 in a LHC I 680 depleted barley mutant indicates that LHC I 680 is involved in binding LHC I 730 to the reaction center core (Knoetzel and Simpson 1991).

The CAB proteins of green algae

The pigment-protein complexes from Chl $a + b$ containing algae (other than *Chlamydomonas*) have been studied very little. The LHC II sequences from *Chlamydomonas*, *Dunaliella* and the terrestrial plants are all highly conserved; these evolutionary relationships will be more thoroughly considered in Chapter 5. The organization of the inner LHC II antennae should be quite similar between the green algae and the terrestrial plants as the same complexes have been identified in each. However, there are indications of novel regulatory mechanisms in the green alga, *Dunaliella*, which involve modifications in the Chl a/b ratio in changing light intensities (Sukenik et al. 1987). Further work may unveil significant differences in the terrestrial plant-green algal CAB organization and regulation. Nonetheless, there are some green algal taxa that have a pigment composition significantly different from those of *Chlamydomonas* and *Dunaliella*; these will be discussed below.

In the green alga, *Codium* sp. (Siphonales), the CAB proteins contain the unusual carotenoid siphonaxanthin (instead of lutein) which increases the absorbance in the 500-550 nm range. The LHC II component has a lower Chl a/b ratio than the terrestrial plant LHC II (0.7) and contains four polypeptides from 27-35.5 kDa (Anderson 1985). Two of these polypeptides (34 and 35.5 kDa) have sizes significantly different from the terrestrial plant CABs. Furthermore, a PS I specific antenna was isolated from this alga. It contained siphonaxanthin, had a Chl a/b ratio of 1.7 and a polypeptide composition resembling the LHC I polypeptides of terrestrial plants (19-25 kDa) (Chu and Anderson 1985). The *Codium* CAB proteins are immunologically related to those of terrestrial plants, despite the pigment differences between them (Anderson et al. 1987).

The Prasinophyte algae (Micromonadophyceae) have unique LHC antennae that contain Chl a, Chl b, and a Chl c-like pigment (Mg 2,4-divinyl pheoporphryin a_5 monomethyl ester)

(Jeffery 1989). The carotenoid prasinoxanthin was also associated with the antenna complex. The LHC complexes from the prasinophyte, *Mantoniella squamata*, contain at least two polypeptides of 20.5 and 22 kDa (Fawley et al. 1986b) and are arranged into larger oligomeric complexes of 80 kDa (possibly trimers) (Rhiel et al. 1993). The presence of a Chl *c*-like pigment and the smaller size suggested a closer relationship to the Chl *a* + *c*-binding proteins; however, sequencing of the a gene encoding this LHC has confirmed its relatedness (though distant) to the CABs (Rhiel and Mörschel 1993). There is some evidence that a unique PS I associated antennae may not exist in *Mantoniella* and that the same protein complex excites both photosystems (Schmitt et al. 1993). The significance of this in terms of regulation and distribution of the complexes remains to be seen.

Euglena gracilis also has Chl *a* + *b*-binding antennae that have sizes in the 26-28 kDa range (Cunningham and Schiff 1986). The predominant LHC has been reported to have a molecular ratio of 12 Chl *a*: 6 Chl *b*: 4 diadinoxanthin: 1 neoxanthin (Cunningham and Schiff 1986b). Interestingly, the xanthophyll diadinoxanthin is more commonly found in the chromophytes rather than in Chl *a* + *b*-containing organisms. Nevertheless, sequencing of genes encoding LHC II and LHC I proteins from *Euglena* has confirmed their relatedness to the CABs (Houlné and Schantz 1988; Muchhal and Schwartzbach 1992). In spite of the sequence similarities, the LHCs from *Euglena* are uniquely translated into large polyprotein precursors from unusually large mRNAs.

The fucoxanthin-chlorophyll proteins (FCPs)

Fucoxanthin is an oxygenic, allenic-xanthophyll (Fig. 1.3) that absorbs in the 450-550 nm range. Its presence in some algae extends the absorption range of the LHC into the green region of the spectrum. This would be particularly useful as the coastal ocean waters are usually limited in the blue wavelengths of the spectrum and are more light limited than the terrestrial plants (Larkum and Barrett 1983). However, there is no correlation between light quality (due to attenuation of specific wavelengths of light at different depths) and the nature of the light-harvesting system used by the algae growing there (Saffo 1987).

Fucoxanthin-chlorophyll proteins occur in the diatoms, chrysophytes, phaeophytes, haptophytes, and some members of the raphidophytes (including *Heterosigma*). The exact molar ratio of the pigments associated with the FCP are not precisely known. However, analyses of complexes isolated using milder detergents suggests that there are 13 Chl a: 3 Chl c: 10 fucoxanthin: 1 violaxanthin for a brown alga FCP (Katoh et al. 1989) and approximately 12.5 Chl a: 5 Chl c: 24 fucoxanthin for a diatom FCP (Friedman and Alberte 1984). It is clear that xanthophylls are much more abundant in these chlorophyll-proteins than in those of the terrestrial plants. Usually, no attempt is made to distinguish between the different Chl c forms (c_1, c_2, c_3), though in one study a Chl c_2/c_1 ratio of 3.1 was observed for a *Phaeodactylum* FCP (Owens and Wold 1986). Both fucoxanthin and Chl c transfer excitation energy to Chl a, indicating their function in the harvesting of light energy (Duval et al. 1983).

Typically, one or two polypeptides with a molecular mass of 15-21 kDa have been identified in FCP-containing fractions (see Table 1.3). Earlier studies reported the isolation of a Chl $a + c$ pigment-protein complex without fucoxanthin (Barrett and Anderson 1980; Alberte et al. 1981; Peyriere et al. 1984; Owens and Wold 1986; Hsu and Lee 1987; Boczar and Prezelin 1989). As the pigment content in these complexes was variable and the polypeptides had the same apparent size as the FCPs, the nature of these complexes is uncertain. However, the Chl $a + c$ complexes were isolated using either the detergent triton X-100 or SDS and there is the possibility that they are the result of pigment loss from the main FCP complex (Hiller et al. 1991). In this study, a specific Chl $a + c$-containing complex lacking fucoxanthin was not found.

Structural relatedness of the FCPs to the CABs has been suggested on the basis of immunological cross-reactions of these polypeptides to antibodies directed towards one of the two LHC types (Caron et al. 1988; Passaquet et al. 1991; Plumley et al. 1993). In this dissertation, antisera specific for the two groups of LHCs were used to investigate the relatedness of the *Heterosigma* FCPs to the CABs and other FCP complexes. The FCPs, like the CABs, are also nuclear encoded, translated on cytoplasmic ribosomes, and processed into the mature polypeptide (Fawley and Grossman 1986). Their relatedness to the CABs was first confirmed by the sequencing of a cDNA encoding the FCP from *Phaeodactylum* (Grossman et al. 1990). This

protein possessed four hydrophobic regions, three of which are present in the mature protein and may potentially form membrane spanning regions. Since then, cDNAs encoding FCPs from another diatom (*Odontella*-gb X81054) a brown alga (Apt et al. 1994) and a haptophyte (LaRoche et al. 1994) have been sequenced, in addition to the *Heterosigma Fcp* cDNA.

Other chromophyte antennae

Other taxa where xanthophylls, other than fucoxanthin, play a significant role in light absorption include the cryptophytes, the xanthophytes and the eustigmatophytes. The cryptophytes possess a Chl a + c_2 containing antenna complex which is abundant in the xanthophyll alloxanthin. The molar ratio of these pigments in an antenna fraction from *Cryptomonas rufescens* has been estimated to be 10 Chl a: 2Chl c: 3.4 alloxanthin (Lichtlé et al. 1987). The polypeptides of these antennae are in the 18-24 kDa size range, similar to the FCPs, and Chl a/c ratios of 1.4 (*Chroomonas*), 1.7 (*Cryptomonas maculata*) and 4.8 (*Cryptomonas rufescens*) have been recorded (Ingram and Hiller 1983; Lichtlé et al. 1987; Rhiel et al. 1987). A small peptide sequence fragment from a cryptomonad LHC (Sidler et al. 1988) showed similarities to the diatom and brown algal FCP sequences.

A 23 kDa LHC from the xanthophyte *Pleurochloris meiringensis* contains Chl a, Chl c and three abundant xanthophylls: diadinoxanthin, vaucheriaxanthin, and heteroxanthin (Wilhelm et al. 1988). These pigments have been reported to exists in the following molar ratio: 1000 Chl a, 224 Chl c, 148 heteroxanthin, 264 diadinoxanthin, and 129 vaucheriaxanthin. A PS I associated antenna with a fluorescence emission different from the main LHC has also been isolated from this alga (Büchel and Wilhelm 1993).

In the eustigmatophytes there is a lack of Chl c while the xanthophyll, violaxanthin, is abundant and plays a significant light-harvesting role. The main LHC fraction from another eustigmatophyte, *Monodus subterraneus*, was enriched in a 23 kDa polypeptide. The pigment ratios of this complex were estimated to be 10 Chl a: 2.8 violaxanthin: 1.3 vaucheriaxanthin ester (Arsalane et al. 1992). The 23 kDa polypeptide is immunologically related to the main FCP from a brown alga indicating a structural relatedness to this group of LHCs (Arsalane et al. 1992).

Table 1.3   Summary of characteristics from chromophyte LHCs

group	organism	accessory pigment	Chl a/c ratio	polypeptides kDa	Refs
Phaeophyta (brown algae)	Acrocarpia	Chl $c_1 + c_2$ fucoxanthin	2	nd	1
	4 taxa		nd	20-23	2
	Fucus		5.6	21	3
	Laminaria		3.3	20	3
	Dictyota		4.3	54	4
Bacilliarophyta (Diatoms)	Phaeodactylum	Chl $c_1 + c_2$ fucoxanthin	2.7	17.5, 18.5	5
			2.3 + 5.0	15	6
			3.0	18, 19, 19.5	7
			2.7	16.4, 16.9	8
			1.8-3.2	15-20	9
	Cylindrotheca		2.8	16, 18	10
	5 taxa		1.5-1.7	≈18	11
Haptophyta	Pavlova	Chl $c_1 + c_2$ fucoxanthin	nd	16.5, 17, 18, 20.5	12
			4.7	21	13
	Isochrysis		1	18, 20, 24	14
Xanthophyta	Pleurochloris	Chl c diadinoxanthin vaucheriaxanthin heteroxanthin	4.5	22-23	15 16
Eustigmatophyta	Nannochloropsis	violaxanthin vaucheriaxanthin-ester	no Chl c	26	17, 25
	Mondus			23	18
	Visheria			20?	18
Cryptophyta	Chroomonas	Chl c_2 alloxanthin	1.4	20, 24	19
	Cryptomonas		1.7	18, 19, 22?	20
Dinophyta	Gonyaulax	Chl c_2 peridinin	1-4	17-24	21
			≈3	19?	22
	Amphidinium		1.7	19	23
	3 taxa		nd	19-20	24

[1](Barrett and Anderson 1980)   [9](Caron and Brown 1987)   [17](Livne et al. 1992)
[2](Caron et al. 1988)   [10](Hsu and Lee 1987)   [18](Arsalane et al. 1992)
[3](Berkaloff et al. 1990)   [11](Brown 1988)   [19](Ingram and Hiller 1983)
[4](Katoh et al. 1989)   [12](Fawley et al. 1987)   [20](Rhiel et al. 1987)
[5](Friedman and Alberte 1984)   [13](Hiller et al. 1988)   [21](Boczar and Prezelin 1987)
[6](Gugliemelli 1984)   [14](LaRoche et al. 1994)   [22](Knoetzel and Rensing 1990)
[7](Fawley and Grossman 1986)   [15](Wilhelm et al. 1988)   [23](Hiller et al. 1993)
[8](Owens and Wold 1986)   [16](Büchel and Wilhelm 1993)   [24](Iglesias-Prieto et al. 1993)
[25](Chrystal and Larkum 1987)

Preliminary peptide sequence information from a LHC from *Nannochloropsis* (eustigmatophyte) (Livne et al. 1992) indicates that these polypeptides are related to the FCPs.

The intrinisic peridinin-chlorophyll proteins (iPCPs)

The dinoflagellate LHCs bind Chl a, Chl c_2 (not c_1) and the xanthophyll, peridinin. Peridinin is brick red in colour and has a C_{37}-skeleton (instead of C_{40} as with other xanthophylls) and an oxidized in-chain methyl group (see Fig. 1.3i). There are two main light-harvesting complexes in the dinoflagellates: an intrinisic peridinin-chlorophyll $a + c$ complex (*i*PCP) and a water-soluble PCP (*s*PCP). The latter LHC will be discussed in the next section. There is some evidence for the occurrence of a Chl $a + c_2$ complex with a high proportion of Chl c_2 (Chl a/c ratio ≈ 0.3) and devoid of peridinin (Boczar et al. 1980; Boczar and Prezelin 1987). However, the status of this complex is uncertain due to the use of SDS in its isolation, variations in pigment content and lack of its detection in other studies (Hiller et al. 1991; Hiller et al. 1993; Iglesias-Prieto et al. 1993). The *i*PCP is the main light-harvesting antenna in the dinoflagellate, *Symbiodinium*, with a Chl a, Chl c_2, peridinin ratio of 1:1:2 and comprising 45, 75 and 70% of the total cellular levels of these pigments, respectively (Iglesias-Prieto et al. 1993). In *Amphidinium*, the Chl a/c ratio is a little higher (1.7) though a similar ratio of Chl a/peridinin occurs (Hiller et al. 1993). In all studies, the polypeptides for the *i*PCPs are 19-20 kDa in size. Immunological cross-reactivity of FCP polypeptides with *i*PCP directed antibodies and the protein sequence of the *i*PCP of *Amphidinium* demonstrates similarity to the FCPs, suggestive of a common evolutionary origin (Hiller et al. 1993; Hiller, unpubl. data-gb Z47563). The individual *i*PCP polypeptides from *Amphidinium* are encoded by a putative polyprotein (Hiller, unpubl. data). This would be similar to coding arrangement of the *Euglena* LHC proteins (Houlné and Schantz 1988). Significantly, photosynthetic members of the euglenophytes and the dinoflagellates have a total of three membranes surrounding the chloroplast.

1.5.3   The extrinisic light-harvesting antennae

The extrinisic light-harvesting antennae are soluble complexes that are easily detached from the membrane. This type of complex includes the soluble peridinin-chlorophyll protein complexes of the dinoflagellates, as well as the phycobilisomes of both the red algae and cyanobacteria.

Soluble Peridinin-Chlorophyll Complex (sPCP)

Soluble peridinin-chlorophyll *a* complexes were the first photosynthetic proteins characterized from the dinoflagellates. Early studies indicated that the complexes were extrinisically associated with the thylakoid membrane, bound significant amounts of the total peridinin and transferred excitation energy efficiently from peridinin to Chl *a* (Prézelin and Haxo 1976; Song et al. 1976). The *s*PCPs also had a molar ratio of Chl *a* to peridinin of 1:4 and were bound to a monomeric complex of 31-35 kDa or to an apparent 15 kDa homodimer, depending on the species (Govind et al. 1990). Recent studies on the *s*PCPs have estimated that they bind 5 and 15% of the total cellular Chl *a* and peridinin, respectively (Iglesias-Prieto et al. 1993), though much higher estimates have been reported. This suggests that, though the *s*PCPs make significant contributions to the absorption of light, the *i*PCPs are the main antennal complexes. Recently the sequence of a gene encoding a *s*PCP from *Symbiodinium* was determined. It encoded a 35 kDa polypeptide with a internal duplication, suggesting it arose from the fusion of genes encoding the 15 kDa *s*PCP form (Norris and Miller 1994). No similarity to any other LHC was found. Interestingly, the resemblance of part of the *s*PCP transit sequence to the transit sequence of a thylakoid lumen localized polypeptide of terrestrial plants has lead to the suggestion that the *s*PCPs may be localized within this compartment (Norris and Miller 1994).

Phycobilisomes

The phycobilisomes are large antennal complexes consisting of many chromophore-binding proteins which are responsible for light absorption in the 450-655 nm range, the

wavelengths where there is poor absorption by chlorophyll. The primary phycobiliproteins making up the phycobilisome are phycoerytherin (PE) ($A_{max} \approx 560$ nm), Phycocyanin (PC) ($A_{max} \approx 620$ nm) and allophycocyanin (AP) ($A_{max} \approx 650$ nm). Each phycobiliprotein has two different subunits (α and β) with a molecular mass in the 17-22 kDa range that form dimers. The α and β polypeptides of each phycobiliprotein share a degree of sequence similarity and are evolutionarily homologous (Zuber 1986). Each subunit binds 1-4 open chain tetrapyrrole chromophores (phycobilins) that are attached by a thioether bond to a cysteinyl residue on the protein. The spectral characteristics of the phycobiliprotein are partly influenced by the protein environment of the chromophore (Glazer 1989), as is the case with other pigment-protein complexes.

Each phycobiliprotein dimer (α,β) is arranged into cyclic trimers which form the building blocks of the phycobilisome. Two phycobiliprotein cyclic trimers ($\{\alpha,\beta\}_3$) are assembled into hexameric protein aggregates ($\{\alpha,\beta\}_6$) of PE and PC, which make up the rod like structures of the phycobilisomes. These are bound to a usually triangular shaped core assemblage of AP. The phycobiliprotein aggregates are assembled and held together with colourless linker polypeptides (Mörschel and Rhiel 1987). The exact composition of PBS rods depends upon the organism, growth and light conditions and has been previously reviewed (Mörschel and Rhiel 1987; Glazer 1989; Gantt 1990; Grossman et al. 1993). Typically, the PBSs are organized in two ways: (1) in a hemi-discoidal fashion with a central core of AP from which the PE and PC rods radiate, or (2) in a hemi-ellipsoidal arrangement which is similar to the above organization except it is twice as thick (Gantt 1981). Both types of pycobilisomes are found in the cyanobacteria and red algae.

The cryptomonads also contain phycobilins (phycoerythrin or phycocyanin) that are located within the thylakoid lumen (Spear-Bernstein and Miller 1989). This phycobiliprotein does not form phycobilisomes but does specifically transfer excitation energy to PS II (Lichtlé et al. 1980).

1.6   Concepts in chloroplast evolution

Under the assumption that the endosymbiotic origin of organelles is generally accepted
(Gray and Doolittle 1982; Taylor 1987), there are two main hypotheses as to the origin of the
different chloroplast types described in section 1.3. The idea of an endosymbiotic origin of the
chloroplast was first proposed by Mereschowsky (1905) where he suggested that the different
plastid types were the result of endosymbioses of different cyanobacteria-like organisms with
nonphotosynthetic phagotrophic protists. These ideas were revised more recently when it was
suggested that the diversity of the plastid types were the direct result of numerous endosymbiotic
events with different prokaryotes already divergent in pigment biosynthesis, antennal systems
and biochemical pathways (Sagan 1967; Raven 1970; Whatley and Whatley 1981); this is
referred to as the polyphyletic view of chloroplast evolution. An alternative hypothesis is that
there was only one primary endosymbiotic event between a cyanobacterium-like organism and a
phagotrophic, nonphotosynthetic eukaryotic host. Subsequent divergence of this organism lead
to the different plastid types observed in the plants and algae today. This view is referred to as
the monophyletic origin of plastids (Cavalier-Smith 1982; Taylor 1987)

I am making a distinction between what I call the primary and secondary endosymbioses
leading to the chloroplasts (see Fig. 1.4). The term primary endosymbiosis refers to the
establishment of a chloroplast from a prokaryotic source. This generally is thought to have
resulted in the generation of an alga with a double membrane around the chloroplast, namely the
red and green algae (Fig. 4.1, top). Debate tends to center around whether the chloroplast from
these two organisms share a common ancestor (monophyletic) or have separate origins
(polyphyletic). With the term secondary endosymbiosis, I am referring to events involved in the
establishment of chloroplasts directly from photosynthetic, eukaryotic sources (Fig. 4.1, bottom).
This is generally thought to have occurred in the algae where there are more than two membranes
surrounding the chloroplast (chromophytes, cryptophytes, *Chlorarachnion*, euglenophytes and
the dinoflagellates), as reviewed in section 1.3 (see McFadden and Gilson, 1995).

Primary endosymbiosis

Photosynthetic   **Phagotrophic**   **Photosynthetic**
prokaryote   **eukaryote**   **eukaryote**

Secondary endosymbiosis

Photosynthetic   **Phagotrophic**   **Photosynthetic**
eukaryote   **eukaryote**   **eukaryote**
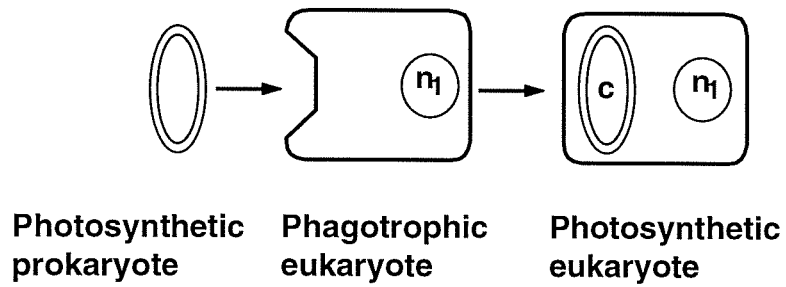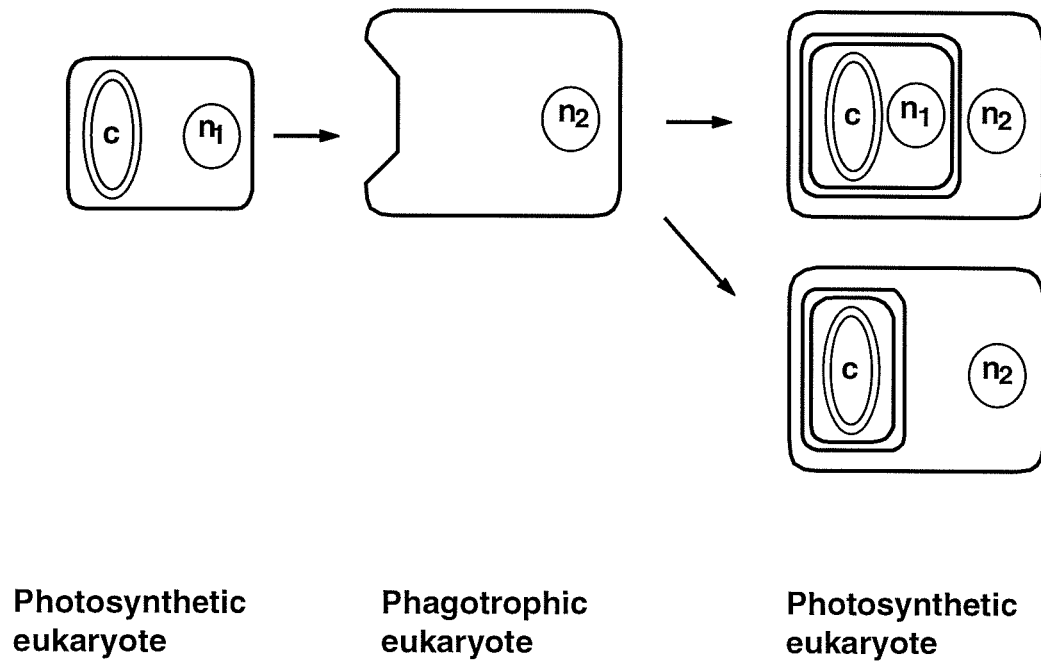
Figure 1.4

     Schematic diagram illustrating the proposed acquisition of chloroplasts (c) through a primary or a secondary endosymbiosis. The nuclei of the different organisms are indicated (n1 or n2). The nucleus of the photosynthetic eukaryote in a secondary endosymbiosis (n1) is thought to have given rise to the nucleomorph (if maintained).

36

Traditional classification of the different algal groups was heavily based on their pigment content. This resulted in the separation of three main taxonomic groups: the Rhodophyta (red algae) with Chl *a* and phycobilisomes, the Chlorophyta (green algae) with Chl's *a* and *b*, and the Chromophyta (coloured algae) with Chl *a, c* and significant amounts of xanthophylls such as fucoxanthin, vaucheriaxanthin and peridinin. The accessory pigments and the proteins binding them have been assigned considerable weight in the speculations as to the number of prokaryotes involved in primary endosymbiotic events. The discovery of a chlorophyll *a* + *b* containing prokaryote (prochlorophyte) (Lewin 1975) with thylakoid stacking similar to green algae and terrestrial plants was interpreted as evidence for a polyphyletic chloroplast origin. Presumably, the green algae would have acquired a chloroplast from a prochlorophyte (or 'green' prokaryote) (Raven 1970). Moreover, the red algae were thought to have acquired a chloroplast from a cyanobacterium based on the presence of phycobilisomes in both these organisms. Along the same line of reasoning, the chromophytes were thought to have separately acquired a chloroplast from a 'yellow' prokaryote (Raven 1970) with chlorophylls *a* and *c* (Whatley and Whatley 1981) or from the anaerobic photoheterotrophic eubacterium *Heliobacterium chlorum* (Margulis and Obar 1985), based on apparent pigment similarities. These comparisons were the foundation upon which the arguments for a polyphyletic view of chloroplast evolution were built.

Though it is generally accepted that the red algal chloroplast evolved from a cyanobacterium-like organism, the evolution of the green algal and chromophyte chloroplast from Chl *a* + *b* and Chl *a* + *c*-containing prokaryotes, respectively, is more controversial. The proposed evolution of the chromophyte plastid from *Heliobacterium* was not widely accepted and has subsequently been disproved by SSU rRNA analysis (Witt and Stackebrandt 1988). The evolution of the green algal chloroplast from a prochlorophyte is still debated and will be discussed in Chapter 3.

Gene organization and phylogenetic analyses of chloroplast encoded gene sequences provided some evidence for a close relationship between the red and chromophyte chloroplasts. Chloroplast gene organization and gene localization of the ATPase subunits (Kowallik 1993), Rubisco large and small subunit (*rbc*L/S) operon (Douglas and Durnford 1989), *psb*A (D1)

37

sequences (Winhauer et al. 1991) and other gene clusters (Reith and Munholland 1993; Douglas 1994b) have been primarily used for examining these evolutionary relationships. In addition, phylogenetic analysis of the *psb*A, *rbc*L/S, *tuf*A (elongation factor Tu), (Morden et al. 1992), and plastid 16s rRNA (Douglas and Turner 1991b) sequences also supports the relationship between the red and chromophyte plastids. This would indicate that the chromophyte plastid may have evolved from an association of a red algal-like ancestor with a phagotropic eukaryotic host, leading to the chromophyte lineage.

The only clear evidence suggesting that a secondary endosymbiosis occurred was with recent work on *Cryptomonas* (Douglas et al. 1991) and *Chlorarachnion* (McFadden et al. 1994; McFadden and Gilson 1995). Like the chromophytes, both of these organisms have four membranes around the chloroplast. However, the cryptophytes and chlorarachniophytes have a membrane bound, nucleic acid containing organelle (called a nucleomorph) located in the space between the outer two CER membranes and the inner two membranes of the chloroplast envelope (the periplastidal space) (Greenwood et al. 1977; Gillott and Gibbs 1980; Hibberd and Norris 1984). It is generally thought that this organelle is the vestigial nucleus of the eukaryotic endosymbiont (Ludwig and Gibbs 1987). In the cryptomonads, the presence of phycobiliproteins (phycoerytherin or phycocyanin) and the storage of starch in the periplastidal space (the former cytoplasm of the putative endosymbiont) lead to the suggestion that the endosymbiont was a red algal-like ancestor (Gillott and Gibbs 1980). This was supported by an analysis of the nuclear SSU rRNA sequence from the nucleomorph and the nucleus of *Cryptomonas Φ*. Phylogenetic analysis showed a loose affiliation of the *Cryptomonas* nucleomorph sequence with the red algae while the cryptomonad nuclear sequence was more distant in the tree (Douglas et al. 1991). The nucleomorph rRNA transcripts were also localized in the nucleomorph and the periplastidal space (McFadden et al. 1994a). In a similar strategy, the nucleomorph rRNA sequence from *Chlorarachnion* was localized in the periplastidal space and was shown to be distinct from the nuclear rRNA sequence (McFadden et al. 1994b). The endosymbiont leading to the chloroplast of *Chlorarachnion* has been hypothesized to be a green alga (Hibberd and Norris 1984); however, phylogenetic studies were inconclusive (McFadden et al. 1994b).

38

1.7   Methods used in molecular phylogeny

There are only a few reports of phylogenetic analyses of the *Cab* genes. Demmin et al. (1989) examined the relationships of the LHC II *Cab* gene family within the angiosperms. From their maximum likelihood analysis of *Lhcb*1 (LHC II type I) and *Lhcb*2 (LHC II type II) sequences, they found that the angiosperm taxa grouped within their traditional taxonomic families. Their trees suggested that the *Lhcb*1 and *Lhcb*2 divergence occurred prior to the monocot/dicot separation. Matsuoka (1990) suggested the *Lhcb*1/2 divergence occurred prior to the angiosperm and gymnosperm separation. Jansson (1994) also examined the evolutionary relatedness of the CAB proteins in a recent review and found there was a close association of CP29 (Lhcb4) and LHC I type I (*Lhca*1). These trees also showed a separation between LHC I (*Lhca*) and LHC II (*Lhcb*). Within the LHC II assemblage, the green algal sequences formed a separate branch from the *Lhcb*1-3 (LHC II types I-III) genes of the terrestrial plants. Unfortunately, little information was given on the alignment, characters utilized or specifics about the method used in the analysis. As no indication of reliability was given, it is not possible to judge which relationships may be significant. Recently, an analysis of a FCP from the haptophyte, *Isochrysis galbana*, suggested it was more related to a tomato LHC I sequence than to a LHC II sequence; therefore, the FCPs were suggested to have been derived from a LHC I-like sequence, before the appearance of the predominant LHC II antennae (LaRoche et al. 1994).

One goal of this study is to examine the evolution of the *Cab* and *Fcp* gene families and relate these gene relationships back to the function of the protein complex (Chapter 5). As well, I hope to get a better idea of the *Cab* and *Fcp* gene relationships in order to determine when they may have separated in relation to the functional separation of LHC I and LHC II. Two methods were used for the determination of phylogenetic relationships amongst the CABs and FCPs: maximum parsimony and distance matrix. Both methods are available in the PHYLIP computer package (Felsenstein 1992). Maximum parsimony is a character method based on the principle that the evolutionary pathway requiring the fewest "steps" is the most likely. Parsimony attempts

to generate a tree which can be explained with the smallest number of mutational events (shortest length). The characters used in parsimony can be of a molecular or morphological nature. In this study, a character is a specific amino acid (e.g. Ala, Ser, etc.) at a particular position (e.g. the 49th position of an alignment). Parsimony evaluates each character site separately on all possible unrooted trees and records the number of changes required to explain the character distribution. The optimal tree is selected by adding the number of changes over each character site, for all possible trees, and choosing the one with the fewest steps (see reviews by Swofford and Olsen, 1991; Hillis et al., 1993 and Stewart, 1993). With large datasets, not all possible tree combinations can be tested as the number of possible trees increases exponentially with an increase in the number of taxa. For example, with 13 taxa or more there will be over 13 billion possible trees, making the calculations impractical (Hillis et al. 1993).

To overcome this, PHYLIP uses a heuristic algorithm which starts with the first two (or three) taxa and creates a tree. The remaining taxa are then added individually in a stepwise fashion to all possible positions on the tree. Each of the trees is evaluated at each step and the shortest tree is kept. This process continues until all the taxa have been added to the tree. In order to improve the chances of finding the most optimal tree, the branches undergo a series of global rearrangements whereby groups (subtrees) are removed and added back to the tree in all possible positions. After each addition, the length of the tree is again assessed and the shorter one is retained. This process continues until no further improvements in the tree topology have been recorded (Felsenstein 1992).   The evaluation of the trees is based on the chosen optimality criteria. This refers to the method by which the evolutionary change of the characters is assessed or weighted (Swofford and Olsen 1990). In this study, an amino acid change is weighted according to the number of mutations required to explain the substitution, which is based on the genetic code. This point is explained further in the methods section of Chapter 5.

Distance methods are based on pairwise comparisons between taxa, the end result being a single value which reflects the dissimilarity (or distance) between the two sequences. Distances are calculated between all possible pairs of taxa and referred to as the distance matrix. With the distance methods, characters are not considered individually between all taxa as they are in the

character based methods (parsimony). No ancestral state is implied or required as no distinction is made between derived or ancestral character states (Sober 1988). The distance matrix calculation can be based on three modes of amino acid substitution in the PROTDIST program, which is included in the PHYLIP package (Felsenstein 1992). In this study, distances were calculated using the PAM (accepted point mutation) 100 matrix of Dayhoff (1978) which weights amino acid changes on the basis of their probable occurrence. This is an empirically determined matrix calculated through the comparison of cytochrome c sequences from different species. This distance matrix is then used to infer a tree by any of several methods. This study routinely used the Neighbor-Joining method (Saitou and Nei 1987) for tree construction as described in Chapter 5.

1.8   Scope of this thesis

This dissertation is primarily concerned with the characterization of the FCP complexes and the genes that encode them. I am also interested in using this information in an analysis of evolutionary relatedness between the different light harvesting antennae. The FCPs are separated and immunologically analyzed for structural relatedness to both the CABs and the FCPs in Chapter 2. The immunological analysis of the LHCs in the red alga *Aglaothamnion neglectum* is presented in Chapter 3. In addition, work done on the red alga, *Porphyridium cruentum*, in collaboration with Beth Gantt and Greg Wolfe at the University of Maryland, is presented in this Chapter. Characterization of the *Heterosigma Fcp* sequence and an analysis of the size and complexity of the nuclear encoded multigene family is presented in Chapter 4. Finally, Chapter 5 involves an analysis of the evolutionary relationships amongst the CABs and FCPs which combines many of the ideas from the previous chapters.

CHAPTER 2

Characterization of the light-harvesting proteins from Heterosigma carterae

2.1 Introduction

As mentioned in the general introduction, there have not been any investigations into the chlorophyll-protein complexes of any raphidophycean alga. Even amongst the other chromophyte algae, there have been relatively few studies characterizing the fucoxanthin-chlorophyll proteins as compared to the CABs of terrestrial plants. As I am interested in the similarities and differences amongst the diverse antennal systems of the algae, I decided to work on a representative of the raphidophytes, *Heterosigma carterae.*

In the initial part of this study I isolated and characterized the different chlorophyll-proteins contributing to the antennae of this organism. This study used two different methods for the fractionation of the thylakoid membrane and subsequent characterization of its components, particularly the FCPs. One method was the fractionation of digitonin solubilized thylakoids on a sucrose gradient, a method that has been used on chromophyte algae with success in the past (Hiller et al. 1991). The other method involved fractionation of solubilized thylakoids by a partially denaturing SDS-polyacrylamide gel electrophoresis (PAGE). This method has been invaluable in the characterization of the terrestrial plant CABs but has, so far, had limited applications for the non-green algal LHCs as the complexes are much more unstable under these conditions.

2.2 Materials and Methods

2.2.1 Heterosigma cultures

An axenic culture of *Heterosigma carterae* was maintained in an artificial sea water mixture as previously described (Cattolico et al. 1976). The media was prepared by adding salts to distilled water at the following concentrations : 0.35 M NaCl, 0.02 M $MgSO_4$, 0.021M $MgCl_2$, 7.8 mM $CaCl_2$, 7.5 mM KNO_3, 0.37 mM KH_2PO_4 and 0.37 mM $NaHCO_3$. To this 1 M Trizma base pH=7.6 (Sigma) was added to a final concentration of 1.9 mM. Stock A (50 mM Na_2EDTA and 9 mM $FeCl_3$) was added at 0.76 ml/l followed by stock B (0.29 mM $ZnCl_2$, 9.7 mM H_3BO_3, 0.12 mM $CoCl_2$, 0.24 mM $CuCl_2$, 2.5 mM $MnCl_2$ and 0.03 mM $(NH_4)_6MO_7O_{24}$), again at 0.76 ml/l. A 0.1 µg/ml vitamin B_{12} solution was then added at 0.38 ml/l. The media was made up to the appropriate volume and an additional 42 ml/l of distilled water was added to account for evaporation during autoclaving. Media was autoclaved at 121°C for 30 minutes. Cells were grown in 1200 ml of media in three litre fernbach flasks. These were grown with continuous agitation on an orbital shaker at 75 rpm. The light was kept on a 12:12 hour light/dark cycle to induce synchronicity (Cattolico et al. 1976). Light levels were maintained at 60 $\mu E/m^2/min$ and the temperature was constant at 18°C throughout the light and dark cycles. Cultures were routinely tested for contamination when inoculated by adding 0.5 ml of culture to 5 ml of nutrient marine media (2.0 g nutrient broth and 1.25 g yeast extract per 250 ml artificial sea water). Cell counts were done using a standard hemacytometer. Cell counts were made after they were killed by adding 0.5% formaldehyde (50 µl/1 ml culture)

2.2.2 Heterosigma thylakoid fractionation

Late log phase cells were harvested at 400 × g for 12 minutes, resuspended in cold 0.33 M sorbitol, 1 mM $MgCl_2$, 50 mM HEPES pH 7.6 and protease inhibitors (1 mM

phenylmethyl sulfonyl fluoride, 5 mM ε-amino-n-caproic acid, 1 mM benzamidine-HCl, 1 mg/ml leupeptin). Protease inhibitors were routinely used in solutions, being added from stock solutions prior to use. Cells were lysed under 4000 kPa (600 psi) nitrogen in a Yeda Press (Yeda Research and Development Co. Ltd. Rehovot, Israel) to release the chloroplasts. The chloroplasts were separated by differential centrifugation in a swinging bucket rotor at 6500 x g for 12 minutes at 4°C. Chloroplasts were washed three times in cold 0.1 M NaCl, 5 mM MgCl$_2$, 20 mM Tricine pH 8.0 (including protease inhibitors) yielding a washed thylakoid fraction. Thylakoids to be used for non-denaturing gel electrophoresis were made up to 10% glycerol prior to quick freezing in liquid nitrogen and storage at -80°C (Allen and Staehelin 1991).

Thylakoids used for sucrose gradient fractionation were used fresh and solubilized with digitonin at a detergent to chlorophyll ratio of 100:1, on ice for four hours with a constant gentle stirring. After centrifugation at 40 000 x g for 30 minutes, the supernatant was loaded onto a 0.3 M-1.2 M linear sucrose gradient on top of 1.3 M and 1.6 M sucrose cushions. Sucrose solutions were made up in 10 mM Tricine pH 8.0 containing 0.05% (w/v) digitonin. Samples were centrifuged for 24 hours at 250 000 x g in a swinging bucket rotor at 4°C. Fractions 2 and 3 (Fig. 2.1) were precipitated at 40 000 x g following dialysis at 4°C in 0.1 M CaCl$_2$, 10 mM MgCl$_2$, 10 mM Tricine pH 8.0 including protease inhibitors (as above). Due to the large amount of detergent at the top of the gradient, Fraction 1 (Fig. 2.1) was pelleted at 100 000 x g following extended periods of dialysis with many changes of the dialysis buffer. Chlorophyll concentrations were determined in 90% acetone using the equations of Jeffrey and Humphrey shown below (Jeffrey and Humphrey 1975).

Chl $a = 11.47$ (A$_{664}$) - 0.4 (A$_{630}$)   Chl $c_1 + c_2 = 24.36$ (A$_{630}$) - 3.73 (A$_{664}$)

2.2.3 Denaturing SDS-PAGE and Western Blotting

For denaturing gel electrophoresis, samples were solubilized in 2% SDS, 65 mM Tris-HCl pH 6.8, 50 mM dithiothreitol, 10% glycerol and heat denatured for 1 minute at 100°C.

Thylakoid and sucrose gradient fractions were loaded on the basis of chlorophyll. Gel slices from non-denaturing PAGE were incubated in 2X sample buffer (4% SDS, 132 mM Tris-HCl pH 6.8, 0.1 M dithiothreitol, 20% glycerol) at room temperature for 2 hours then heated to 80°C for 20 minutes. Polypeptides were separated on 12-16% or 7.5-15% SDS polyacrylamide gels (acrylamide: bis-acrylamide 37.5:1) containing 0.05% SDS and 1.32 M Tris-HCl pH 8.8, with a 2 cm stacking gel containing 5% acrylamide, 0.1 M Tris-HCl pH 6.1 and 0.1% SDS. Gels were run for 18 hours at 4°C with the Laemmli buffer system (Laemmli 1970). Coloured molecular mass standards (Amersham) were used to estimate molecular mass.

Proteins were electrotransferred to nitrocellulose in 50 mM sodium acetate pH 7.0 overnight at 200 mA and 4°C. As a guideline, 1/5th the amount of chlorophyll loaded on gels to be stained was used for the same samples destined to be used for western blotting. Western blotting was carried out as previously described (White and Green 1987). Western blots were reblotted after stripping the nitrocellulose membrane in 0.1 M glycine-HCl pH 2.2, 20 mM Mg-acetate, 50 mM KCl (Legocki and Verma 1981) followed by reblocking in 3% Hipure liquid gelatin (Norland Products Inc. New Brunswick, N.J.) in phosphate buffered saline (1.37 M NaCl, 27 mM KCl, 81 mM Na_2HPO_4, 15 mM KH_2PO_4, pH 7.4). Proteins were Coomassie stained for $\geq$ 2 hours using 0.1% Coomassie brilliant blue-R, 50% methanol and 7% glacial acetic acid. Gels were destained in several changes of a 20% methanol, 7% glacial acetic acid. Polyclonal antibodies used in this study were anti-CP1a (α-CP1a), specific for barley CP I plus LHC I (White and Green 1987) and anti-FCP (α-FCP), specific for *Phaeodactylum tricornutum* fucoxanthin-chlorophyll *a* + *c* (FCP) protein complex (Fawley and Grossman 1986), which was provided by Dr. Art Grossman. Other antibodies include the α-PsaD antibody specific for a PS I associated subunit (also called PS I subunit #2) (Bengis and Nelson 1975) and α-D1, specific for the PsbA polypeptide of PS II, provided by L. McIntosh.

45

2.2.4 Non-Denaturing Gel System

Thylakoids were solubilized with a mixture of 0.9% octylglucoside, 0.9% decyl maltoside and 0.2% lithium dodecyl sulphate (in 2 mM Tris-maleate pH 8.0, 10% glycerol and with protease inhibitors) and resolved on a non-denaturing gel system according to Allen and Staehelin (Allen and Staehelin 1991) except that a 7% acrylamide gel was used with an acrylamide to bisacrylamide ratio of 150:1. A stacking gel was not used because it resulted in degradation of the pigment-protein complexes. Samples were solubilized on ice for 30 minutes at an anionic detergent to chlorophyll ratio of 30:1 with occasional mixing, then centrifuged in an microfuge for 20 minutes at 4°C. Samples were electrophoresed at 10 mA for 1.5 to 2.5 hours at 4°C. Estimations of molecular mass were done using non-denatured, high molecular mass markers (Pharmacia). Gel bands were excised and electrophoresed on a denaturing gradient gel as described above. Samples to be used for fluorescence data were excised from the gel and quick frozen in liquid nitrogen prior to storage at -80°C.

2.2.5 Spectroscopy and Fluorescence measurements

Absorption spectra were recorded on a Cary 210 Spectrophotometer at room temperature. P700 content was measured from the sucrose gradient fractions directly by monitoring the recovery of absorption at 700 nm after photo-oxidation by saturating red light with 1.7 mM ascorbate and 0.075 mM methylviologen present in the reaction mixture (Marsho and Kok 1972).

Fluorescence emission spectra were recorded with a Perkin Elmer LS50 fluorometer with the 77°K low temperature attachment and red sensitive photomultiplier. Excitation wavelength was 440 nm and the excitation and emission slit widths were adjusted to 10 nm and 5 nm, respectively. Spectra shown are an average of three scans. A 530 nm cut off filter helped to remove Rayleigh scatter in the 620 nm range. Gel slices from the non-denaturing gel system

were fitted in the cuvette with 60% glycerol and frozen in liquid nitrogen prior to measuring. Emission spectra were corrected for the drop in photomultiplier sensitivity in the 600-800 nm range using an averaged correction factor provided by Perkin Elmer. Excitation spectra were recorded from similarly prepared samples at 77 K. Emissions from the excitation spectra were detected at 680 nm in all samples. The excitation and emission slit widths were 2.5 nm and 10 nm, respectively. Scan rate was 300 nm/minute and the spectra shown are an average of two scans.

2.3 Results

2.3.1 Fractionation of digitonin-solubilized membranes by sucrose gradient centrifugation

Thylakoid membranes solubilized with digitonin were resolved into three major fractions on a sucrose gradient (Fig. 2.1). The top dark brown fraction (fraction 1) was rich in fucoxanthin and chlorophyll c as demonstrated by a broad shoulder from 488-540 nm and a prominent shoulder at 460 nm, respectively (Fig. 2.2A). Fraction 1 was removed from the 21% sucrose level and contained approximately 53% of the total chlorophyll. It also showed visible red fluorescence upon excitation with long wavelength UV light, indicating the detachment of the light-harvesting complex from the reaction center. A Chl a emission maximum of 675 nm, with a secondary peak at 732 nm, was recorded (Fig. 2.2B, solid line). A small peak at 637.5 nm is probably the result of the partial uncoupling of Chl c fluorescence which is preventing complete transfer of excitation energy from Chl c to Chl a. Residual P700 activity was detected in fraction 1, giving a Chl a/P700 ratio around 1200.

Sucrose Concentration   **Appearance**   **Chl a / c**   **Total Chl (%)**   **Abs. Maxima**

0.2 M

1   Brown   4   53   672, 440, 460*,
488*

2   Green-brown   14   25   677, 437, 460*,
496*

1.2 M
1.3 M   } 3 light-brown   10   11   675, 437, 460*,
1.6 M   496*
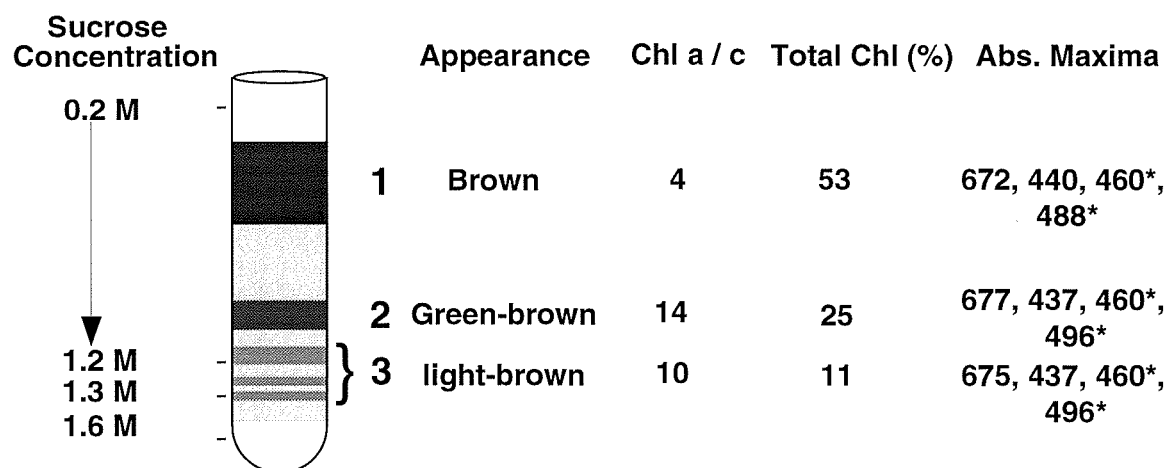
Figure 2.1

Schematic representation of sucrose gradient fractionation of digitonin solubilized
thylakoids, with Chl a / c ratios, percentage of total chlorophyll in each fraction and
absorption maxima data for the three major fractions. Asterisk in absorption data
indicates a shoulder.

Fractions 2 (30% sucrose level) and 3 (34% sucrose level) contained 25 and 11% of the
total chlorophyll, respectively (Fig. 2.1). Both contained significant amounts of carotenoid as
there is a prominent absorption at 496 nm in both fractions (Fig. 2.2A). An absorbance shoulder
at 460 nm also indicates the presence of Chl c though the Chl a/c ratios were 14 and 10 for
fraction 2 and 3, respectively. This indicated significant amounts of antennae were still
associated with the complexes. Both fractions (2 and 3) were enriched in PS I with Chl a / P700
ratios of 340 and 420. Both fractions 2 and 3 had emission maxima at 687 nm and significant
shoulders at 717 nm, though fraction 2 had the greatest fluorescence in the 717 nm region
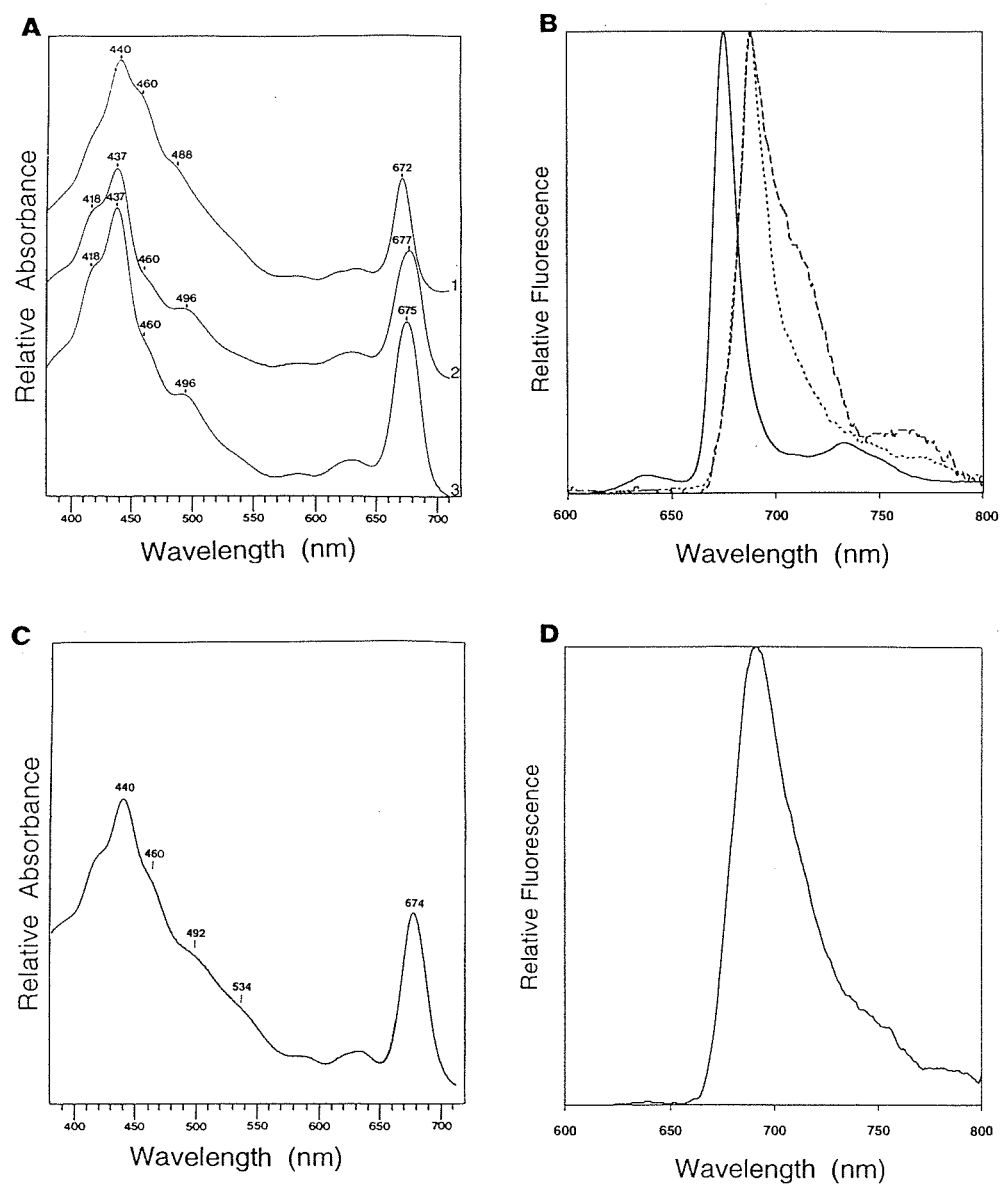
Figure 2.2

   Spectral characteristics of the sucrose gradient fractions. (A) Room temperature absoption and (B) fluorescence emission spectra. Spectra are fraction 1 (—), fraction 2 (---), fraction 3 (···). A room temperature absorption spectrum (C) and fluorescence emission spectrum (D) of thylakoid membranes is also shown. Emission spectra were excited with 437 nm light at 77°K.
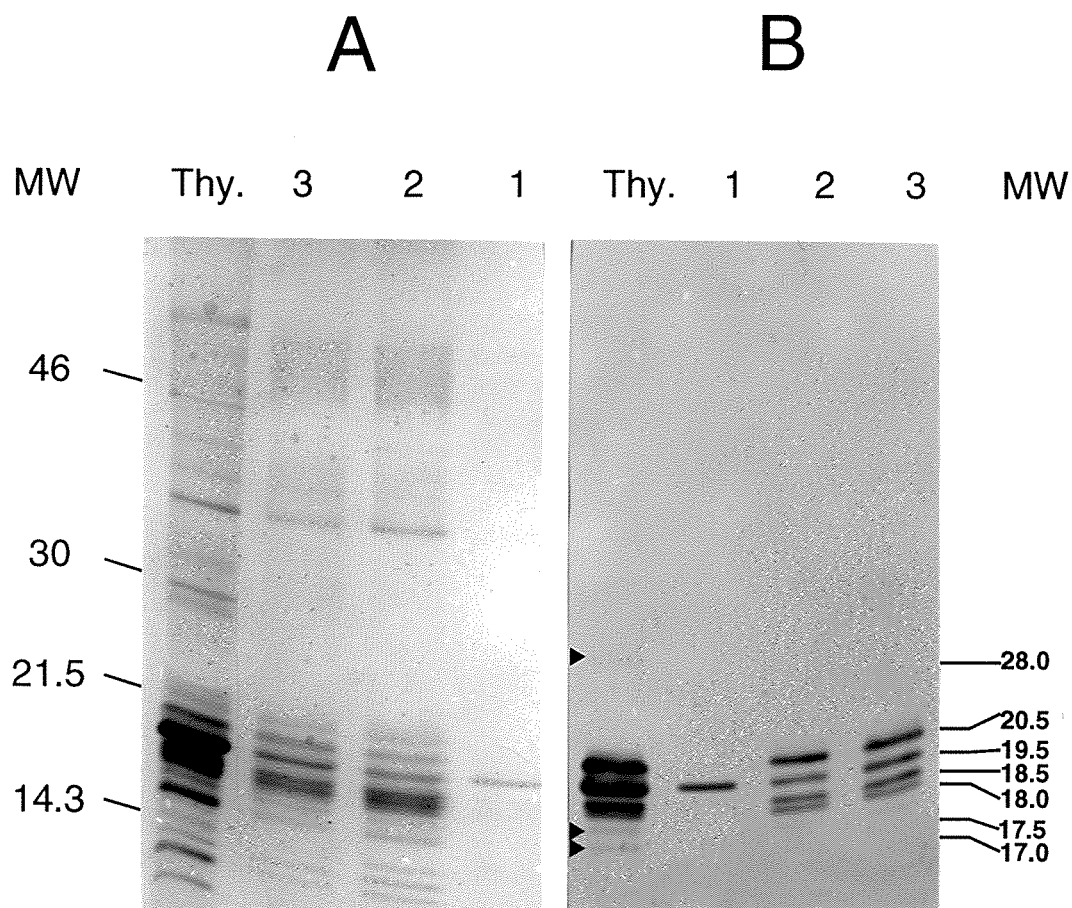
Figure 2.3

    Polypeptides of digitonin sucrose gradient fractions separated by denaturing 12-16% gradient SDS-PAGE. (A) Stained with Coomassie blue. Thy.: Whole thylakoids; 1: Fraction 1; 2: Fraction 2; 3: Fraction 3. (B) Western blot immunoprobed with the α-FCP antibody. Lanes labeled as in panel A. MW, molecular mass standards in kDa. Arrowheads on left of Panel B denote faint bands at 28.0, 17.0 and 17.5 kDa in whole thylakoids (Thy.) only.

(Fig. 2.2B). Fluorescence in the 717 nm area is attributed to fluorescence from PS I, indicating that fractions 2 and 3 are enriched in PS I. The strong emission at 687 nm is probably from the antennae complexes associated with this fraction. In contrast, thylakoids had a fluorescence emission maximum at 691 with a broad shoulder towards 740 nm (Fig. 2.2D). Room

50

temperature absorbance peaks at 440 and 674 nm, along with prominent shoulders at 460, 492 and 534 nm, were also characteristic of the thylakoids (Fig. 2.2C). Thylakoids solubilized with digitonin have a blue shifted fluorescence emission maximum that is more susceptible to changes in the assay buffer conditions (not shown). Because of this, fluorescence emission maxima have to be interpreted with caution.

SDS-PAGE (Fig. 2.3A) showed that there was a single polypeptide in Fraction 1. It cross-reacted with an antibody specific for the FCP from the diatom *Phaeodactylum tricornutum* (Fig. 2.3B). The purity of this fraction was ideal for obtaining tryptic fragments from the FCP, as will be discussed in Chapter 4. Though one polypeptide was usually found in this fraction, some extractions removed smaller amounts of the other three main FCP polypeptides, that are obvious in the thylakoid fraction (Fig. 2.3B).

Fractions 2 and 3 showed a similar polypeptide pattern with a number of bands of 16-22 kDa, a sharp band at 37 kDa and diffuse bands in the 49-55 kDa range (Fig. 2.3A). The four polypeptides estimated as 20.5, 19.5, 18.5 and 18.0 kDa cross-reacted strongly with the α-FCP antiserum (Fig. 2.3B). Two minor polypeptides at about 17.5 kDa and 16.5 kDa were faintly immunostained (lower arrowheads). The FCP antibody also detected a polypeptide with an apparent molecular mass of 28 kDa, found only in the thylakoid fraction (upper arrowhead, see also Fig. 2.4A). Note that in this Figure and subsequent figures, the apparent molecular masses determined by SDS-PAGE are used as labels to identify distinguishable polypeptides, and are not meant to imply accurate molecular mass determinations.

Using an antibody specific for barley CP1a (PS I core complex plus its corresponding light-harvesting polypeptides (White and Green 1987)), a different subset of cross-reacting polypeptides was found (Fig. 2.4). The α-CP1a antibody detected five major bands at approximately 16, 17.5, 18.5, 19 and 21.5 kDa in fractions 2 (Fig. 2.4C, left) and 3, while an additional four bands of about 18, 20, 22, 28 kDa could be resolved in thylakoids (Fig. 2.4A, left). To identify which of the polypeptides reacting with the α-CP1a antiserum were also immunostained with the α-FCP antiserum, the immunoblots shown in the left panels of Figure

2.4A-C were stripped and reblotted with the α-FCP (shown in Fig. 2.4A-C, right panels). The bands of 20.5, 19.5, 18.5, and 18.0 that were prominent in thylakoids immunostained only with α-FCP (Fig. 2.3) are heavily stained in Figure 2.4A (right panel) and are clearly distinguished from the bands labeled 21.5 and 22.0 above them and the 16.0 band below which only cross reacts with the α-CP1a antiserum. Similar results were obtained with fraction 2 when it was reblotted with α-FCP (Fig. 2.4C, right panel). Note that the major light-harvesting polypeptide in fraction 1, which was immunostained with the α-FCP, was not detected using the α-CP1a (Fig. 2.4B). Only four polypeptides at 28, 18.5, 18 and 17.5 kDa appeared to react with both antisera. Results of the immunoblotting with the two antisera are summarized in Table 2.1. These results show that there are up to 12 polypeptides in the FCP/CAB family in *Heterosigma* , a larger number than previously reported for any chromophyte alga.
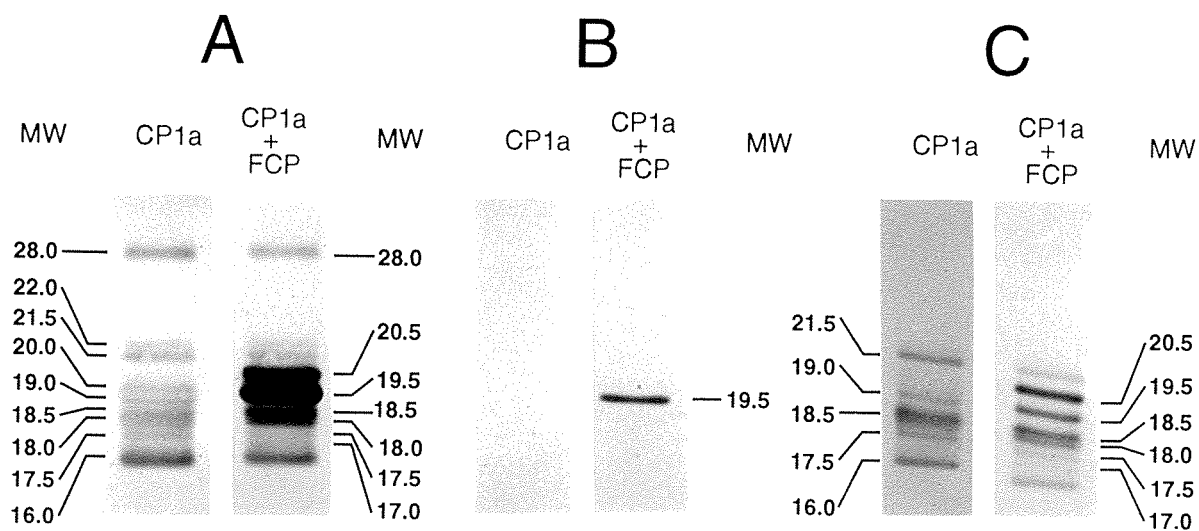
Figure 2.4

    Western blot of sucrose gradient fractions immunoprobed with α-CP1a (left panels) then stripped, blocked and immunoprobed with α-FCP (right panels) on the same blot. (A) Whole thylakoids (B) Fraction 1 from sucrose gradient (C) Fraction 2 from sucrose gradient. Approximate molecular masses (kDa) are used as labels to distinguish individual bands.

Table 2.1 Summary of cross-reactivity with the α-CP1a and α-FCP antisera

antiserum	Molecular Mass (kDa)											
	28.0	22.0	21.5	20.5	20.0	19.5	19.0	18.5	18.0	17.5	17.0	16.0
α-CP1a	+	+	+	-	+	-	+	+*	+*	+	-	+
α-FCP	+	-	-	+	-	+	-	+	+	+	+	-

* An apparent single band at 18-18.5 kDa may be a doublet

53

2.3.2 Fractionation by Non-Denaturing PAGE

Pigment-protein complexes were isolated from *Heterosigma* thylakoids solubilized in 0.9% octylglucoside, 0.9% decyl maltoside and 0.2% lithium dodecyl sulphate, and separated by means of the non-denaturing gel system of Allen and Staehelin (Allen and Staehelin 1991) (Fig. 2.5). With this system, eleven pigment-protein complexes were resolved (Fig. 2.5A). Bands 1 and 10 were resolved into two Complexes (1b, 10b) with longer periods of electrophoresis (Fig. 2.5B), but this resulted in some degradation of the central pigment-protein complexes. The first ten pigment-protein complexes were green and lacked noticeable fucoxanthin while Complex 11 was a brown fraction making up approximately 40% of the total protein.
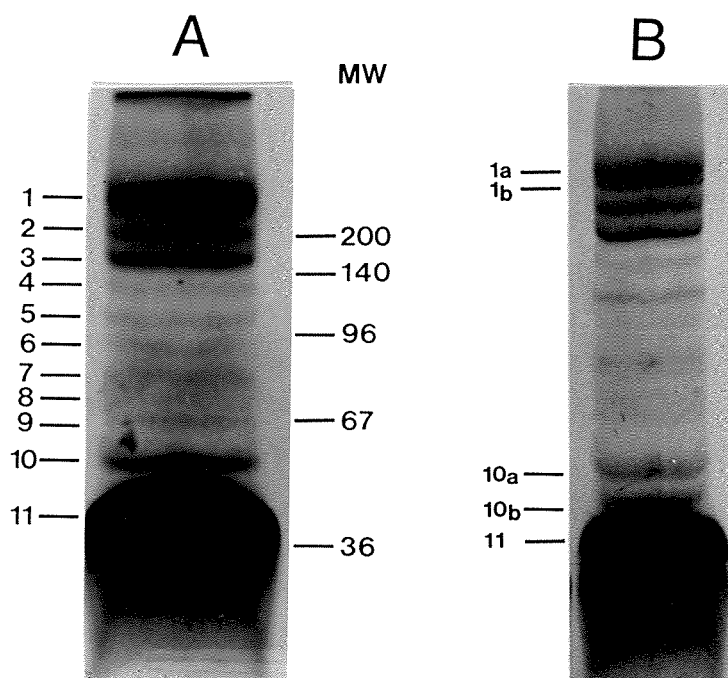
Figure 2.5

Unstained 7% polyacrylamide gel separating pigment-protein complexes of thylakoids solubilized with 0.9% octylglucoside, 0.9% decyl maltoside, 0.2% lithium dodecyl sulfate (Allen and Staehelin 1991). (A) Electrophoresed for 1.5 hours at 10 mA at 4°C. (B) Similar gel electrophoresed for 2.5 hours. Apparent molecular masses are given in kDa.

54

A

MW   M   1   2   3   4   5   6   7   8   9   10   11   Thy

46

30

21.5

14.3

B

Thy   1   2   3   4   5   6   7   8   9   10a   10b   11   Thy

32

21

14

D1

}FCP

psaD

MW   1   2   3   4   5   6   7   8   9   10a 10b 11

53 —

CP I
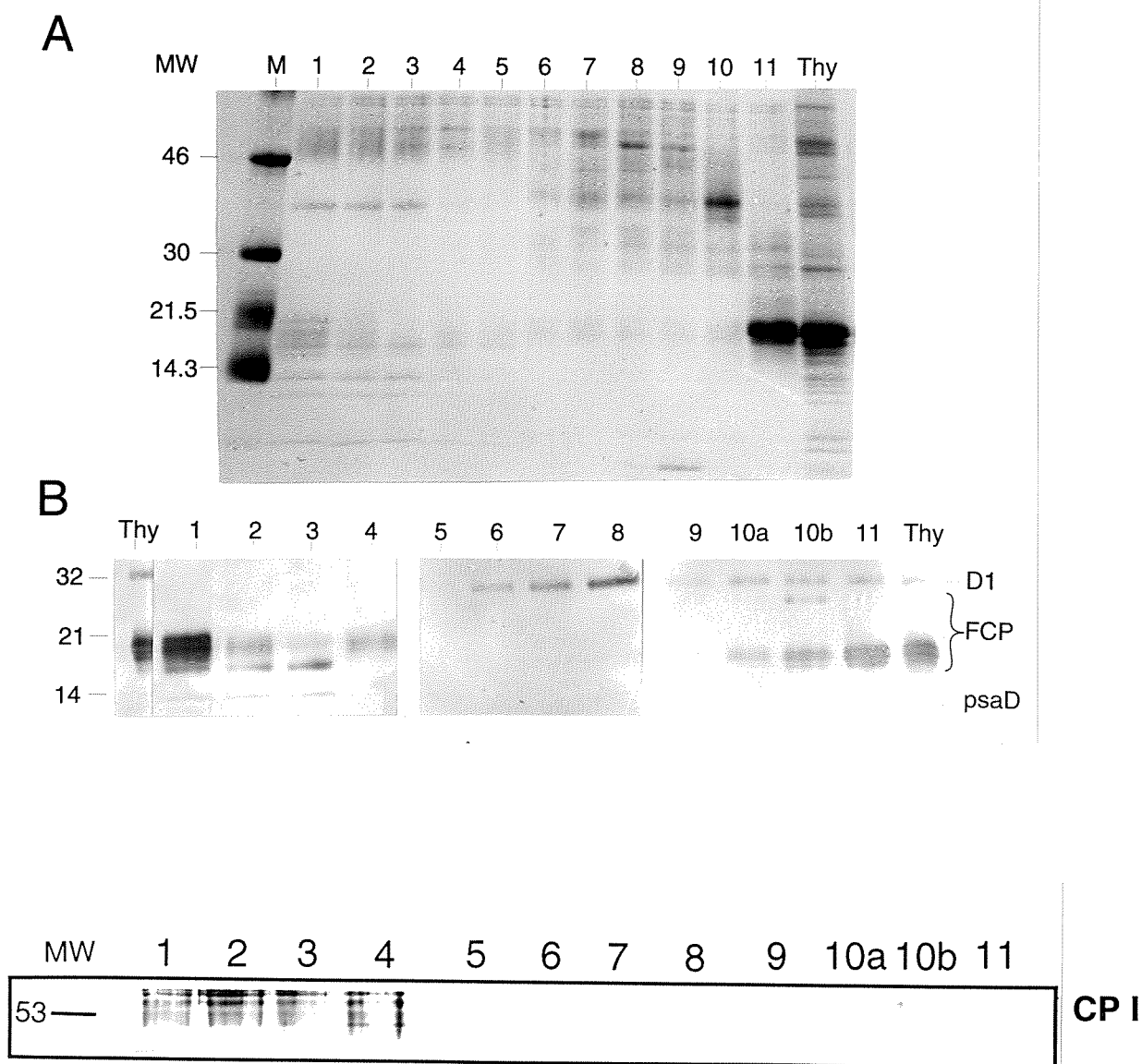
Figure 2.6

Analysis of isolated pigment-protein complexes. (A) 7.5-15% denaturing SDS-PAGE of pigment-protein Complexes 1-11 separated from non-denaturing PAGE. (B) Western blot analysis of the same pigment-protein complexes. Antibody used in sequential immunoprobing and area of cross reactivity indicated on the right; Molecular masses in kDa on the left. The bottom panel (CP I) is a western blot of the same gel fractions with the α-CP1 (PS I core complex) antiserum.

Denaturing SDS-PAGE of Complexes 1-3 (Fig. 2.6A) showed a broad stained band of about 53 kDa, a 37 kDa band, and a number of sharper bands in the molecular mass range (10-21 kDa) typical of the non-pigmented subunits of PS I (Golbeck and Bryant 1991). An immunoblot of samples from a similar gel was sequentially probed with several antibodies to determine the composition of the various bands in Figure 2.5. Immunodetection of the PS I-D (*psa*D gene product, Fig. 2.6B), and CP I polypeptides (Fig. 2.6, bottom panel) indicated that pigment-protein Complexes 1-3 were PS I complexes. PS I core polypeptides were also detected in Complex 4 but were lacking in the other fractions. When blots were probed with the α-FCP antibody, there appeared to be a number of antennal polypeptides in the 17-20.5 kDa range associated with the PS I core complex (Fig. 2.6B). The immunologically detected LHCs in Complexes 1-3 had the same molecular mass as those in Complex 11 and may be similar to the LHC polypeptides found in sucrose gradient fraction 2 and 3. However, unlike sucrose gradient fractions 2 and 3, there did not appear to be any fucoxanthin or Chl *c* associated with Complexes 1-3 (Fig. 2.7). This may have been the result of the detergents used in the extraction and the electrophoretic forces partially denaturing the complexes but allowing some LHCs to remain associated with the PS I core.

Complex 1 had the long wavelength absorption maximum at 677 nm typical of PS I (Fig. 2.7A) and a long wavelength chlorophyll *a* fluorescence emission maximum at 717 nm (Fig. 2.7B), similar to that attributed to the PS I specific light-harvesting complex of land plants (Haworth et al. 1983; Murata and Satoh 1986). It also had a second fluorescence emission maxima at 676 nm which may indicate uncoupled chlorophyll resulting from the detergent treatment (Fig. 2.7B). Excitation spectra of Complex 1 showed a Chl *a* peak at 437 nm and a second minor peak at 535 nm (Fig. 2.7C).

Complexes 6-10a appeared to be PS II-related, as they had a number of polypeptides migrating as somewhat diffuse bands in the 30-50 kDa range. An antibody specific to D1 (PS II reaction center polypeptide encoded by *psb*A gene) (Fig. 2.6B) showed that it was found in all these fractions but was absent from Complexes 1-5. Complexes 5-7 had long wavelength
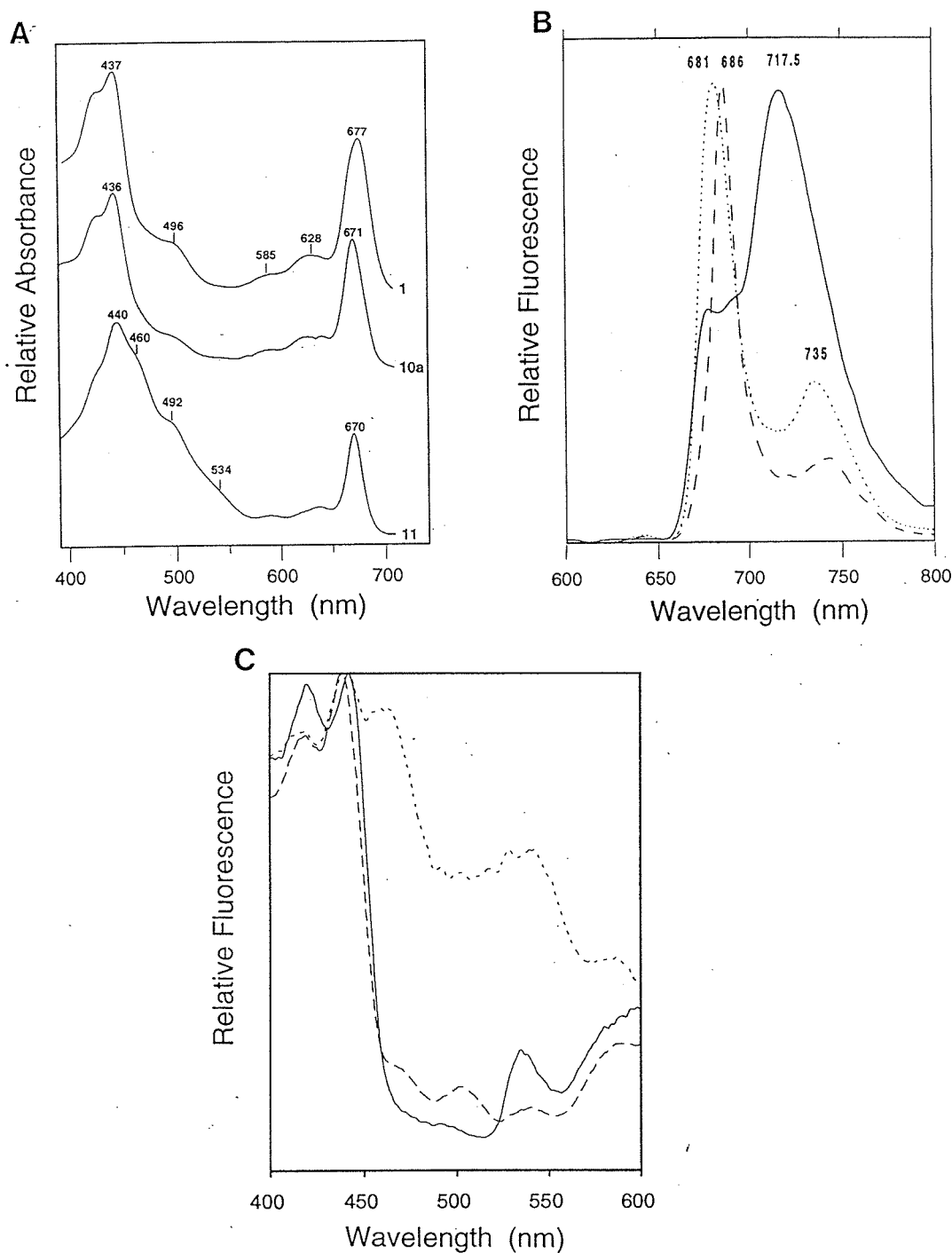
Figure 2.7

Spectral characteristics of pigment-protein Complexes 1, 10a, and 11 in gel slices from the non-denaturing SDS-PAGE. (A) Room temperature absorption spectrum. Spectra are offset for clarity. (B) Fluorescence emission spectra at 77 K of pigment-protein complexes. Spectra are Complex 1 (——); Complex 10a(_____); Complex 11 (......). (C) Fluorescence excitation spectra of the same three complexes and thylakoids at 77 K. Emission was detected at 680 nm.

maxima at 674 and Complexes 8-10b at 671 nm (data not shown; except 10a, Fig. 2.7).

Complex 10a had a fluorescence emission maximum at 686 nm (Fig. 2.7) which is typical of

PS II core complexes (Murata and Satoh 1986; Brown 1988). An excitation spectrum of

Complex 10a has a main Chl a peak at 437 nm and minor shoulder at approximately 470 and

502 nm, indicating minor amounts of Chl c and a carotenoid (probably fucoxanthin). These

peaks are probably from minor amounts of FCPs detected in this fraction (Fig. 2.6B). Variability

in the resolution of pigment-protein Complexes 4-9 was observed between different preparations

of thylakoids, possibly the result of different cell culture densities which could have altered light

conditions.

Complex 10b, detectable after longer periods of electrophoresis, was enriched in the

28 kDa polypeptide that cross reacted with the α-FCP (Fig. 2.6B). It had absorbance peaks

typically associated with Chl a, Chl c, and fucoxanthin but since it also contained some of the

lower molecular mass light-harvesting polypeptides, I cannot say whether the 28 kDa

polypeptide binds these pigments or not. This polypeptide may be analogous to the 31 kDa

pigment-protein complex containing only Chl a characterized from another chromophyte,

Ochromonas (Gibbs and Biggins 1991).

Complex 11 contained the majority of the fucoxanthin and had a low Chl a/Chl c ratio as

estimated from the peaks at 440 nm and 460 nm respectively (Fig. 2.7A). It appeared to be

similar to fraction 1 from the sucrose gradient, but was not as pure. Rather than a single

polypeptide, it had the three major light-harvesting polypeptides at 20.5, 19.5 and 18.5 kDa that

cross reacted with the diatom α-FCP antibody. Moreover, polypeptides in the 21-33 kDa range

were also weakly stained with Coomassie blue in this fraction (Fig. 2.6A). One-third the amount

of Complex 11 was loaded onto the gel in Figure 2.6 in order to prevent overloading; therefore,

comparison of absolute protein amounts can not be made. Complex 11 had a fluorescence

emission maximum at 681 nm at 77°K (Fig. 2.7) which is comparable to the LHC II of land

plants (Murata and Satoh 1986). The excitation spectrum of Complex 11 has peaks at 437 and

460 nm from Chl a and Chl c, respectively. As well, a broad area from 480 to 540 nm results

from excitation energy transfer of xanthophylls (fucoxanthin) to Chl a (Fig. 2.7C). This indicates that the accessory pigments are still coupled to Chl a. Although some D1 is immunodetected in Complexes 10b and 11, it would be premature to conclude that the 28 kDa polypeptide or any of the FCPs are preferentially associated with PS II, as it is impossible to rule out comigration of individual polypeptides in this region of the gel.

An orange free pigment zone migrated just ahead of Complex 11. Absorbance spectra indicated that it contained carotenoids and a small amount of chlorophyll (data not shown). No polypeptides were detected following Coomassie staining.

2.4   Discussion

Sucrose gradient separation following digitonin solubilization has been successfully used in the isolation of light-harvesting complexes from a number of algae (Berkaloff et al. 1990; Hiller et al. 1991; Arsalane et al. 1992). This fractionation technique usually results in a light-harvesting antennal fraction at the top of the gradient and a few additional denser pigment-protein complexes. I have found that the raphidophycean alga, *Heterosigma*, like other chromophyte algae, has a predominant fucoxanthin-chlorophyll a/c pigment-protein complex released by digitonin solubilization. This complex has a single polypeptide with an apparent molecular mass of 19.5 kDa and spectral characteristics comparable to the predominant LHC from other chromophytes (Hiller et al. 1991). It appears to be the more abundant of four predominant FCPs in the thylakoids as determined immunologically with the α-FCP antiserum. Since it is easily dissociated from the core complexes, this suggests it may be peripherally located and possibly analogous to land plant LHC II. In addition, the 19.5 kDa polypeptide is preferentially removed in most digitonin extractions, suggesting it is even more distal to the other predominant FCPs. In contrast, the FCP fraction obtained by non-denaturing SDS-PAGE

59

(Complex 11) appeared to contain all the major FCP polypeptides. The detergent used for the solubilization of thylakoids in the non-denaturing gel system were obviously more 'penetrating' and this resulted in a more vigorous extraction of the FCPs.

The denaturing SDS-PAGE system used allowed for the resolution of up to 12 separate cross-reacting LHC related polypeptides in *Heterosigma* . They were in the same size range (15-22 kDa) as those reported from other chromophytes (Hiller et al. 1991). Most published work has reported one to four light-harvesting polypeptides (Hiller et al. 1991), although as many as six polypeptides from four chromophyte species have been reported to cross react with an antibody raised to *Chlamydomonas* (Chlorophyceae) LHC (Plumley et al. 1993). Differences in the number of polypeptides detected may partly be due to the different electrophoretic systems used to resolve the complexes and the nature of the antiserum. In order to rule out the possibility that some of the immuno-reactive bands were the result of proteolytic cleavage of larger polypeptides, *Heterosigma* thylakoids were isolated and incubated at 37°C, in the presence or absence of protease inhibitors, with no difference in the number of LHC bands detected. Whole cells solubilized directly in 2X SDS sample buffer also showed the same pattern as thylakoids (data not shown). The FCP antenna family in *Heterosigma* may therefore be as complex as the CAB antenna family in land plants (Green et al. 1991; Green et al. 1992).

It is also interesting to note that these light-harvesting polypeptides were detected using LHC specific antibodies from a different class in the Chromophyta and antibodies from terrestrial plants. Other studies using antibodies specific for land plant and chlorophyte LHCs show cross-reactivity with various members of the Chromophyta (Caron et al. 1988; Passaquet et al. 1991; Plumley et al. 1993); others show cross-reactivity within the Chromophyta (Fawley et al. 1987). These results indicate the presence of commonly conserved antigenic determinants associated with all light-harvesting polypeptides, suggestive of a common evolutionary origin. This structural similarity is confirmed by sequences of the FCP genes from the diatom, *Phaeodactylum tricornutum* (Grossman et al. 1990). Protein sequences of tryptic fragments from a Cryptomonad (Sidler et al. 1988), the dinoflagellate *Amphidinium* (Hiller et al. 1993) and

60

from *Heterosigma* (Green et al. 1992, Chapter 4) also demonstrate the apparent structural similarity between the Chl $a + b$ binding proteins and the Chl $a + c$ binding proteins.

Solubilized chromophyte thylakoids have previously been fractionated by non-denaturing PAGE, especially in sodium deoxycholate (Caron and Brown 1987; Brown 1988) or Deriphat 160 gel systems (Boczar et al. 1980; Peyriere et al. 1984; Boczar and Prezelin 1989; Knoetzel and Rensing 1990). I was unable to obtain satisfactory results with either of these gel systems or with the non-denaturing gel systems used successfully with land plants (Camm and Green 1989; Thornber et al. 1991). This indicates the labile nature of these complexes as compared to the LHCs of terrestrial plants; a problem that is impeding characterizations of the number and types of chlorophyll-proteins in the chromophytes. However, a modification of the non-denaturing gel system devised by Allen and Staehelin (Allen and Staehelin 1991) proved to be successful in the separation of *Heterosigma* pigment-protein complexes, allowing the preservation of a number of large complexes with apparent molecular masses of over 200 kDa. This system represents an improvement over other electrophoretic separation techniques for chromophyte algal pigment-protein complexes, being able to separate several PS I fractions, a number of PS II fractions, and a dominant LHC fraction.

An important feature of the non-denaturing electrophoretic separation technique is the ability to isolate light-harvesting antennal proteins still associated with the core complexes. Complexes 1-3 were PS I fractions and appeared identical except that the slowest migrating appeared to have larger amounts of associated light-harvesting polypeptides. Complexes 2 and 3 lacked the majority of the LHC and tended to retain high levels of the lower cross reacting polypeptide (17 kDa) suggesting it may be closely associated with the PS I core complex. Previous studies on isolated PS I complexes of chromophyte algae found a 715-720 nm fluorescence emission peak which is usually assumed to be due to PS I reaction center in association with its light-harvesting antenna (Brown 1988; Berkaloff et al. 1990). As well, a PS I specific antenna with different fluorescence emission characteristics has been identified in the xanthophyte alga, *Pleurochloris* (Büchel and Wilhelm 1993). It appears that the presence of

61

a light harvesting complex associated with PS I is a common feature of the Chromophyta (Berkaloff et al. 1990), as is the case with green algae and land plants where a number of unique antennal proteins in the size range of 21-24 kDa are specifically associated with PS I (Mullet et al. 1980; Haworth et al. 1983; Lam et al. 1984b).

Complexes 1-3 do contain associated LHC-related polypeptides though the total amount is low compared to the overwhelming occurrence of these polypeptides in Complex 11. In addition, these polypeptides do not appear to bind significant amounts of fucoxanthin or Chl c. As these polypeptides are of comparable sizes to the other LHCs in Complex 11 and in thylakoids, it would seem that these polypeptides are partially denatured and have lost the accessory pigments without being removed from their association with the core complex. This may be a result of the detergents used, the forces exerted during electrophoresis, or a combination of both. This would not be too surprising as the LHCs are quite susceptible to degradation. Because of the possibility of a nonspecific interaction with the high molecular mass complex and an inability to distinguish between LHCs associated with this PS I fraction and those in the LHC fraction (Complex 11), I can not yet assign any specific LHC polypeptides exclusively to PS I in *Heterosigma* .

The ability to resolve a number of PS I and PS II complexes is comparable to the results obtained with the green alga, *Chlamydomonas*, (Allen and Staehelin 1991) though there are differences in the associations of the LHCs with these complexes. The number of LHCs resolved also appears to differ, illustrating the differences between the chromophytes, green algae and land plants. At the present time I am unable conclude, with any certainty, the nature of the complex organization in the thylakoid with regard to the localization of PS I and PS II or whether the pigment-protein complexes separated on the non-denaturing gel system represent different environments within the thylakoid membrane.

The recent immunocytochemical localization of FCP and PS I complexes within the thylakoids of members of the Chromophyta (Lichtlé et al. 1992; Pyszniak and Gibbs 1992; Lichtlé et al. 1992b) show that PS I and PS II are not as highly segregated within appressed and

non-appressed regions as they are in terrestrial plants. Though there was only a slight preference for PS I in the 'nonappressed' regions of the chromophyte thylakoid, the FCPs were homogeneously distributed throughout all parts of the membrane. In *Heterosigma*, there is an association of a large number of antennae polypeptides with the PS I enriched fractions from the sucrose gradient. It is unlikely that these are all specific PS I antennae as most of the prominent LHC polypeptides detected in thylakoids are present in these fractions. At the moment it is not clear which of these polypeptides may be preferentially associated with PS I to the exclusion of a PS II association. The consistent association of the FCPs with these lower fractions suggests that the separation of the main antennae is not as sharply defined as is the case with LHC I and LHC II in terrestrial plants. The homogeneous distribution of the FCPs in appressed and nonappressed thylakoid regions of chromophytes may explain their prevalent association with the PS I enriched fractions of the sucrose gradient. This would agree well with work showing that the excitation energy captured by the main antennae of a diatom was equally distributed to both photosystems (Owens 1986b). However, this remains speculative at the moment since the association of the FCPs with PS I may be a result of contamination during the fractionation procedure.

An immunological characterization of LHC related-polypeptides in red algae

3.1 Introduction

This Chapter is concerned with the immunological characterization of LHC proteins from two red algae. In this Chapter I will first describe the immunological analyses I did with *Aglaothamnion*. This will be followed by the immunological work done with *Porphyridium* in collaboration with Beth Gantt's group at the University of Maryland. Greg Wolfe, in the lab of Beth Gantt, was the first to demonstrate that a PS I fraction from the red alga *Porphyridium* contained the core complex and an array of smaller polypeptides in the 11-24 kDa range, typical of the PS I polypeptide distribution of chlorophytes and terrestrial plants (Wolfe et al. 1992; Wolfe et al. 1994b). In his work, PS I and PS II fractions from *Porphyridium* were isolated. The spectral characteristics and the immunological detection of D2, CP43 and CP47 were used to identify the PS II fraction. The discovery of a putative LHC I complex in *Porphyridium* lead to a collaboration with our lab in order to examine the immunological relatedness of these red algal chlorophyll-proteins to other antennae. This collaboration showed that these polypeptides were indeed structurally related to the CABs and FCPs of the terrestrial plants and diatoms, respectively (Wolfe et al. 1994). In addition, I examined a more distantly related red alga, *Aglaothamnion neglectum*, to look for similar immunologically related polypeptides to determine whether the CAB-related antennae were a general occurrence amongst the red algae or, alternatively, if they were unique to the primitive red algal class represented by *Porphyridium*.

Light-harvesting polypeptides resembling the CABs of the Chl a + b-containing

organisms have not been previously discovered in the red algae. Many of the earlier attempts at isolating pigment-protein complexes involved solubilization of the thylakoids with SDS followed by polyacrylamide gel electrophoresis. This frequently resulted in the release of a considerable amount of free chlorophyll, though a PS I fraction usually remained intact (Hiller and Larkum 1981; Redlinger and Gantt 1983). More recently, a PS I fraction was isolated from the unicellular red alga, *Cyanidium caldarium*, and was found to contain a high molecular mass band and four smaller polypeptides in the 13-18 kDa range (Yurina et al. 1991). In a separate study, the thylakoid membrane composition of *Porphyridium* was analyzed by fractionation of detergent solubilized thylakoid membranes on a sucrose gradient. This resulted in the separation of PS I and PS II fractions (Marquardt and Ried 1992). Neither of the above two studies reported the occurrence of LHC polypeptides associated with either PS I or PS II. The latter study used a partially-denaturing gel system to examine the thylakoid composition which would not be expected to resolve the LHC polypeptides. In addition, the lack of LHC detection in either of these studies may have been due to the degradation of the complexes as a result of the methods used to fractionate the thylakoids (Wolfe et al. 1992).

3.2 Materials and Methods

3.2.1 Aglaothamnion cultures

Aglaothamnion neglectum Feldmann-Mazoyer is a filamentous red alga (class Rhodophyceae) originally collected off the shores of Hawaii and belonging to the subclass Florideophycidae—order Ceramiales (in the family Ceramiaceae). An axenic culture of *Aglaothamnion* was provided by Dr. Kirk Apt to whom I am grateful. These cultures were maintained in the same artificial sea water medium described in Chapter 2. No difference was seen when *Aglaothamnion* was cultured in Provasoli's enriched sea water media as described by Magruder (1984), so the artificial media was used. These algae were kept on a 16 hour light: 8

hour dark cycle at 24°C and 30 µE/m^2/min, with constant bubbling of air through 0.2 µm millipore filters to maintain an axenic culture. The cultures were kept in six litre flasks with four litres of media. Bubbling air at 4000 cm^3/min with an aquarium pump was sufficient to keep the cultures agitated. In order to subculture the algae and to maintain an even growth of the tissue, the algae were fragmented in a sterilized blender for 10-15 seconds before inoculating fresh cultures.

3.2.2 Aglaothamnion neglectum thylakoid fractionation

Aglaothamnion was harvested by pouring the culture through four layers of cheese cloth. Thylakoids were prepared using a modification of a method used for *Porphyridium* (Wolfe et al. 1992). The material was quick frozen in liquid nitrogen and ground to a fine powder in a pre-frozen mortar. This powder was resuspended in cold 50 mM NaPO$_4$ pH 7.0, with the protease inhibitors (1 mM phenylmethyl sulfonyl fluoride, 5 mM ε-amino-n-caproic acid, 1 mM benzamidine-HCl, 1 mg/ml leupeptin). The extract was then put through a pre-chilled French press at ≈ 1300 psi. This was repeated three times and the effluent was kept on ice between runs. The extract was centrifuged at 1000 × g for 10 minutes at 4°C to remove unbroken cells and cellular debris. The supernatant was then centrifuged at 28 000 rpm in a SW28 rotor for 45 minutes at 4°C. The green pellet was resuspended in cold 10 mM NaPO$_4$ pH 7.6, with protease inhibitors (PIs) included in the buffer. The fraction was then centrifuged through a sucrose step gradient (0.5 M/ 0.8 M/ 1.6 M sucrose steps in 10 mM NaPO$_4$ pH 7.0, plus PIs) at 27 000 rpm in a SW28 rotor for three hours at 4°C. The pellet was removed and resuspended in cold 10 mM NaPO$_4$, 150 mM NaCl, plus PIs. If necessary, the sample was then quick frozen in liquid nitrogen and stored at -80°C for use at a later time.

Thylakoids were resuspended in 4 ml ice-cold 1 M NaBr and centrifuged at 15 000 rpm in a SS34 rotor for 10 minutes. The pellet was resuspended in 1 ml cold 50 mM HEPES pH 7.4, 10 mM NaCl, 10% glycerol, and 0.75 M sucrose (+PIs). Thylakoids were solubilized in β-Dodecylmaltoside (β-DM) at a detergent:chlorophyll a ratio of 24:1 for 2 hours at 4°C, in the

dark. Solubilized thylakoids were diluted 1:1 with 50 mM HEPES pH 7.4 (+ PIs) and loaded onto a linear 15-30% (w/v) sucrose gradient (in 50 mM HEPES pH 7.4, 10 mM NaCl, 10% glycerol, 0.04% (w/v) β-DM and PIs). The gradient was centrifuged in a SW41 rotor at 36 000 rpm for 20 hours. Pigmented bands were dialyzed against 2 × 2 liters of ice-cold 50 mM HEPES pH 7.4, 10 mM NaCl plus PIs. Samples were stored at -80°C until required. Chlorophyll concentrations were determined in N-N-dimethyl formamide using the following equation: Chl =(A_{664}) 10.65 (ε=83.9 mM^{-1}cm^{-1}) (Moran 1982).

3.2.3 SDS-polyacrylamide gel electrophoresis

Aglaothamnion thylakoid samples were denatured as described in Chapter 2. Proteins were separated on 14% acrylamide gels (acrylamide: bis-acrylamide 37.5:1) containing 0.8 M Tris-HCl pH 8.8 and 0.1% SDS. A 2 cm stacking gel containing 5% acrylamide, 0.1 M Tris-HCl pH 6.1, and 0.1% SDS was used. Gels were typically run at 18 mA for 17 hours at 4°C. Electrophoresis buffer used was the standard SDS-Tris-glycine buffer described in Chapter 2. Staining, electrotransfer to nitrocellulose and western blotting of the polypeptides were done as described in Chapter 2 (section 2.2.3). The α-D1 (from pea) and α-OEC3 (from spinach) were a gift from Ann Eastman. The α-OEC2 antiserum was donated by Dr. Ekramadoullah. The spinach α-PsaD and α-PsaF antisera (from spinach) were from Bengis and Nelson (1975).

	percent chlorophyll	absorbance maxima
I	14%	417, 435, 482*, 672 nm
II	86%	421, 437, 493, 678 nm
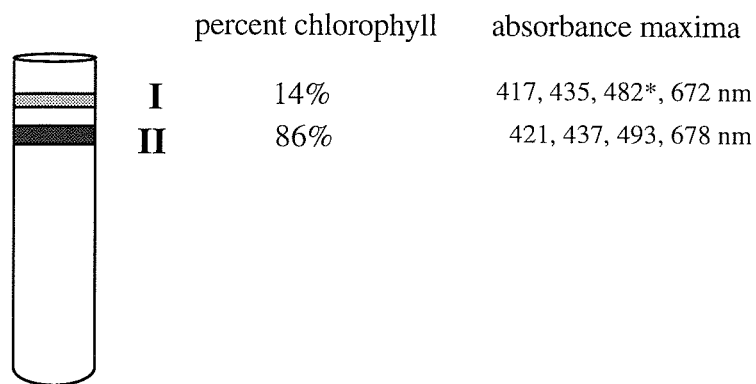
Figure 3.1.

Schematic diagram of linear 15-30% Sucrose gradient used to fractionated β-dodecyl maltoside solubilized thylakoids of *A. neglectum*. Resolved chlorophyll-binding complexes are indicated as fraction I or II. The percentage of chlorophyll and the absorption maxima of each fraction are given. An (*) indicates the presence of an absorbance shoulder.

3.3 Results

Fractionation of the thylakoids of *Aglaothamnion neglectum* on a 15-30% linear sucrose gradient resolved two green bands: band I and band II (fig. 3.1). The top fraction (I) was light green and contained only 14% of the total chlorophyll present on the gradient. The darker green bottom fraction (II) contained the majority of the chlorophyll (86%). A very small reddish-brown pellet present at the bottom of the gradient contained very little or no chlorophyll. Room temperature absorption spectra of the two fractions were taken (Fig. 3.2). Fraction I had absorbance maxima at 417 nm, 435 nm and with a long-wavelength absorbance maximum of 672 nm. A broad shoulder at 482 nm was also present. Fraction II had an absorption maxima in the 493 nm range and a long-wavelength form of chlorophyll *a* at 678 nm. Soret peaks at 437 nm and 421 nm were also observed. The distinct absorbance properties in the 480-550 nm region between the two fractions indicates differences in carotenoid distribution.
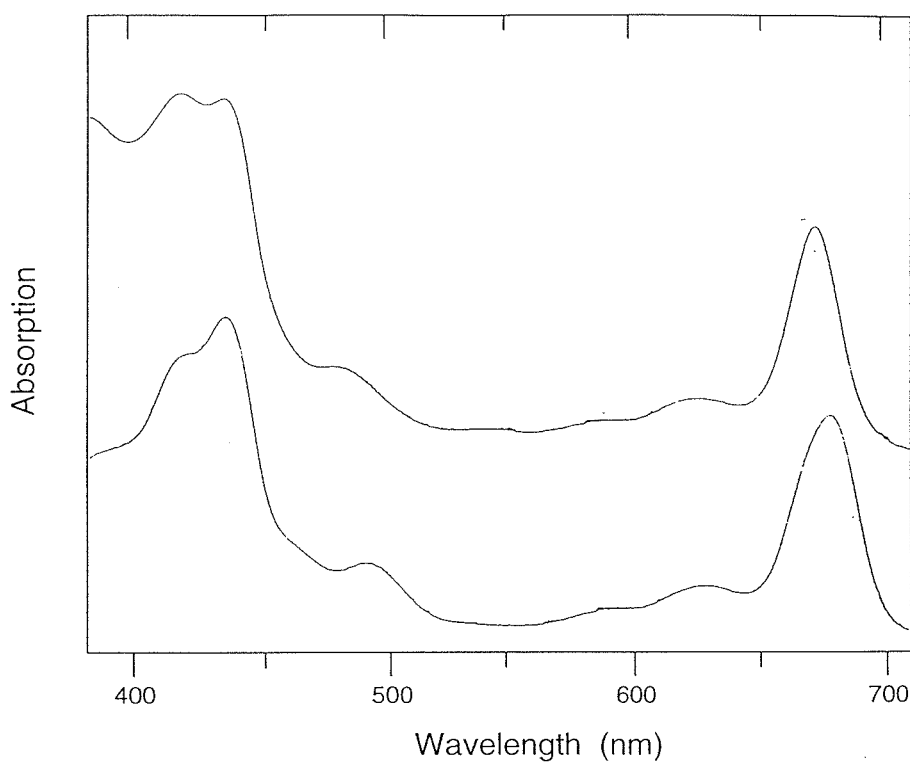
Figure 3.2

Room temperature absorption spectrum of sucrose gradient fractions I (top spectra) and II (bottom spectra). Spectra are offset for clarity.

In Figure 3.3 (panel B), the polypeptide composition of fraction II and thylakoids from *A. neglectum* are shown along with a thylakoid preparation from *Heterosigma* and a PS I fraction from spinach. Analyses were limited to fraction II and to thylakoids of *A. neglectum* because the presence of excessive amounts of detergent prevented proper resolution of the polypeptide constituents in fraction I. Polypeptides in the size range of 18-22 kDa, 30-34 kDa, and 40-100 kDa were especially abundant in the thylakoid lane (fig. 3.3 B, lane 6). Fraction II had a distinct band at 66 kDa and diffuse polypeptides in the 18-22, 30-34 and 40-46 kDa regions (fig. 3.3 B, lane 5). The LHCs of *Heterosigma carterae* in the 15-22 kDa range were the most abundant thylakoid proteins (fig. 3.3 B, lane 7; Chapter 2). In the spinach PS I preparation there were a number of polypeptides in the 14-69 kDa range but the 21-27 kDa antennae were the most prevalent (fig. 3.2 B, lane 8).

69

A western blot with the same fractions was immunostained with the α-CP1a antiserum (fig. 3.3 A). This antibody (described in Chapter 2) cross-reacted with four polypeptides in the spinach PS I enriched fraction, the two lower polypeptides belonging to LHC I (25 and 22 kDa) and the two upper polypeptides (26.5 and 27 kDa) to contaminating LHC II. This antiserum recognizes epitopes from both LHC I and LHC II polypeptides (White and Green 1987). The *H. carterae* thylakoid fraction cross-reacts with up to six polypeptides in the size range of 17-23 kDa (fig. 3.3 A, lane 3). The pattern of cross-reacting polypeptides appears a little different from similar immunoblots in Chapter 2 because a different gel system was used to resolve them.

Of particular interest was the detection of four cross-reacting polypeptides (19.0, 18.5, 18.0 and 17.5 kDa) in the *A. neglectum* thylakoid fraction (fig. 3.3 A, lane 2). In addition, there may be a fifth immunoreactive polypeptide with a molecular mass of 19.5 kDa, as the 19.0-19.5 kDa band appears to be a doublet. Fraction II contains two cross-reacting polypeptides with a size of 19.0 and 19.5 kDa. The other cross reacting polypeptides were probably removed during fractionation. It is interesting to note that these polypeptides are significantly smaller than the corresponding polypeptides in terrestrial plants and green algae (21-24 kDa). In terms of size, they more closely resemble the FCPs from the chromophytes (16-21 kDa). The α-CP1a antibody also cross-reacts with the core complex of PS I, as seen in the 66 kDa area of all four fractions (fig. 3.3 B).

Using an antibody specific to the pea D1 protein, the 32 kDa core complex polypeptide of PS II (D1) was detected in both fraction II and in the thylakoid fraction from *Aglaothamnion* (fig. 3.4, lanes 1 and 2). The D1 polypeptide was also detected in the *Heterosigma* thylakoid fraction (lane 3) but was absent from the spinach PS I preparation (lane 4). Though fraction II contains both PS I and PS II specific polypeptides, it appears to be enriched in PS I complexes.

Immunological analyses have demonstrated that D1 and CP47 (PS II core antenna) are structurally conserved in *Synechocystis* (cyanobacterium), *Prochlorothrix*, *Heterosigma*, and *Aglaothamnion* (data not shown). The immunological similarities of these polypeptides in the red alga, *Porphyridium* have also been demonstrated (Marquardt and Ried 1992). The putative
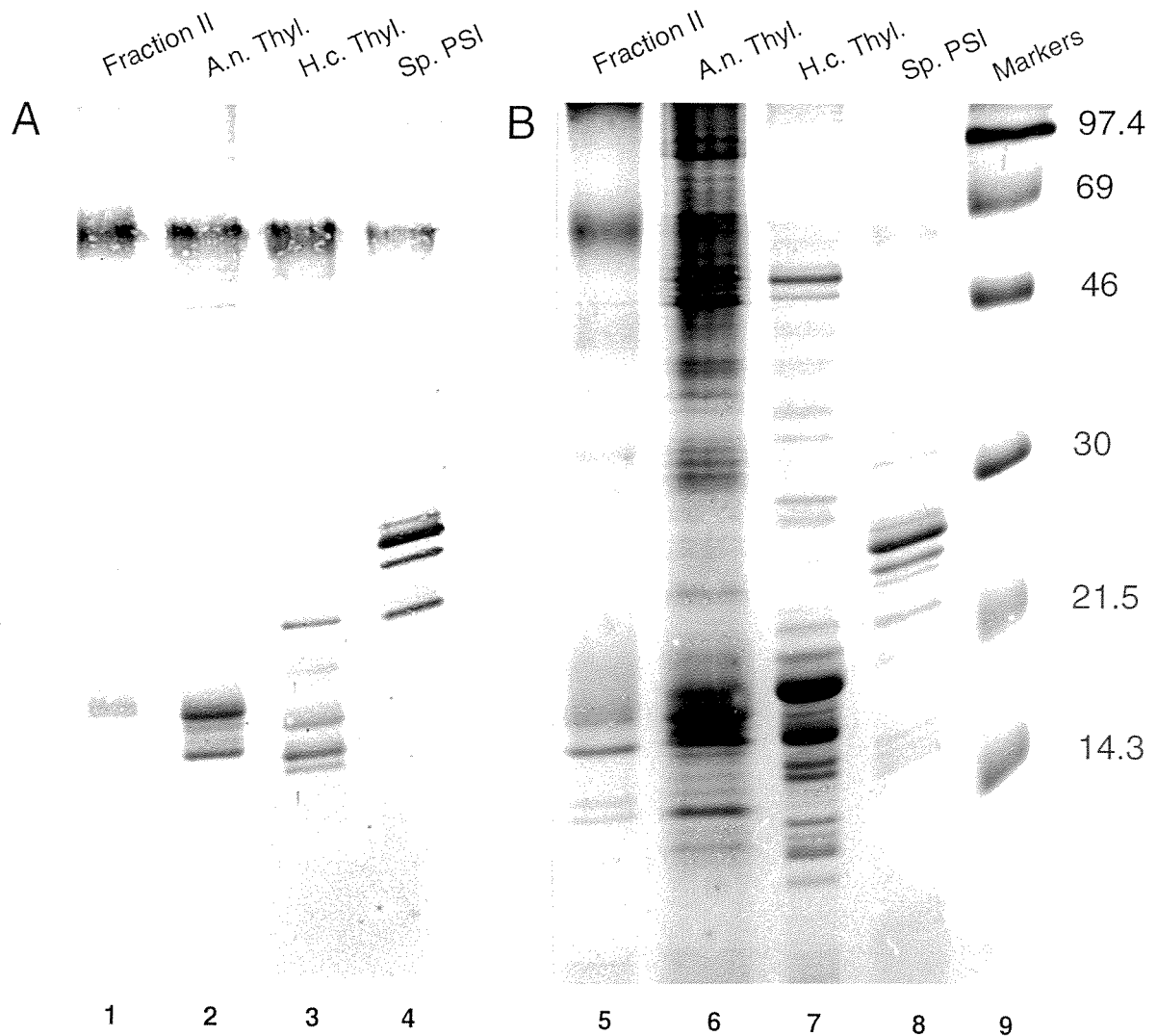
Figure 3.3

Composition and immunological analysis of polypeptides in *Aglaothamnion* thylakoids and fraction II. A) Western blot immunostained with the α-CP1a antiserum (described in text). Samples are; 1, *Aglaothamnion* fraction II; 2, *Aglaothamnion* thylakoids; 3, *Heterosigma carterae* thylakoids; 4, spinach PS I fraction. B) Gel stained with coomassie blue. Samples are; 5, *Aglaothamnion* fraction II; 6, *Aglaothamnion* thylakoids; 7, *Heterosigma carterae* thylakoids; 8, spinach PS I preparation; 9, markers. Size of markers in kDa are indicated on the right.

A.n. A.n. H.c. Sp. M
F II Thy. Thy. PSI

⟋ 30 KDa

1   2   3   4   5

α-D1
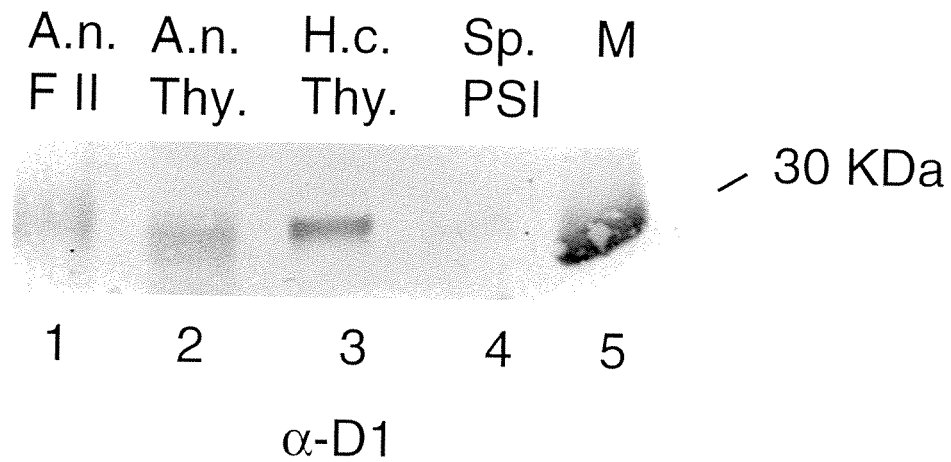
Figure 3.4

Immunological detection of the D1 protein in *Aglaothamnion* sucrose gradient fraction II (lane 1), *Aglaothamnion* thylakoids (lane 2), *Heterosigma* thylakoids (lane 3) and in a spinach PSI fraction (lane 4). The molecular mass marker (30 kDa) is in lane 5. An anti-D1 antiserum (a-D1) derived from spinach is used.

ferredoxin docking and PsaC stabilizing protein, PsaD (PS I subunit II, *psa*D), was also present and immunologically reactive to a spinach PsaD specific antibody in *Synechocystis*, *Heterosigma*, and *Aglaothamnion*(fig. 3.5, *psa*D). However, there were considerable variations in size (14-22 kDa) and the reactions were weak in the *Aglaothamnion* and *Heterosigma* thylakoid lanes (fig. 3.5, *psa*D). In contrast, only *Heterosigma* and *Aglaothamnion* thylakoids showed a cross-reaction with an antibody directed to the putative plastocyanin docking protein, PsaF (PS I subunit III), of spinach (fig. 3.5, *psa*F). A polypeptide immunologically related to PsaF was not detected in *Synechocystis*, though it may have been removed during isolation since it is extrinisically associated with the membrane.
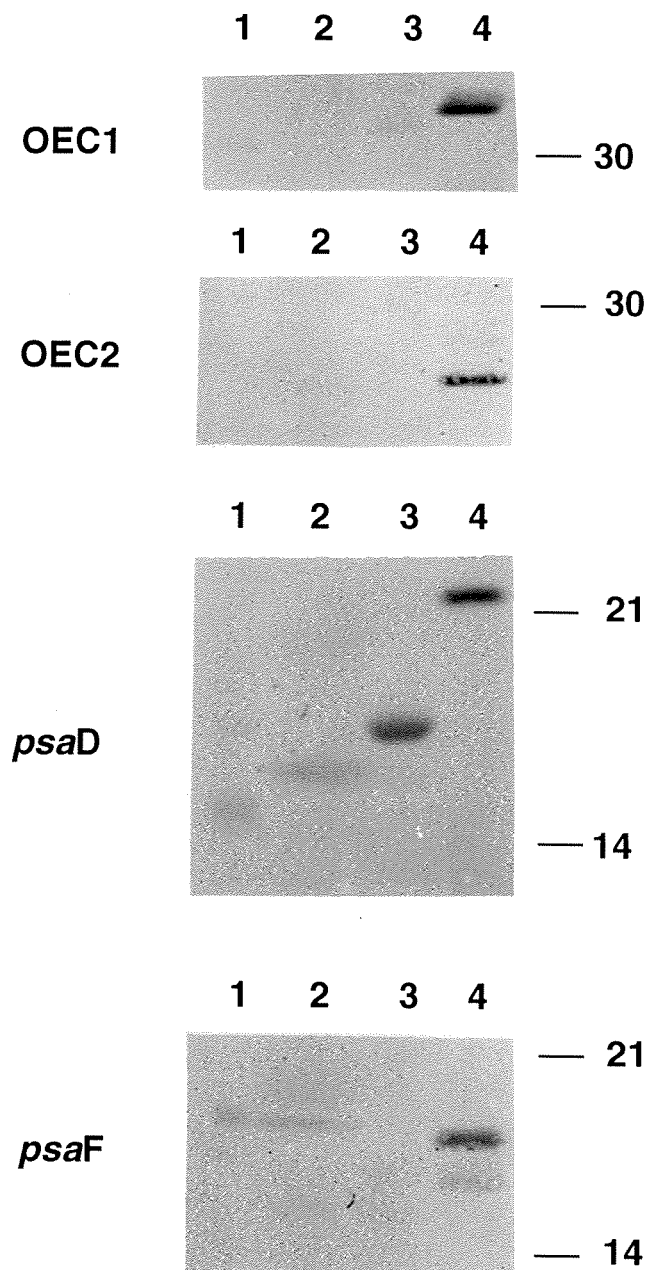
Figure 3.5

Immunological analyses of thylakoids from *Heterosigma* (lane 1), *Aglaothamnion* (lane 2), *Synechocystis* (lane 3), and spinach (lane 4). Western blots were immunostained with the anti-OEC1 (oxygen evolving complex 1), the anti-OEC2 (oxygen evolving complex 2), the anti-*psa*D (PS I subunit II) and anti-*psa*F (PS I subunit III) antisera as indicated. Molecular mass markers are indicated on the right.

Western blots were done to examine the presence or absence of polypeptides making up the oxygen evolving complex (fig. 3.5). This was attempted immunologically using spinach derived antibodies specific to the OEC1 (33 kDa) and OEC2 (23 kDa) polypeptides. Since the OEC2 polypeptide is not present in cyanobacteria (Stewart et al. 1985), it was thought that an examination of the red algae for the presence or absence of this complex would provide some useful phylogenetic information. The OEC1 polypeptide was only very weakly detected in *Synechocystis* and *Heterosigma* and this antibody did not react significantly with anything in the *Aglaothamnion* thylakoid fraction (fig. 3.5, OEC1). In addition, only the spinach thylakoid fraction cross-reacted with the OEC2 antibody, even though four times the normal amount of chlorophyll was loaded on the gel (fig. 3.5, OEC2). Unfortunately, I am unable to conclude if *Aglaothamnion* or *Heterosigma* have a protein homologous to the OEC2 polypeptide as the lack of immunological cross-reaction could be due to sequence divergence and not complete absence.

Greg Wolfe et al. (1992, 1994) at the University of Maryland analyzed the pigment-protein complexes of the unicellular red alga, *Porphyridium cruentum*. They were the first to demonstrate the presence of a putative chlorophyll-binding complex analogous to LHC I of terrestrial plants. In a collaboration with this group, we were able to analyze immunologically the structural relatedness of these polypeptides to those from the CAB and FCP family. This alga is a member of the class Bangiophycidae (order Porphyridiales), which is considered to be primitive with respect to the Florideophycidae, of which *Aglaothamnion* is a member. Figure 3.6 shows the results of immunoblotting two identical gels with different antisera; the α-CP1a polyclonal antibody and the α-FCP antibody (described in Chapter 2). At least five polypeptides in the *P. cruentum* PS I fraction cross-reacted with the α-CP1a antibody; they had apparent molecular masses of 19.5, 20, 22, 23 and 23.5 kDa (Fig. 3.6, panel B, lane 3). There are two additional polypeptides cross-reacting in the thylakoid lane at 19 and 20.5 kDa (Fig. 3.6, panel B, lane 4). The spinach PS I fraction showed a number of cross-reacting polypeptides (22-29 kDa) from LHC I and contaminating LHC II and CP29 (Fig. 3.6, lane 2, panel B) as the sample was overloaded. A major cross-reacting band at 65 kDa in all four lanes is the core complex of PS I. It is significant that thylakoid proteins from the cyanobacterium, *Nostoc*, lack
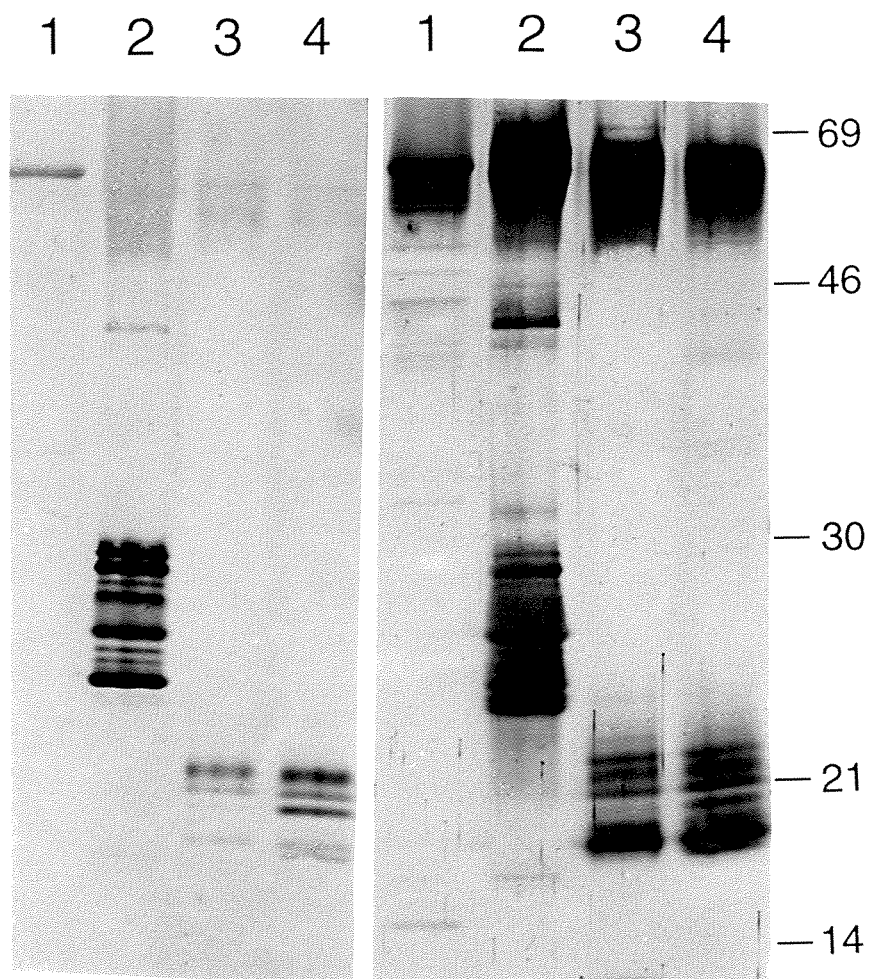
Figure 3.6

Immunological analysis of polypeptides from: 1, *Nostoc* thylakoids; 2, spinach PS I fraction; 3, *Porphyridium cruentum* PS I fraction; 4, *Porphyridium cruentum* thylakoid fraction. Molecular mass markers are on the right. Antisera used included the α-FCP antiserum specific for a diatom FCP complex (panel A) and the α-CP1a antiserum (panel B).

75

cross-reactivity with the α-CP1a antiserum in the size range typical for a LHC (Fig. 3.6, lane 1, panel B).

A second antibody, α-FCP, was used to assess relatedness of the low molecular mass polypeptides of *Porphyridium* to the FCPs of the chromophytic algae (Fig. 3.6, panel A). The α-FCP antibody cross-reacted with five polypeptides from the *P. cruentum* thylakoid fraction (Fig. 3.6, panel A, lane 4). Interestingly, the 19 and 19.5 kDa bands appear less immunologically reactive to this antiserum compared to their reaction with the α-CP1a antiserum. The 20 and 23.5 kDa polypeptides did not appear to cross-react at all with the α-FCP antiserum. These observations indicated a degree of structural divergence between some of the polypeptides. At least three polypeptides (19.5, 22, and 23 kDa) in the *P. cruentum* PS I fraction cross-reacted with the α-FCP antiserum (Fig. 3.6, panel A, lane 3) . Again, the 20.5 kDa polypeptide in the PS I complex was only weakly detected and must have been removed during purification.

These results show that the complex pattern of polypeptides immunologically related to the LHCs in the unicellular red alga, *Porphyridium*, is comparable to that in the filamentous red alga, *Aglaothamnion*, except for differences in the apparent size of the polypeptides. This may be due to the use of a different SDS-gel system. No polypeptides in the 18-23 kDa range were detected in the thylakoids of the cyanobacterium, *Nostoc*, with the α-FCP antiserum (Fig. 3.6, panel A, lane 1). This supports the idea that the cyanobacteria do not possess intrinisic antennae related to those of algae or terrestrial plants. However, a large molecular mass polypeptide (69 kDa), of unknown identity, was weakly immunostained with the α-FCP antiserum.

Thylakoid and PS I fractions from the Chl *a* + *b*-containing prochlorophyte, *Prochlorothrix hollandica* (a gift from Georg van der Staay), were immunoblotted with the barley α-CP1a antiserum (Fig. 3.7, lanes 1 and 2) to search for evidence of any CAB related LHCs. Neither of these fractions showed any immunologically reactive polypeptides in the size range of a typical LHC (18-34 kDa). However, the CAB polypeptides were detected in the spinach thylakoid control, as expected (fig. 3.7, lane 3). In both the thylakoid and the PS I fractions of *Prochlorothrix*, the core complex of PS I was detected at 65 kDa.

76
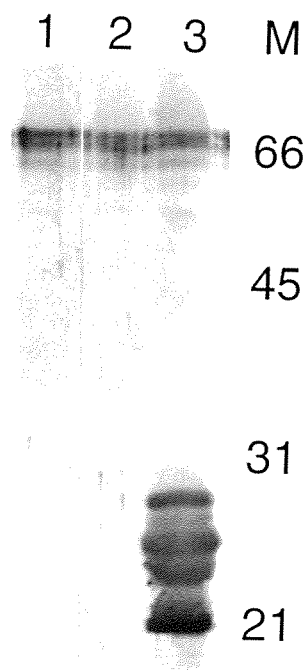
Figure 3.7

Immunological analysis of *Prochlorothrix hollandica* PS I (1) and thylakoid (2) fractions. A spinach control (thylakoid fraction) is also shown (3). Molecular mass standards (M) are shown on the right. The blot was probed with the α-CP1a antiserum.

3.4 Discussion

Analyses with the unicellular alga, *P. cruentum*, demonstrated that there are several polypeptides that are immunologically related to both the CABs and FCPs (Wolfe et al. 1994). Furthermore, some of these are specifically associated with a purified PS I complex (Wolfe et al. 1992; Wolfe et al. 1994). This is significant as it is the first demonstration of such membrane intrinisic light-harvesting antennae within the Rhodophyta. Despite recent analyses of red algal thylakoid membrane polypeptide composition, CAB-related LHCs have not been previously reported. The LHCs in *Porphyridium* also cross-react to different extents with the CAB and FCP-specific antisera. This indicates that though the different polypeptides share common immunologically reactive epitopes, there is structural variability between these antennal proteins; this may be related to different functions. These observations suggest closer evolutionary relationships between the red, green and chromophytic plastids than previously thought. Recent protein sequencing data from these polypeptides have confirmed their relatedness to the LHC I proteins of terrestrial plants (Beth Gantt, unpubl. data).

When I examined a second red alga, *Aglaothamnion neglectum*, up to five polypeptides in the thylakoid membrane fraction were antigenically related to the Chl $a + b$-binding proteins of terrestrial plants. Fewer cross-reacting polypetides were found in the sucrose gradient fraction, probably a result of selective loss during the detergent solubilization procedure. It has been immunologically determined that fraction II from the sucrose gradient contains polypeptides from both PS I and PS II; therefore, it is not possible to conclude whether they are associated specifically with PS I or PS II. By analogy with *Porphyridium*, I would expect an association of these LHCs with PS I in *A. neglectum*. However, this does not exclude the possibility that CAB/FCP-related polypeptides are associated with PS II. In both red algae, there were fewer detectable LHCs in the detergent solubilized fractions as compared to the thylakoid fractions. This potential for degradation may be one reason these polypetides were not previously reported.

It is significant that CAB/FCP-related polypeptides were found in representatives of both

the Porphyridiales (*Porphyridium*) and the Ceramiales (*Aglaothamnion*), as they represent very diverse lineages separated by large evolutionary distances (Garbary and Gabrielson 1990). This suggests that these LHC-related polypeptides are not limited to a specific taxon; likely being a common feature of the red algae. I suspect that similar polypeptides will be detected in other red algal orders.

Plastids are thought to have evolved from endosymbiotic cyanobacteria and/or prochlorophytes; therefore, it was significant to find a lack of CAB/FCP-related polypeptides in *Nostoc* and *Prochlorothrix*. This is particularly significant with the prochlorophytes as it has been demonstrated that they possess Chl $a + b + c$-binding antennae. The lack of immunological relatedness of these antennae has been demonstrated by others (Hiller and Larkum 1985; Bullerjahn et al. 1990). In addition, preliminary sequence information from the prochlorophytes, *Prochloron* (Hiller & Larkum, unpubl.), *Prochlorococcus* (LaRoche & Partensky, unpubl.) and *Prochlorothrix* (van der Staay & Green, unpubl.) shows that the 34 kDa antennae are related to the *isi*A gene product and to the inner Chl a antenna, CP43. There were no similarities to the CABs or the FCPs. This proves that the prochlorophyte LHCs evolved independently of the CABs and FCPs.

The presence of CAB/FCP-related polypeptides in the red algae and their absence in the cyanobacteria and prochlorophytes is significant in terms of both the evolution of the light-harvesting antennal proteins and the evolution of the chloroplast. The significance with regard to the evolution of the light-harvesting proteins as an extended gene family will be dealt with in Chapter 5. Much of the remaining discussion will deal with the implications of these observations for the different hypotheses of chloroplast evolution (discussed in section 1.6), especially with regards to the red and green algae. In analyzing chloroplast evolution, both chloroplast and nuclear characters can be used (see table 3.1). Plastid characters, including plastid encoded genes and plastid localized gene products encoded in the nucleus (such as the CABs and FCPs), were probably present in the original endosymbiont and comparisons will normally reflect the ancestry of the chloroplast. Cytoplasmic/nuclear characters will generally reflect the ancestry of the original phagotrophic host. Comparing evidence of both types is

necessary to develop a better understanding of the possible evolutionary pathways leading to the chloroplast.

Many of the plastid encoded characters point toward a monophyletic scheme of chloroplast evolution. These include the plastid 16s rRNA, the *psbA, tufA* and *atp*B genes (see table 3.1) which have been reviewed by Morden et al. (1992). Recent phylogenetic analysis of the plastid encoded *atp*B gene (H^+-ATPase, β-subunit) suggests that there is a deep division between the green and non-green algal chloroplast lineage's, though they cluster together within the cyanobacterial line, indicating a single origin for the chloroplast (Kowallik 1993; Douglas and Murphy 1994).

Table 3.1   Some genes commonly used to infer phylogenetic relationships amongst photosynthetic organisms

gene	protein /function	location
*psb*A	D1-core PS II reaction centre protein	Chloroplast
*atp*B	β-subunit of chloroplast ATPase	Chloroplast
*rbc*L	rubisco large subunit	Chloroplast
*rbc*S	rubisco small subunit	Chloroplast /nuclear
*tuf*A	elongation factor Tu	Chloroplast
16s rRNA	small subunit rRNA	Chloroplast
*Gap*A/B	glyceraldehyde-3-phosphate dehydrogenase	nucleus
18s rRNA	small subunit rRNA	nucleus

However, there is evidence that is consistent with a polyphyletic interpretation of chloroplast evolution. A seven amino acid deletion in the C-terminus of D1 (*psb*A gene) in *Prochlorothrix*, the green algae and in the terrestrial plants but not in the D1 protein of cyanobacteria, red algae and the chromophytes is suggestive of a polyphyletic origin of the chloroplast (Morden and Golden 1989; Golden et al. 1993). This emphasizes a character shared by a prokaryotic/ chloroplast pair containing Chl *a* and *b* to the exclusion of another prokaryotic/

80

chloroplast pair containing PBSs or Chl $a + c$ —satisfying the criterion for a polyphyletic origin. However, it has recently been reported that the prochlorophyte, *Prochloron didemni*, lacks this deletion, indicating that it is not a common occurrence amongst the prochlorophytes (Lockhart et al. 1993).

It is reasonable that insertion/deletion (indel) events would be considered good phylogenetic indicators as two independent deletion events in an identical area seems unlikely. However, I argue that the significance of this particular indel may have been over estimated. Besides having representative sequences from relatively few taxa, there are two known variations in the D1 indel location; one in the green alga, *Chlamydomonas reinhardtii* (Erickson et al. 1984), and the other in a euglenophyte, *Euglena gracilis* (Karabin et al. 1984). In these examples, there is an eight and 16 amino acid deletion at the C-terminal end of D1, respectively. The fact that these algae possess additional or separate deletions different from the other Chl $a + b$-containing organisms suggests that there are lower constraints on the 9-16 amino acids at the C-terminal end. Significantly, this region of D1 is post-translationally removed during processing into the mature form. In addition, the C-terminal region is neither required for protease binding nor for recognition (Taguchi et al. 1993). Phylogenetic analysis of the whole *psb*A gene is consistent with the monophyletic view of chloroplast evolution with *Prochlorothrix* clustering with the cyanobacteria, separate from the eukaryotic algal groups (Morden et al. 1992). All things considered, it seems that the deleted region has an ambiguous evolutionary history of questionable importance. Furthermore, based on the accumulation of molecular evidence from plastid/cyanobacterial characters, a direct evolutionary relationship between the prochlorophytes and the green algal chloroplasts are now considered less likely (Turner et al. 1989; Morden et al. 1992; Golden et al. 1993; Swift and Palenik 1993).

Plastid characters that do offer strong evidence for a polyphyletic origin of the chloroplast are the plastid encoded *rbc*L and *rbc*S (rubisco large and small subunit) data which clearly show the Chl $a + b$-containing eukaryotes in one lineage and the red algae and chromophytes in the other. In these studies, the cyanobacteria are closely related to the Chl $a + b$-containing lineage, while the most similar prokaryotic ancestor to the red algal/

chromophyte line is the β-purple bacterium, *Alcaligenes eutrophus* (Douglas et al. 1990).
However, due to the lack of phycobilisomes, Chl *a*, and oxygenic photosynthesis, a β-purple
bacterium is not a logical choice for a hypothetical chloroplast ancestor. Several hypotheses
have been put forth to explain the relatedness between the β-purple bacterial and chromophyte
*rbc*L/S sequences: (1) lateral gene transfer of the *rbc*L/S operon from a purple bacterium to an
alga leading to the red algal/chromophyte lineage (Boczar et al. 1989b); (2) the possibility that
two organisms provided genes to the phagotrophic host (Assali et al. 1990); (3) since some
purple bacteria possess two *rbc*L genes, it has been suggested that the original eukaryotic
photoautotroph had two copies of the *rbc*L/S operon with the green algal and chromophyte
lineage's each retaining a different copy; (4) finally, transfer of *rbc*L from an α-purple bacterium,
the proposed mitochondrial ancestor, to the chloroplast ancestor has also been suggested (Martin
et al. 1992). The significance of the *rbc*L data will remain an issue that will need to be resolved
through the analysis of other data.

The evidence supporting or refuting either hypothesis of chloroplast evolution (as
discussed in section 1.6 and above) has been plagued by conflicting and ambiguous evidence.
Deciding between alternative hypotheses of chloroplast evolution can be done by applying the
following two criteria, as emphasized by Reith and Munholland (1993): first, the polyphyletic
view of chloroplast evolution would be supported when characters (morphological or molecular)
are found in one chloroplast type that are shared with a putative prokaryotic ancestor, to the
exclusion of another chloroplast type-prokaryote pair. Second, if chloroplast evolution is to be
deemed monophyletic then the different chloroplast types should all share a certain trait or
character, which is more related to each other than either is to a prokaryotic (cyanobacterial)
ancestor (Reith and Munholland 1993). This trait or character could then be interpreted as
having been derived within a single lineage, following the primary endosymbiotic event that lead
to the first chloroplast containing eukaryote.

The immunological relatedness of the red algal polypeptides to the CABs and FCPs is
significant because it provides evidence that links organisms possessing the three major antennal
systems: the Chl *a* + *b*-containing antennae, the Chl *a* + *c*-containing antennae, and the

phycobilisomes. Previously, these characters formed the basis of an algal taxonomic system that separated the Chl $a + b$ containing green algae (Chlorophyta), the Chl $a + c$ containing chromophytic algae (Chromophyta), and the PBS containing red algae (Rhodophyta) into major divisions. This finding also demonstrates that the thylakoid membrane intrinisic LHCs and the soluble phycobilisome types of antennal systems can be present in the same organism.

The lack of detectable, immunologically related LHCs from the cyanobacterium, *Nostoc*, and the prochlorophyte, *Prochlorothrix*, suggests that the intrinisic light-harvesting antennae evolved after the primary endosymbiotic event that gave rise to a photosynthetic eukaryote. The presence of CAB/FCP related LHCs in all eukaryotic algae, to the exclusion of the putative chloroplast ancestors (cyanobacteria and prochlorophytes), is consistent with the monophyletic view of chloroplast evolution because it seems unlikely that structurally similar proteins would evolve independently in the red, green and chromophyte algae from prokaryotic precursors. These data also support the idea that both the PBS and the CAB/FCP antennal systems existed in the first photosynthetic eukaryote from which the green algae and red algae diverged. Divergence of the intrinisic LHCs with the gain of Chl c (in the chromophytes and prasinophytes) and the loss of PBSs (in the chlorophytes and chromophytes) could explain the differences in antennal systems presently observed (Cavalier-Smith 1982).

There can be alternative interpretations or hypotheses invoked to explain the LHC distributions between the prokaryotes and eukaryotes. One alternative is that the cyanobacteria and prochlorophytes both contained CAB/FCP-related antennae but they were selectively lost over time while being retained in the immediate chloroplast ancestor (Bryant 1992). Such a scenario does not necessarily exclude a monophyletic view of chloroplast evolution. Alternatively, the different LHCs may have evolved independently following separate chloroplast acquisitions (polyphyletic) leading to the major algal divisions. However, this alternative requires the acceptance of additional evolutionary steps and seems less likely. Nevertheless, the independent evolution of LHC proteins from related cyanobacterial precursors is possible and could explain the overall low degree of similarity between the CABs and FCPs. Some possible molecular sources from which the LHC antennae may have evolved are discussed

in section 5.4.6 (Chapter 5).

The acceptance of a monophyletic chloroplast origin requires the assumption that the red and green algae are related and have diverged from a common ancestor. Such a relationship was proposed by Cavalier-Smith (1981, 1987) who grouped the red and green algae together in the kindom Plantae. In order to demonstrate a evolutionary relationship between the red and green algae it will be necessary to show similarities between nuclear encoded characters in addition to the chloroplast characters.

More recently, the nuclear encoded (but chloroplast localized) glyceraldehyde-3-phosphate dehydrogenases (*GapA/B*) from two different red algae have been sequenced and phylogenetic analyses suggests that the red and green algae form a monophyletic group with a cyanobacterial ancestor (Zhou and Ragan 1993; Liaud et al. 1994). These studies also indicate that the green algal and red algal lineages separated very early in evolution. Moreover, the transit peptide from the *GapA/B* precursors in red algae resembles the chloroplast targeting transit peptides of terrestrial plants (Zhou and Ragan 1993; Liaud et al. 1994). *In vitro* transport studies with the transit peptide of the nuclear encoded γ-subunit of phycoerytherin in *Aglaothamnion* has shown that it can direct rubisco (*RbcS*) into pea chloroplasts (Apt et al. 1993), suggesting the red and green algae have a similar chloroplast import mechanism. This may be more easily explained by a monophyletic origin of the red and green algal chloroplasts due to the perceived difficulty in acquiring import capabilities (Cavalier-Smith 1982; Cavalier-Smith et al. 1994b).

A better idea of the relationships between the red and green algae can be achieved through the analysis of nuclear encoded characters from the two groups. This can give information on the nature of the eukaryotic 'host' for the chloroplast. The most widely utilized nuclear character to date has been the small subunit rRNA (SSU, 18s) (Bhattacharya et al. 1990; Douglas et al. 1991; Hendricks et al. 1991; Maier et al. 1991; Cavalier-Smith et al. 1994b) . The nuclear encoded large subunit rRNA (LSU, 23s) has been used infrequently (Perasso et al. 1989) due to its larger size and the smaller dataset. Most of these studies did not support a common origin of the red and green algae. A recent maximum likelihood analysis of nuclear encoded

SSU rRNA sequences, however, showed that the red and green algae grouped together on the same branch as would be expected for a monophyletic origin (Cavalier-Smith et al. 1994b). This grouping does not occur when a distance matrix method is used, which was the case in all the studies mentioned above (Cavalier-Smith, pers. comm.; Cavalier-Smith et al. 1994). Comparing the studies mentioned above, the relationships between the red algae and other groups were not consistent. This was probably due to the inclusion of different taxa, the number of characters included in the analysis and the limitations of the method used. Many of these studies also did not give an estimate of the tree node reliability, so it is hard to determine how consistent the deeper branches of these trees are.

Overall, there is accumulating evidence suggesting a monophyletic origin of the chloroplast, though the question is far from settled. If true, this would mean that the red and green algae share a common ancestor and that the acquisition of a chloroplast from a prokaryotic ancestor occurred only once. Divergences over a long period of time would have resulted in the differences between the two groups today. The presence of CAB and FCP-related LHCs in the red algae provides some evidence for the monophyletic origin of the different chloroplast types, though other alternative evolutionary pathways can be envisaged. At the moment, there is insufficient evidence to clearly decide whether chloroplasts with two surrounding membranes have arisen through one or multiple endosymbiotic events. The analysis of more characters from a diverse array of red, green and chromophytic algae along with cyanobacteria should help to determine relationships amongst these groups and help to resolve conflicting interpretations of some data.

CHAPTER 4

Characterization of Fcp *cDNAs from* Heterosigma carterae

4.1 Introduction

This section examines cDNAs encoding the FCP family of proteins in *Heterosigma*. At the beginning of this project only a single chromophyte algal FCP sequence had been characterized (Grossman et al. 1990). As the chromophytes consist of many distinct phyla, I thought this research would begin to fill a gap in the literature and aid in our understanding of the relationships between the different antennae and their diversity. As well, it would be an important piece of information for assessing the evolutionary relationships amongst the FCPs and the CABs of the terrestrial plants and green algae. In this chapter I will first describe the cloning and sequencing of a number of cDNAs encoding FCPs. I will then examine the complexity of the *Fcp* gene family and concentrate on the relationships of the *Heterosigma* sequence to other FCPs and the CABs. A cDNA library was constructed and screened with a nucleic acid probe in order to clone cDNAs encoding the FCPs. This will complement the protein characterization work in Chapter 2. In addition, I was interested in the processes involved in the targeting and translocation of proteins into the chloroplast. As these organisms possess two additional membranes around the chloroplast, an examination of the leader sequences of nuclear encoded, chloroplast localized precursors should provide some clues on the nature of the transport mechanism in the chromophytes.

4.2 Materials and Methods

4.2.1 Tryptic fragment sequencing

Maintenance of *Heterosigma* cultures and preparation of thylakoids have been previously described in Chapter 2. Sixty µg of chlorophyll (four lanes at 15 µg/lane) from sucrose gradient fraction one (F1, Fig. 4.2; chapter 2) was separated on a 12-17% gradient gel (Chapter 2) and transferred to nitrocellulose (Biorad) in 50 mM sodium acetate pH 7.0 at 200 mA, for 18hours at 4°C. The transferred band was stained with amido black (0.1% amido black, 45% H_2O, 45% methanol, and 10% acetic acid) and destained in 50% methanol/10% acetic acid. The protein band was cut out from the membrane, digested on the support with trypsin (Aebersold et al. 1987) and separated by narrow-bore reverse phase HPLC on a Waters peptide analyzer equipped with a Vydac C-4 column. Individual peptides were collected manually and sequenced using standard pulsed-liquid phase or solid-phase sequencing procedures (Aebersold et al 1990).

4.2.2 Heterosigma DNA and RNA isolation

Genomic DNA was isolated from 4 liters of late log phase cells. Cells were harvested by centrifugation at 1500 × g for 10 minutes. The pellet was resuspended in 2 ml extraction buffer (0.05 M Tris-HCl pH 8.0 and 0.05 M EDTA pH 8.0) that was added while vortexing gently. 16 ml of lysis buffer (2% sarkosyl, 10 mM EDTA pH 8.0, and 20 mM Tris-HCl pH 7.6) was then added and gently mixed. Proteinase K (8 mg) was added and swirled at 37°C for 1 hour to break down proteins. Four ml of 5 M NaCl was added and mixed. This mixture was extracted with an equal volume of Tris equilibrated phenol (pH 7.6) for 30 minutes at room temperature. Following centrifugation at 9000 × g for 20 minutes, the supernatant was extracted with an

equal volume of chloroform:butanol (4:1), centrifuged as above and repeated. The DNA was ethanol precipitated with an equal volume of 95% ethanol and spooled out with a glass rod. The DNA was resuspended in T.E. pH 8.0 and re-extracted twice with phenol:chloroform, as above. The supernatant was ethanol precipitated and resuspended in a total of 11 ml T.E. pH 8.0. DNA was purified by CsCl density gradient ultracentrifugation, using standard protocols (Sambrook et al. 1989).

Total RNA was isolated from 4 liters of log phase cells of *Heterosigma* using a guanidinium thiocyanate extraction method (Chromczynski and Sacchi 1987). Cells were harvested at 1500 × g for 10 minutes and resuspended in guanidinium solution (4 M guanidinium thiocyanate, 25 mM sodium citrate pH 7, 0.5% sarcosyl, and 0.1% β-mercaptoethanol) at 1 ml per 0.1 g cells, followed by hand homogenization in a 5 ml Potter-Elvehjem homogenizer. The following solutions were then added to the homogenized extract: 1/10th volume of 2 M sodium acetate pH 4, 1 volume water saturated phenol, and 1/5th volume chloroform:isoamyl alcohol (49:1). The sample was mixed following each addition. The sample was vigorously vortexed for 10 seconds, cooled on ice for 15 minutes and centrifuged at 10 000 × g for 20 minutes. An equal volume of isopropanol was added to the supernatant in order to precipitate the nucleic acids. After centrifugation at 10 000 × g for 20 minutes the pellet was resuspended in 1.5 ml guanidinium solution and again precipitated with an equal volume of isopropanol. Precipitated samples were resuspended in T.E. and stored at -80°C. The poly A$^+$ mRNA used in the construction of the cDNA library was isolated from total RNA using an oligo (dT) cellulose (Pharmacia) column (Sambrook et al. 1989).

4.2.3 cDNA library construction and screening

The cDNA library was constructed from 5 μg of poly A$^+$ mRNA. The poly A$^+$ mRNA was considered to be of high quality since *in vitro* translation using a wheat germ system (Promega) yielded many proteins in the 14-70 kDa range. The library was constructed using the

lambda ZAP II kit from Stratagene according to the manufacturers instructions. The synthesis of cDNA was initiated with a poly-T primer with an Xho I adapter and after the second strand synthesis, the ends were blunt ended and EcoR I linkers were ligated to the cDNA. Directional cloning was achieved by ligating the cDNAs into an Xho I/EcoR I cut λZAP vector.

Screening of the amplified library was done using a nucleic acid probe created through a polymerase chain reaction. The amino acid sequence information from the tryptic fragments (Fig. 4.2) was used to create degenerate primers for amplification of a *Fcp*-specific gene product. The degenerate primers were based on tryptic fragments T1 (TVEIK) and T3 (YDLAGDQ) as shown in Figure 4.2. This would give an amplified fragment of approximately 100 bp if the alignments with the diatom FCP were correct. The primers included a nonspecific adapter sequence with an Xba I or Pst I restriction site in the 5' region (indicated in bold). The sequences of the primers used were as follows: 5'-**GCTCTAGA** AC(G/C/T/A) T(T/G/C/A) GA(A/G) AT(T/C/A) AA(A/G) CA(T/C) GG-3' for P1 and 5'-**CGCTGCAG** TG(A/G) TC(A/G/C/T) CC(A/G/C/T) GC(A/G/C/T) A(G/A)(G/A) TCA (A/G)A -3' for P2. The 36 cycle PCR reaction profile was as follows: cycle 1→4, 95°C-2 min., 52°C-1.5 min., and 72°C-15 sec.; cycle 5→24, 95°C-1 min., 52°C-1.5 min., and 72°C-30 sec.; cycle 25→36, 95°C-1 min., 52°C-1 min.; and 72°C-1 min.. The reaction included 2.0 mM MgCl$_2$, 100 pmol of each primer (P1 and P2), 50 mM KCl, 10 mM Tris-HCl pH 8.3, 0.2 mM dNTPs and 0.5 mg *Heterosigma carterae* genomic DNA as the target sequence. The reaction was done under mineral oil with 0.5 units of Taq polymerase (AmpliTaq, Perkin-Elmer-Cetus). The reaction mixture was extracted with phenol-chloroform (1:1), ethanol precipitated and run on a 4% Nusieve agarose gel (Mandel Sci.) along side ΦX174-Hae III markers (NEBL). A 118 bp fragment was isolated using Whatman DE-81 paper and directly labeled with α-^{32}P-dCTP (3000 mCi/mmol Amersham) using a random primer labeling kit (BRL), according to the manufacturers instructions. The cDNA library was screened at high plaque densities (10 000 pfu/plate) for the first round of screening. Further rounds of screening were done at low densities so that individual plaques were well separated. Plaque lifts, denaturation, fixing and hybridization were

done as descibed in the Stratagene cDNA kit instruction manual and according to Sambrook et al. (1989).

Prospective positive clones went through three to four rounds of screening until plaques were homogeneous. The recombinant pBluescript phagemid (with insert) was excised from the λ-ZAP vector using the ExAssist/SOLR system (Stratagene). Miniprep plasmid DNA was prepared using an alkaline lysis method (Sambrook et al. 1989) and was often sequenced directly after RNAse treatment, two phenol-chloroform extractions and ethanol preciptation. Large scale plasmid preparations were also done using an alkaline lysis method followed by plasmid precipitation with polyethylene glycol (Sambrook et al. 1989). Double stranded sequencing with ^{35}S-dATP was done using the dideoxy chain terminating method with a T7 polymerase (Pharmacia). Standard sequencing gel receipes and running conditions were used (Sambrook et al. 1989). Sequencing gels were lifted directly off plates onto Whatman 3MM paper, without fixing, and dried for 45 minutes at 80°C under vacuum.

4.2.4 Rapid Amplification of cDNA Ends (RACE)

As the isolated cDNA clones were truncated near the N-terminus, it was necessary to use alternative methods to clone this region. The 5' ends of the truncated cDNA clones were determined using the rapid amplification of cDNA ends (RACE) technique (as illustrated in Fig. 4.1) using a modification of previously described methods (Frohman 1990; Jain et al. 1992; Schuster et al. 1992). Reverse transcription of 1 μg of poly-A$^+$ mRNA (heated at 70°C, 10 min and quenched on ice) was done using 200 units superscript II reverse transcriptase (BRL), 200 μM dNTPs, 20 units RNAse inhibitor (RNasin, Promega), 10 mM DTT, and 10 pmol of a gene specific primer (gsp1) located near the 5' end of the truncated cDNA clone [5'AATGAAGCCGATGGTCT3']. The reverse transcription reaction was incubated at 42°C for 60 minutes followed by a 50°C incubation for 15 minutes. The cDNA was treated with RNAse H (Pharmacia) at 42°C for 15 minutes to remove the RNA template. The reverse

90
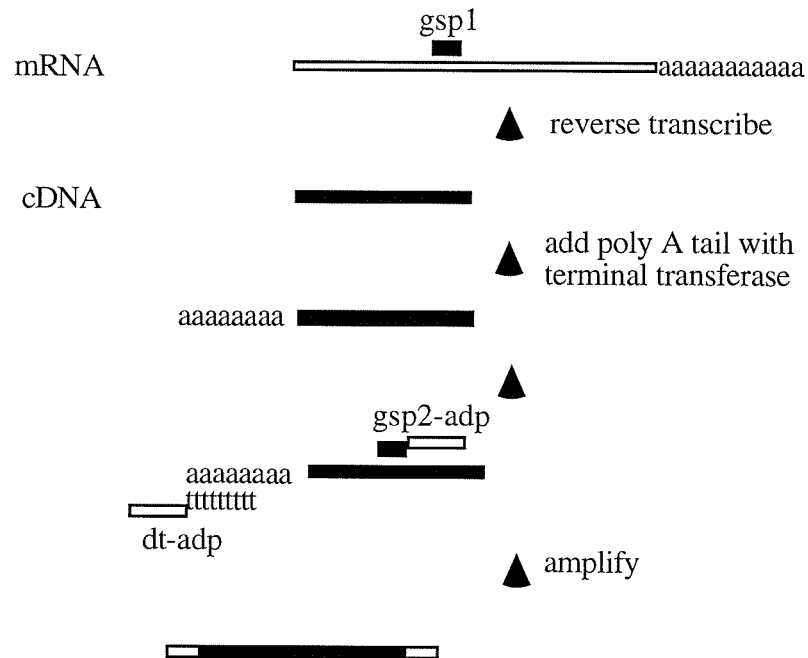
Figure 4.1 Rapid amplification of cDNA ends (RACE) technique,
illustrated. The gene specific primers are gsp1 and gsp2.
The adapter portion of the primer, adp, is indicated.
Mechanism is explained in the text.

transcription primer and dNTPs were removed by ultra-filtration through ultrafree-MC

centrifugation filters (30,000 NMWL, Millipore) (Jain et al. 1992). Following a washing step in

the ultra-filtration unit, the samples were concentrated under vacuum to approximately 10 μl. A

poly-A tail was then added to the extended cDNA by using terminal deoxy transferase

(Pharmacia) in a standard PCR buffer (10 mM Tris-Hcl pH8.3, 50 mMKCl, 25 mM MgCl$_2$) with

2 mM dATP (Schuster et al. 1992).

Amplification of the cDNA was done using a poly T primer with a 5' adapter (indicated

in bold) [5' **GACTCGAGTCGACATCGTTTTTTTTTTTTTTTTT** 3'] and a second, nested

91

gene specific primer (gsp2) closer to the 5' end of the truncated clone, also with a 5' adapter

[**GACTCGAGTCGACATCG**AGCAGGCAGAGCAGACA3']. The amplification reaction was

carried out as before (section 4.2.3) but with 2.5 mM $MgCl_2$ and 10 pmol of each primer. The

reaction was denatured at 95°C for 5 min. and brought down to 80°C when 2 units of Taq

polymerase were added and the reaction layered with oil. The reaction then continued with the

following profile: cycle 1, 55°C-5 min., 72°C-20 min.; cycle 2→37, 95°C-45 sec., 55°C-1 min.,

72°C-2 min.; cycle 38, 95°C-45 sec., 55°C-1 min., 72°C-10 min. The amplified product was

isolated, as previously described, and cloned in a pBluescript II vector (Stratagene) that was cut

with EcoR V and tailed with ddTTP (Holton and Graham 1991).

4.2.5 Southern blots

For a Southern blot, 4 μg of genomic DNA was digested with an appropriate restriction

enzyme using the conditions recommended by the manufacturer. In order to ensure complete

digestion, 10 units of enzyme were added at one hour intervals for approximately three hours,

mixing gently at each addition to avoid mechanical shearing. DNA was run on large (20 X

28 cm) 0.8% agarose gels in a Tris-Borate-EDTA electrophoresis buffer (Sambrook et al. 1989)

with 100 μg/ml ethidium bromide. Gels were transferred onto Hybond-N nylon membrane

(Amersham) by capillary action in 20X SSC, as described in Sambrook et al. (1989). Nucleic

acids were fixed onto the membrane by baking for 2 hours at 80°C. Membranes were

prehybridized in Church buffer (0.25 M sodium phosphate, 7% SDS, 1 mM EDTA) for ≥ 1 hour

at 65°C (Church and Gilbert 1984). Hybridization buffer was changed prior to adding denatured

probe and then incubated for ≥ 16 hours. Membranes were typically washed at low stringency

(1X SSC/0.1% SDS at 65°C) three times (15 minutes each) or at a higher stringency wash

(0.1X SSC/ 0.1%SDS at 65°C), as required.

4.2.6 Northern blots

RNA was quantitated spectrophotometrically, and an appropriate amount precipitated and resuspended in a RNA loading buffer. Samples were heated at 70°C for 10 minutes and quenched on ice for 5 minutes prior to loading the gel. A 1.5% agarose-formaldehyde gel made with a 1X MOPS buffer (20 mM MOPS pH7, 5 mM sodium acetate, 1 mM EDTA) and 6% (v/v) formaldehyde (2.2 M) was used. The gel was run in a 1X MOPS buffer with 2.2 M formaldehyde (Douglas et al. 1990). RNA was transferred to Hybond-N, fixed and hybridized as in section 4.2.5.

4.3 Results

4.3.1 Identification and characterization of the Fcp1 *and* Fcp2 *cDNAs*

A cDNA expression library was constructed from *Heterosigma* mRNA. Because of this, the initial several rounds of screening were attempted with both the α-FCP and α-CP1a antisera. However, this method was not successful and was abandoned. An alternative was to screen the library with a nucleic acid probe. A heterologous *Fcp* gene probe from a diatom did not hybridize to *Heterosigma* genomic DNA; therefore, I had to obtain protein sequence information directly from a *Heterosigma* FCP in order to generate primers for the amplification of a homologous nucleic acid probe.

The sequences of seven tryptic fragments were obtained from the 19.5 kDa polypeptide in fraction 1 (F1) of the sucrose gradient (Fig. 4.2, F1). This polypeptide was easily purified from the other three dominant FCPs normally present in the thylakoids (Fig. 4.2, Chapter 2). The 19.5 kDa polypeptide was immunologically related to the diatom FCP sequence (Fig. 4.2,

F1). A total of 60 amino acid positions were obtained from the sequencing of seven tryptic fragments. Only two of these tryptic fragments (T1 and T2) could be unambiguously aligned with the *Phaeodactylum* FCP sequence and both were within the first putative membrane spanning region. The sequence information from tryptic fragments T1 and T2 was used to create degenerate primers for the amplification of a *Fcp* specific probe; this was used to screen the cDNA library.

<div>

Thy   F1   Tryptic Fragments

T1-	T V E I K
T2-	X S M L A Y L G$^?$F L V T F A G$^?$
T3-	L P G A Y D L A G D Q F S S L P T
T4-	S/L P G A Y D L A G D Q F
T5-	E E L/A E/A G/A G$^?$X E A
T6-	I I N S L L G S P V E F
T7-	I I N S L L G S P V E/D F N A G F H$^?$

</div>

30 — (marker)
21.5 — (marker)
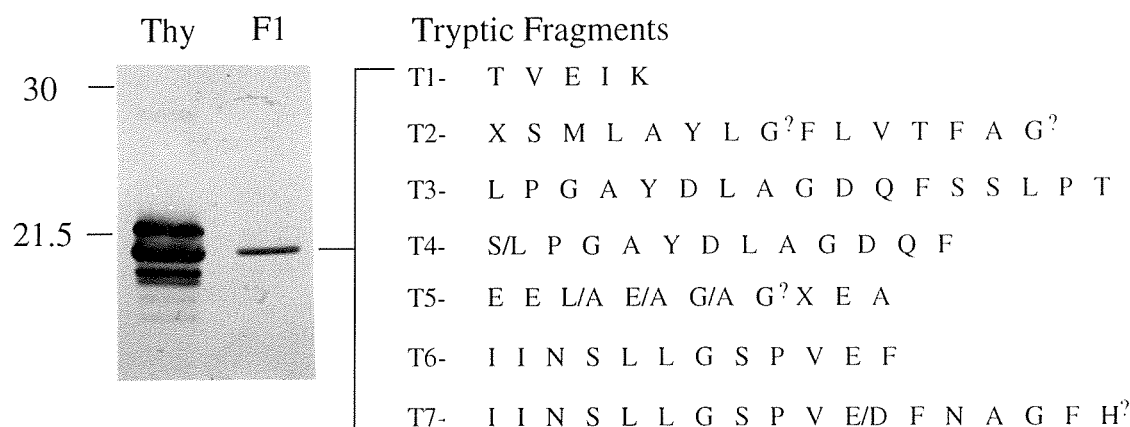
Figure 4.2

Tryptic fragment sequences determined from the abundant protein in fraction 1 (F1) of the sucrose gradient (Chapter 2). The left panel shows a western blot of a thylakoid fraction (Thy) and fraction 1 (F1) immunostained with the α-FCP antiserum. Molecular mass standards in kDa are indicated on the left. Sequences from the tryptic fragments T1-T7 are indicated on the right. Ambiguous amino acids are indicated by a question mark.

More than 90 000 recombinant cDNA clones were screened with the PCR amplified probe resulting in the detection of 3 clones. However, all were truncated upstream of the first putative membrane spanning region. A re-screening with one of the truncated cDNA clones detected an additional 30 positive clones of which 15 were isolated; most were smaller than the original. The RACE technique was used to amplify the 5' end of the *Fcp* cDNA directly from poly-A$^+$ mRNA using a reverse transcription reaction followed by PCR amplification (see Fig. 4.1, methods). A number of nearly identical fragments differing in only a few nucleotides, mainly within the third codon position, were obtained from the amplification reaction. The RACE product that was identical to the truncated cDNA clone in the 200 bp overlapping region was used to generate the full length sequence. The sequencing strategy and the proportion of the full length sequence determined from the RACE product and the cDNA clone are indicated in Figure 4.3.

The cDNA encoding the FCP (*Fcp*1) is 858 base pairs (bp) long with an open reading frame of 625 nucleotides (Fig. 4.4). This gives an immature polypeptide with 210 amino acid residues which is typical for the FCPs (Grossman et al. 1990; Apt et al. 1994). The proposed transit peptide cleavage site is at serine 36 (Fig 4.4), by analogy to the *Phaeodactylum* processing site (Bhaya and Grossman 1991). With the cleavage site at Ser 36, the mature protein would have a calculated molecular mass of 18.9 kDa. This is very close to the estimated size of 19.5 kDa based on the SDS polyacrylamide gel (Fig. 4.2; Chapter 2). A typical polyadenylation signal at the 3' end of the cDNAs (AATAAA) was not present in any of the cDNA clones.

In line with the proposed nomenclature for the *Cab* genes (Jansson et al. 1992), I use the term *Fcp*1 to refer to a specific cDNA/gene type (type 1) that includes all the nearly identical members. As I have only a single complete cDNA of *Fcp* type 1, this term refers directly to the sequence shown in Figure 4.4.

The tryptic fragment sequences were identical to the inferred protein sequence in all positions except for two from fragment T2. These residues are indicated by square brackets [] in Figure 4.4. The first conflict has a Tyr [Y] instead of the inferred Ile 82 (I). However, the latter
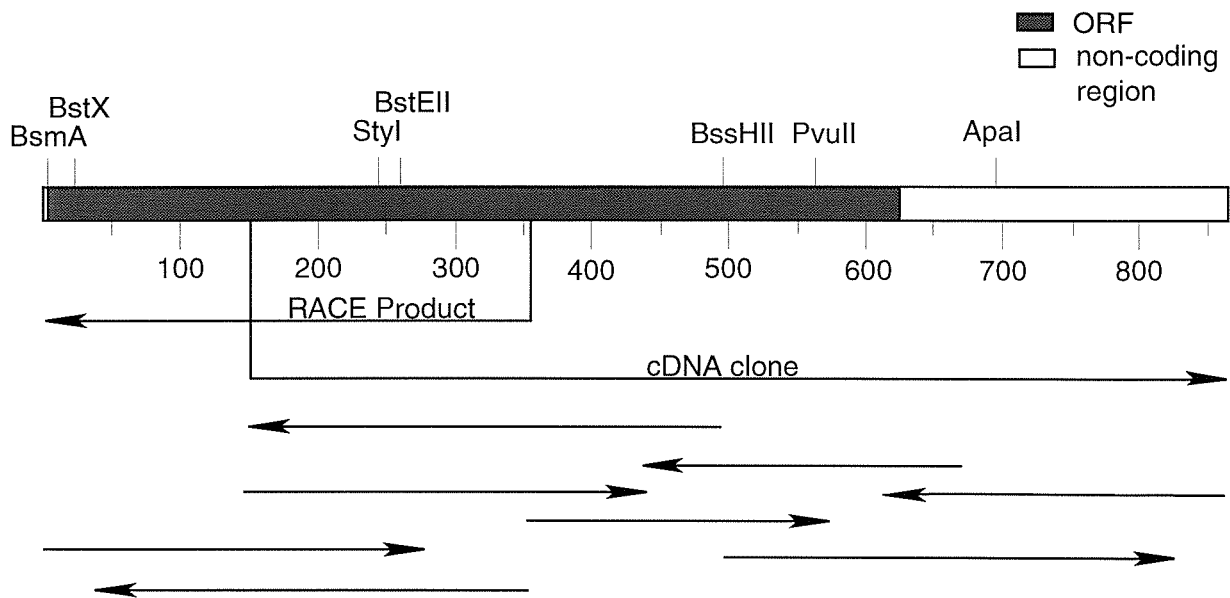
Figure 4.3

Schematic representation of the *Fcp*1 cDNA and the sequencing strategy used. Shading represents the FCP open reading frame (ORF). The sections of the *Fcp* sequence determined from the RACE product and the cDNA clone are indicated. Key restriction sites are given and the arrows indicate the sequencing done on both strands.

```
 •+1/1                                      •28/10
aaa atg tct ctc aag ctc gcc acc ctc gct gcc gcc ctc atg ggt gcc tcc gcc ttc gtg
    met ser leu lys leu ala thr leu ala ala ala leu met gly ala ser ala phe val


 •58/20                                     •88/30
gcc ccc aac aag atg ggc gtg gcc aag tcc tcc tcc gcc ctg aag atg agc ttc gag aac
ala pro asn lys met gly val ala lys ser ser ser ala leu lys met ser phe glu asn
                                                                            Δ
 •118/40                                    •148/50
gag atc ggc gtc cag gcc cct ctg ggc ttc tgg gac cct ctg ggc ctt ctg gat gag gcc
glu ile gly val gln ala pro leu gly phe trp asp pro leu gly leu leu asp glu ala


 •178/60                                    •208/70
gac cag gag cgg ttc gac cgc ctc cgt acc gtc gag atc aag cac ggc cgc atc tcc atg
asp gln glu arg phe asp arg leu arg thr val glu ile lys his gly arg ile ser met
                                     T   V   E   I   K               X   S   M
 •238/80                                    •268/90
ctt gcc atc ctt ggc cac ctg gtg acc acc gct ggt gtg cgt ctg cct ggc gct tac gac
leu ala ile leu gly his leu val thr thr ala gly val arg leu pro gly ala tyr asp
 L   A  [Y]  L   G   F?  L   V   T  [F]  A   G?          L   P   G   A   Y   D
 •298/100                                   •328/110
ctg gct ggc gat cag ttc tcc agc ctg ccc acc ggc ctg aag gct ctg tct gct ctg cct
leu ala gly asp gln phe ser ser leu pro thr gly leu lys ala leu ser ala leu pro
 L   A   G   D   Q   F   S   S   L   P   T
 •358/120                                   •388/130
gct gct ggt gtg gcc cag acc atc ggc ttc att ggt ctg att gag ctg ggt ttt gct cag
ala ala gly val ala gln thr ile gly phe ile gly leu ile glu leu gly phe ala gln


 •418/140                                   •448/150
atc aag gag gag ctg gag gct gac tgc gag gcc cgc atg gac gct gcc ggg tgg gat gac
ile lys glu glu leu glu ala asp cys glu ala arg met asp ala ala gly trp asp asp
         E   E   L?  E?  G?  G?  X   E   A
 •478/160                                   •508/170
gag aag aag gac tcc aag cgc gca att gag ctg aac aac ggc cgt gcc gcc cag atg ggt
glu lys lys asp ser lys arg ala ile glu leu asn asn gly arg ala ala gln met gly


 •538/180                                   •568/190
atc ctt gcc ttg atg gtc cac gag cag ctg gac aac aac cct tac atc atc aac tct ctg
ile leu ala leu met val his glu gln leu asp asn asn pro tyr ile ile asn ser leu
                                                         I   I   N   S   L
 •598/200                                   •628/210
ctg ggc tcc cct gtg gac ttc aac gct ggc ttc taa aca gtg ttt ttt ttt cct tct cca
leu gly ser pro val asp phe asn ala gly phe OCH
 L   G   S   P   V  E/D  F   N   A   G   F   H?


aaa att ttt gac ttt ctc ttg acc ttc tga ctt ttc tgg gcc cag gct ctt tct agt gtt
ggg ttg ctg tgc tgg gtt agt tcg cag ttt gta gcc tga tgc gtg ttt cat cat ata gac
tct ttt gtt ggt atc ttg act tga ctt ctt ttg ctt gac tgc tca caa tct ttc ata cac
ttt gaa cta tag tac cct tct ttc •858
```

Figure 4.4

Nucleotide sequence of the *Fcp*1 cDNA from *Heterosigma*. The nucleotide and amino acid positions are given above the first nucleotide of the codon (•). Tryptic fragment sequences are shown in bold below the appropriate amino acid. Ambiguous amino acids are indicated by a question mark (?). Amino acids not matching the derived gene sequence are indicated by square brackets []. A putative processing site is indicated by an open triangle (Δ).

is consistent with the other FCP sequences as there is an Ile and a Val at an homologous position in the FCPs from both *Macrocystis* and *Phaeodactylum*, respectively. The other conflict involves a Phe [F] in the sequence of T2 but a Thr 89 (T) in the inferred protein. At homologous positions in the other FCP sequences, there is either a glutamate (E), a glutamine (Q) or an arginine (R) residue. These conflicts could be the result of sequencing ambiguities. Alternatively, since the FCPs are encoded by several nearly identical multigene family members, the differences may reflect true sequence polymorphisms.

A second cDNA clone from the first round of screening was significantly divergent from the first cDNA (*Fcp1*). An alignment of this second clone (*Fcp2*) with *Fcp1* is shown in Figure 4.5. Because *Fcp2* is not full length, the alignment starts at amino acid position 73. The two clones are 75% identical at the nucleotide level and 70% identical at the amino acid level. When the chemical similarities of the amino acids are considered, *Fcp1* and *Fcp2* are approximately 82% similar. Most of the differences between the two clones occur in the connectors linking the putative membrane spanning regions (MSR). Several changes in the second MSR were also observed, though most were conservative substitutions. On the other hand, few amino acid changes are observed in the first and third MSRs. Because this sequence shows significant differences from the main *Fcp1* it may represent a different *Fcp* type; therefore, it is referred to as *Fcp2* (*Fcp* type 2). However, such a designation requires further characterization.

A wide band with an average size of 0.98 kb is detected using the *Fcp1* clone to probe total RNA (Fig. 4.6); this is consistent with the size of the cDNA clone. The wide hybridization signal indicates that there is probably a collection of related transcripts with small size variations. This is consistent with the variable sizes of the 3' non-coding region in different *Fcp* cDNA clones (not shown). The strong hybridization signal obtained in a short period of time (8 hrs) indicates the *Fcp* genes are abundantly expressed.

An analysis of the codon usage is given in table 4.1A. In the third position of the codon there is an overall preference for pyrimidines (C + T), representing over 67% of the codons

98

```
•217/73                                            •247/83
 K   H   G   R   I   S   M   L   A   I   L   G   H   L   V   T   T   A   G   V
aag cac ggc cgc atc tcc atg ctt gcc atc ctt ggc cac ctg gtg acc acc gct ggt gtg
aag cat ggc cgc atc tct atg ctt gcc atc ctt ggc cac atc ctc acc act gcc ggt gct
 K   H   G   R   I   S   M   L   A   I   L   G   H   I   L   T   T   A   G   A


•277/93                                            •307/103
 R   L   P   G   A   Y   D   L   A   G   D   Q   F   S   S   L   P   T   G   L
cgt ctg cct ggc gct tac gac ctg gct ggc gat cag ttc tcc agc ctg ccc acc ggc ctg
cgc tgg cct ggt gcc gtg gat ctg tcc ggc aag aca tac gcc gag atc cct gct ggt atc
 R   W   P   G   A   V   D   L   S   G   K   T   Y   A   E   I   P   A   G   I


•337/113                                           •367/123
 K   A   L   S   A   L   P   A   A   G   V   A   Q   T   I   G   F   I   G   L
aag gct ctg tct gct ctg cct gct gct ggt gtg gcc cag acc atc ggc ttc att ggt ctg
aag gcc ctt ggt gcc ctt cct ttt gct ggc gtc tgc cag att gtg gcc ttc att ggt ctg
 K   A   L   G   A   L   P   F   A   G   V   C   Q   I   V   A   F   I   G   L


•397/133                                           •427/143
 I   E   L   G   F   A   Q   I   K   E   E   L   E   A   D   C   E   A   R   M
att gag ctg ggt ttt gct cag atc aag gag gag ctg gag gct gac tgc gag gcc cgc atg
att gag ctt ggt ttc tcc aag tgc cag gat gac gtg gca gcg ttc tgc gag ggc aag atg
 I   E   L   G   F   S   K   C   Q   D   D   V   A   A   F   C   E   G   K   M


•457/153                                           •487/163
 D   A   A   G   W   D   D   E   K   K   D   S   K   R   A   I   E   L   N   N
gac gct gcc ggg tgg gat gac gag aag aag gac tcc aag cgc gca att gag ctg aac aac
gat gag cag ggc tgg gat gag gcc aag aag gac tcc aag cgc gcc att gag ctg aac aac
 D   E   Q   G   W   D   E   A   K   K   D   S   K   R   A   I   E   L   N   N


•517/173                                           •547/183
 G   R   A   A   Q   M   G   I   L   A   L   M   V   H   E   Q   L   D   N   N
ggc cgt gcc gcc cag atg ggt atc ctt gcc ttg atg gtc cac gag cag ctg gac aac aac
ggc cgt gct gct cag atg ggt atc ctc gct ctg atg gtg cac gag acc atc aac aac gat
 G   R   A   A   Q   M   G   I   L   A   L   M   V   H   E   T   I   N   N   D


•577/193                                           •607/203
 P   Y   I   I   N   S   L   L   G   S   P   V   D   F   N   A   G   F   Z
cct tac atc atc aac tct ctg ctg ggc tcc cct gtg gac ttc aac gct ggc ttc taa aca
cct tat gtg atc aac tct ctc ctc ggt gcc cct gtg gac ttc aac gcc ggc ttc taa ttt
 P   Y   V   I   N   S   L   L   G   A   P   V   D   F   N   A   G   F   Z


gtgttttttttttccttctccaaaaattttttgactttctct»
tcctttatcctgtaacataatttttaaatgacccatagag#
```

Figure 4.5

Comparison of the *Fcp*1 sequence (fig.4.4) with the *Fcp*2 partial sequence. *Fcp*1 amino acid and nucleotide sequences (top two lines) are displayed starting at amino acid 73, according to the labeling in figure 4.4. The partial *Fcp*2 nucleotide and amino acid sequences are in the bottom two lines of each row. Nucleotide/amino acid positions are indicated above the appropriate residues. Lines above and below selected residues indicate the potential membrane spanning regions. Amino acids and nucleotides in bold indicate areas differing between the two sequences. The # at the end of *Fcp*2 indicates the start of the poly A-tail. The » symbol at the end of *Fcp*1 indicates the presence of additional sequence that is not shown.

99

10µg 20µg

— 4.5

— 2.4

— 1.4

— 0.24
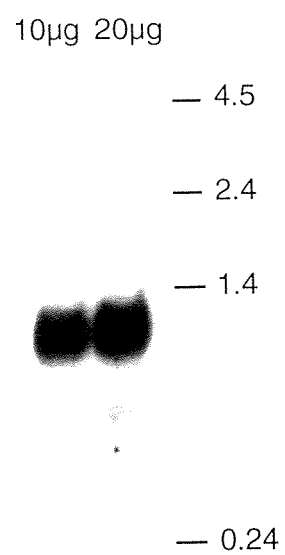
Figure 4.6

    Northern blot of *Heterosigma* total RNA probed with the *Fcp*1 cDNA probe. 10 and 20 µg of total RNA were loaded. Molecular mass markers (kb) are shown.

Table 4.1A   Codon usage of the *Heterosigma Fcp* 1 cDNA

210 codons   MM : 22314 Dalton

TTT phe F	1	TCT ser S	3	TAT tyr Y	-	TGT cys C	-
TTC phe F	8	TCC ser S	8	TAC tyr Y	2	TGC cys C	1
TTA leu L	-	TCA ser S	-	TAA OCH Z	1	TGA OPA Z	-
TTG leu L	1	TCG ser S	-	TAG AMB Z	-	TGG trp W	2
CTT leu L	4	CCT pro P	6	CAT his H	-	CGT arg R	3
CTC leu L	5	CCC pro P	2	CAC his H	3	CGC arg R	4
CTA leu L	-	CCA pro P	-	CAA gln Q	-	CGA arg R	-
CTG leu L	18	CCG pro P	-	CAG gln Q	7	CGG arg R	1
ATT ile I	3	ACT thr T	-	AAT asn N	-	AGT ser S	-
ATC ile I	9	ACC thr T	6	AAC asn N	8	AGC ser S	2
ATA ile I	-	ACA thr T	-	AAA lys K	-	AGA arg R	-
ATG met M	8	ACG thr T	-	AAG lys K	10	AGG arg R	-
GTT val V	-	GCT ala A	12	GAT asp D	3	GGT gly G	6
GTC val V	3	GCC ala A	17	GAC asp D	10	GGC gly G	13
GTA val V	-	GCA ala A	1	GAA glu E	-	GGA gly G	-
GTG val V	6	GCG ala A	-	GAG glu E	13	GGG gly G	1

Table 4.1B   Third Codon Position Data

nucleotide group	total (%)	G + C%
Pyrimidines	**67**	
C	71	
T	29	
		80
Purines	**33**	
G	97	
A	3	

101

(Table 4.1B). The pyrimidine bias is particularly evident in the alanine and glycine codon usage. Cytosine accounts for 71% of the codons in the third position when a pyrimidine occurs. Likewise, purines (G + A) occur only 33% of the time in the third codon position, though there is an extreme bias (97%) for guanine over adenine. Overall, there is a strong bias for guanine and cytosine in the third position, occurring 80% of the time. Comparatively, the G + C content of the entire *Fcp*1 cDNA is 63%, though it is only approximately 40% in the non-coding region. The codon usage in the partial *Fcp*2 clone does not appear to be significantly different from that in the *Fcp*1 clone.

4.3.2 Characterization of the Fcp gene family

Fifteen truncated cDNA clones of various lengths were isolated and sequenced. Six were derived from unique genes as determined by differences in the untranslated region at the 3' end of the cDNA (Figure 4.7). Of these six, five were nearly identical in the coding region and were therefore considered to be the same type (*Fcp*1*1-5). Preliminary sequencing evidence indicates there is at least one additional unique *Fcp* cDNA clone (*Fcp*1*6). The asterisk followed by a number indicates a different cDNA representative of the same gene type based on sequence comparison. However, as these cDNAs are not full length, the designation of these cDNAs as being nearly identical to the full length *Fcp*1 cDNA clone requires further analysis. The values to the right of Figure 4.7 indicate the number of identical copies isolated for each of the cDNA clones. Three copies of *Fcp*1 (*Fcp*1*1 in Fig. 4.7) were isolated. In addition, six clones with an untranslated region identical to *Fcp*1*3 were isolated and two equivalent cDNA clones of another type were observed (*Fcp*1*5, Fig. 4.7). The isolation of these different cDNAs indicates that the major FCP protein is encoded by a multigene family as in the terrestrial plants (Green et al. 1991) and diatoms (Grossman et al. 1990).

Hybridization of the *Fcp*2 probe to genomic DNA detected 4-7 bands, depending on the restriction enzyme used (Fig. 4.8, lane 2). When the *Fcp*1 cDNA was used as a hybridization

coding region

No.copies
detected

  E F N A G F **Z**

Fcp1*1 ...GACTTCAACGCTGGCTTC**TAA**acagtgtttttttcctctccaaaaattttgactttctcctgaccttctgactttctgggcc»   3

Fcp1*2 ...GACTTCAACGCTGGCTTC**TAA**atcaagtctctaattgtaacattagtgtgtctgtttttggcctcgtttcattttggagctggctgaca»   6

Fcp1*3 ...GACTTCAACGCCGGTTTC**TAA**atacaatctagttggtctggatcgcgatcccgcctctgggttggcgttgaggatgccatagtggact»   2

Fcp1*4 ...GAGTTCAACGCCGGCTTC**TAA**gcgatgttcctgttctcatgacttgcaaattttcaactgaatgacttcttgtaacctttcggg»   1

Fcp1*5 ...GACTTCAACGCGCTGGCTTC**TAA**gcagtgtttttctcttctcctgacccctgacctctgactttctgggcttgggcc»   1

Fcp2 ...GACTTCAACGCCGGCTTC**TAA**ttccttccttatcctgtaacataatttaaatgaccccatagag#   1

Figure 4.7

Sequence alignment of truncated cDNA clones starting 18 bp before the stop codon (bold). Coding region is shown in upper case, while the noncoding region is in lower case. The number of identical cDNA clones isolated for each is given. A (#) denotes the site of the poly A-tail. A (») indicates only a portion of the 3' end has been shown.
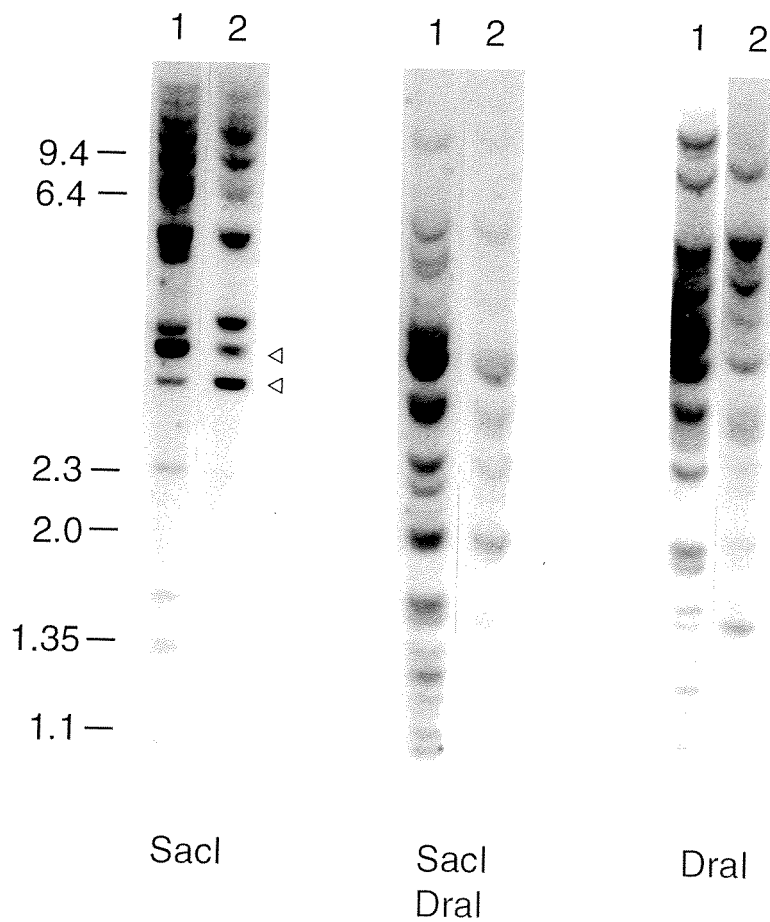
Figure 4.8

Southern blots of *Heterosigma* genomic DNA probed with the *Fcp*1 cDNA (1) or the *Fcp*2 cDNA (2). Genomic DNA was digested with Sac I, Dra I or both Sac I and Dra I, as indicated. Molecular mass markers (kb) are indicated on the left. Open arrows indicate two DNA fragments that appear to hybridize to both *Fcp* cDNA probes.

probe there were approximately 11 to 17 fragments detected (lane 1). These fragments ranged in size from 1 to ≥18 kb. However, not all the hybridization signals were of the same intensity. This could be due to multiple gene copies on the fragment and/or to the close migration of more than one DNA segment. At this washing stringency (1X SSC, 65°C) the *Fcp*1 and *Fcp*2 probes appear to hybridize to fragments of the same size such as those indicated by the open triangles in Figure 4.8. This could occur as the result of the two genes being linked on the same fragment. Alternatively, hybridization to genes of intermediate sequence divergence could account for fragments of weaker intensity and for some of the shared hybridizing bands in both lanes. In addition, bands with strong hybridization signals may have multiple copies of one gene type on that fragment. Overall, there are multiple copies of both *Fcp* gene types (*Fcp*1 and *Fcp*2), the *Fcp*1 type having more members. This indicated the existence of a very large *Fcp* multigene family in *Heterosigma*.

A series of genomic Southern blots were done in order to investigate the complexity of this multigene family more thoroughly. Figure 4.9 shows a single blot that was sequentially washed at increasingly stringent conditions, as labeled. The probe was a Sty I/BssH II fragment from *Fcp*1 (see Fig 4.3), which included sequences from the middle portion of the putative MSR1 to the middle of MSR3. At the low stringency wash (1X SSC, 65°C) there are nine prominently hybridizing bands in the 3.6-18 kb size range and a few smaller hybridizing bands (1.4-3.3 kb) in the Sac I digest. The Dra I (D) digest shows over 20 hybridizing bands, as does the Sac I/Dra I double digest. At the next level of stringency (0.1X SSC, 0.18 M Na$^+$) there was little obvious change in the number of hybridizing bands or in the signal ratio between them in the Sac I digest. In the Dra I digest one hybridizing band was lost (4.1 kb) and the signal intensity of five other bands decreased relative to the others (12, 8, 5.5, 5.4, and 3.6 kb). In the double digest, three hybridizing bands were removed (8, 5.8 and 2.2 kb) and a fourth had a reduced signal (3.8 kb). The hybridization conditions described above were both below the theoretical T_m of the probe, which was 93°C at 0.18 M Na$^+$ (1X SSC) and 77°C at 0.018 M Na$^+$ (0.1X SSC).
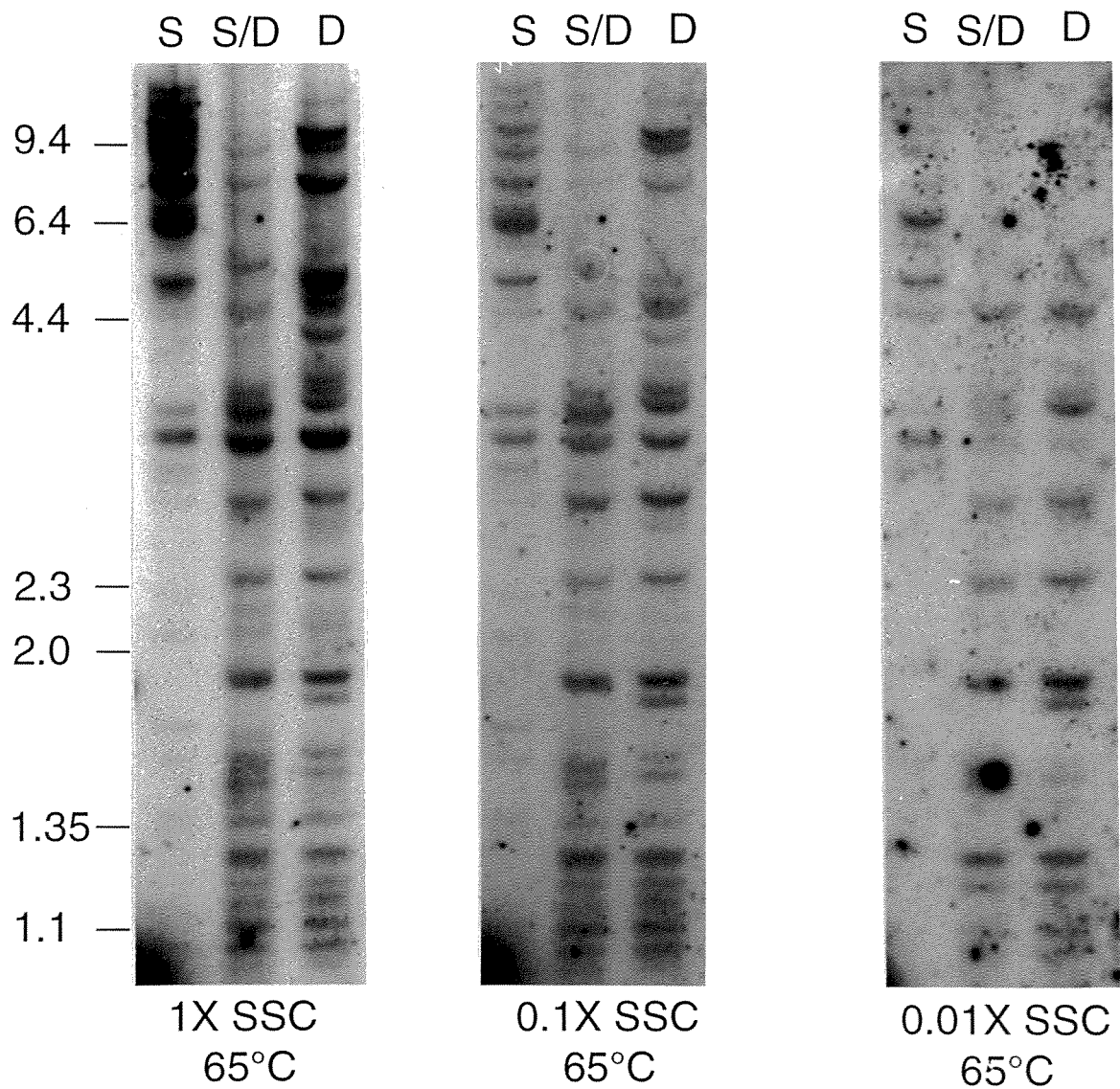
Figure 4.9

Southern blots of *Heterosigma* genomic DNA probed with a ^{32}P-labeled Sty I/BssH II *Fcp*1 cDNA fragment. The blot was successively washed under increasingly stringent conditions as indicated on the bottom of each panel. Molecular size markers (kb) are indicated on the left of the first panel. The restriction enzymes used were as follows: S, Sac I; S/D, Sac I/Dra I; D, Dra I.

The final washing condition (0.01X SSC, 65°C) exceeded the theoretical T_m of the probe (61°C at 0.0018 M Na$^+$). After this wash, the probe was selectively removed from a number of fragments. Three fragments with a strong signal (6.6, 5.5, and 3.6 kb) remained in the Sac I digest after the 0.01X SSC wash, along with one weakly hybridizing fragment (4.5 kb). In the Dra I lane, 8 prominent hybridizing bands remained with a size of 4.5, 3.8, 3.0, 2.5, 1.9, 1.8, 1.3, and 1.2 kb. Six main bands were detectable in the Sac I/Dra I digest (4.5, 3, 2.5, 1.9, 1.3, and 1.2 kb). Though it is difficult to get an accurate estimate of gene copy number using genomic Southerns, there appear to be 6-8 copies of the *Fcp*1 gene. This is in good agreement with the number of characterized cDNA clones.

There seem to be as many as 20 or more related *Fcp* genes present on the nuclear genome of *Heterosigma*, as seen in the number of hybridizing fragments at the lower stringency washes. These may represent different *Fcp* gene types similar to the gene types in the *Cab* family. However, the presence of additional hybridizing bands due to the existence of pseudogenes cannot be ruled out. Though care was taken to achieve complete digestion and the pattern was repeatable, some of the weakly hybridizing bands may represent intermediates of an incomplete digestion. Neither Sac I or Dra I cut within the full length *Fcp*1 cDNA clone, however, internal restriction sites in other genes could cause an over-representation of detectable bands.

4.3.3   Characterization of the FCP protein sequence

The FCP protein is predicted to have three membrane spanning regions using the Kyte and Doolittle scale of amino acid hydrophobicity (Kyte and Doolittle 1982) with a window size of 19 (K-D plot, Fig. 4.10A). The hydrophobic regions of the mature protein detected in the hydropathy plot include residues 73-96, 106-141, 173-192, and the C-terminus (200-210) (as labeled in Fig. 4.4, 4.10, 4.12). A fourth hydrophobic region is also detected in the amino terminus of the protein (residues 1-28), which is entirely within the predicted transit sequence

107

(TS) domain. The membrane spanning regions (MSR), as labeled in Figure 4.10A, were predicted by comparison to the K-D plot and by comparing the FCP sequence with the CABs (Fig. 4.13). The regions that span the thylakoid membrane in the pea LHC II structure (Kühlbrandt et al. 1994) are conserved in the FCPs; therefore, a similar topology was thought to be a reasonable assumption (see Fig. 4.15 for model). A feature of the prediction is the presence of distinct hydrophilic domains at the start of the putative membrane spanning regions 1 and 3 (Fig. 4.10, 4.15). These areas may protrude from the bilayer in a continuation of the α-helical structure, as appears to be the case with the pea LHC II (Kühlbrandt and Wang 1991; Kühlbrandt et al. 1994). Figure 4.10B shows the distribution of acidic and basic amino acids within the protein by vertical bars. Acidic amino acids (top panel (A) of Fig. 4.10B) which include glutamic acid (full bar) and aspartic acid (intermediate bar). The positions of the basic amino acids (bottom panel (B) of Fig. 4.10B) are also shown and include arginine (full bar), lysine (intermediate bar), and histidine (short bar). There is an enrichment for acidic and basic amino acids in the regions preceding, and at the start of, the predicted MSRs which are located on the stromal side of the membrane. Stroma exposed areas are indicated by shading in Figure 4.10. The areas exposed to the thylakoid lumen are indicated by the hatched regions (Fig. 4.10). Few acidic and basic residues are present in the predicted lumen exposed sections of the FCP (Fig. 4.10B). There are a few occurrences of acidic and basic residues within the membrane exposed portion of the protein, some of which are probably involved in the binding of chlorophyll (discussed later).

The first and third putative membrane spanning regions of the FCP are related as is the case for the CABs; first recognized in a tomato *Cab* gene by Hoffman et al. (1987). This internal similarity between the two regions is suggestive of a gene duplication. These regions (MSR1 and MSR3) are approximately 49% similar to one another (Fig. 4.11), comparable to the degree of relatedness between MSR1 and MSR3 of the CAB proteins. With the CABs, similarity between MSR1 and MSR3 extends into the domains immediately preceding the start of the putative transmembrane α-helix. This is not the case with the FCPs.
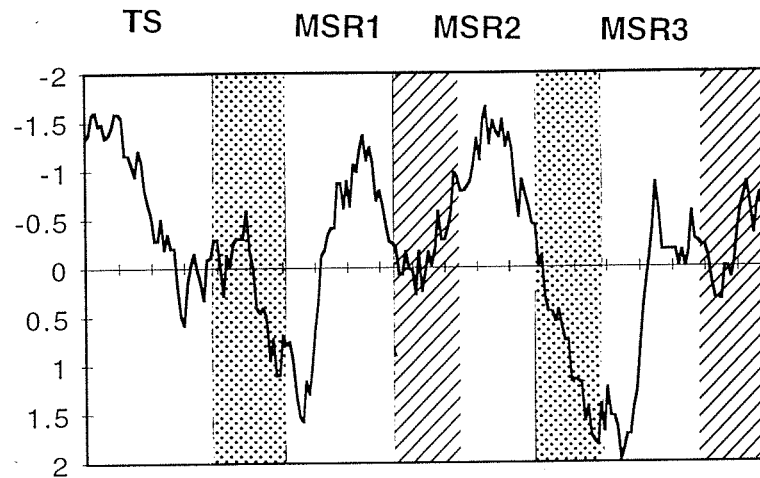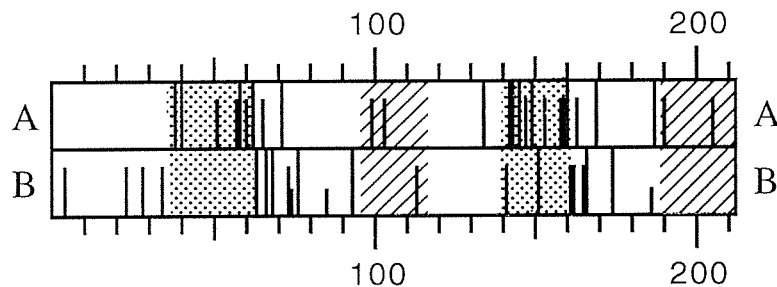
Figure 4.10

Topological analysis of the *Heterosigma Fcp*1 full length sequence. A) Kyte-Doolittle hydropathy plot done using a sliding window of 19 amino acids. Hydrophobic areas are assigned a positive value and hydrophilic ones are negative The transit sequence (TS) and the putative membrane spanning regions (MSR1-3) are labeled and roughly correspond to the clear areas. B) an acidic-basic map of the FCP protein. The position of the acidic amino acids, are indicated in the top panel (A) by a full vertical bar (E-glutamic acid) and an intermediate vertical bar (D-aspartic acid). The bottom panel (B) shows the position of the basic amino acids with a full bar (R-arginine), an intermediate bar (K-lysine), and a short bar (H-histidine). In all three panels, shading represents regions of the protein exposed to the stroma of the chloroplast. Hatches indicate areas that are exposed to the thylakoid lumen.

```
MSR1    61  DQERFDRLRTVEIKKHGRISMLAILGHLVTTAGVRLP    95
MSR3   159  DDEKKDSKRAIEL NNGRAAQMGILALMVHEQLDNNP    193
```

Figure 4.11

    Alignment of the first and third putative membrane spanning regions of the *Heterosigma* FCP. Similar amino acids are in bold. Numbers indicate the positions of the first and last amino acids shown, relative to the entire protein.

    The alignments in figures 4.12 and 4.13 were done with MACAW (Schuler et al. 1990) and adjusted by hand to maximize similarity. The Chl *a* + *c*-binding proteins (FCPs/*i*PCPs) are very well conserved between the different algal taxa. The *Heterosigma* FCP sequence is roughly 77% similar to the other known algal FCP sequences. The boxed regions in Figure 4.12 represent similarities between the FCPs and the *i*PCPs. Shaded regions are amino acids conserved only in the FCPs, excluding the *Isochrysis* sequence. The greatest similarity between the Chl *a* + *c*-binding proteins is within the putative membrane spanning regions; particularly MSR1 and MSR2 (Fig. 4.12). The N-terminal portion preceding MSR1 is also very highly conserved, though the analogous region in front of MSR3 is not. This is contrary to the relationships between the stromal exposed regions in front of MSR1 and MSR3 in the CABs (see Fig. 4.13).

    Amino acid similarities between the CABs, FCPs and *i*PCPs are indicated by the boxed regions in Figure 4.13. The *Heterosigma* FCP sequence is approximately 40% similar to both the tomato LHC I and LHC II sequences, when the unambiguously aligned positions are compared (see Fig. 5.1, chapter 5). For comparison, the LHC I and LHC II sequences from tomato are approximately 57% similar using the same amino acid positions in the calculation. The *Heterosigma* FCP sequence shows the greatest similarity to the green algal and terrestrial plant CAB sequences, almost exclusively within the predicted membrane spanning regions of the mature protein, primarily within the first and third MSRs. However, the sequence conservation between the CABs and the FCPs does extend into the stromal side of the first MSR

```
                1           10          20          30          40          50          60
                |           |           |           |           |           |           |
iPcp-Ac                                                     FESELGVAA PTGFWDPLGL AKDGSMKABK
Fcp-Hc    MSLKLAT FAAALM--GA SAFVAPNKMG VAKSSSALKM SFENEIGVQA PLGFWDPLGL LDEADQERHD
Fcp1-Mp   MKSAVMAVA CAAAPGFRGP SAFNGAALTT SAKACSAMKM SFESEIGAQA PLGFWDPLGL LADADQERFE
Fcp-Ls                                                                  DPLGL LADADOE-FF
Fcp-Os      MKL--- AIAALLAGSA AAF-APAQ-- SGKASTALNM AFESELGACP PLGFFDPLGM LADADQERFD
Fcp1-Pt     MKF--- AVFAFLLASA AAF-APAQQ- SARTSVATNM AFENEIGAQQ PLGYWDPLGL VADGDOEKHD
Fcp-Ig     MMTL--- ASLPSTAIAG LASAAPKVQ- -PRMAANDEF AM-GLPGGAN ILGEFDPAGF LKGKDKLEVY
Fcp-Pl                                                          A PLGFYDPLGL LDNADEEYFE
```

MSR 1

```
                70          80          90          100         110
                |           |           |           |           |
iPcp-Ac   RRRASEIKHG RIAMLATMGY ITPELTGKFP G--YLSPSMG LKYEDIPNGL GAI
Fcp-Hc    RLRIVEIKHG RISMLATLGH LVTTAGVRLP G--AYDL-AG DQFSSLPTGL KAL
Fcp1-Mp   RLRYVEVKHG RIAMLATAGH LTQQN-TRLP G--MLSNSAN LSFADMPNGV AAL
Fcp-Ls    RLRYVEVKHG RIAMLATAGH LTQQN-TRLP G--MLSNSA
Fcp-Os    RLRYVEVKHG RIAHVAFLGQ IVTRNGIHLS GNIDYAGNSF DSFPNGWAAI SGP
Fcp1-Pt   RLRYVEVKHG RICMLAVAGY LTQEAGIRLP G--DIDYS-G TSFESIPNGF AAL
Fcp-Ig    RLRBAETTHG RVAMLASLGF MVQEK-FH-- --PLFSGDNG PAIEQIP--- ---
Fcp-Pl    RM YVEL                                                 L
```

MSR 2

```
                120         130         140         150
                |           |           |           |
iPcp-Ac   SKVPAVG WAQIIAYAFY CEQSQDQSEG SAGEAGDFGF KVLT------ ---------- -
Fcp-Hc    SALPAAG VAQTIGFIGL IELGFAQIKE ELEAD--CBA RM-------- DAAGWD---- -
Fcp1-Mp   SKIPPAG LAQIFAFIGF LELAVMKNVE GS---FPGDF TLG----GNP FGASWD---- -
Fcp-Ls                                                          WD---- -
Fcp-Os    DAIPQAG LLQIVAFVGI LELAVMKDVT G-EGEFPGDF R------NGA LDFGWD---- -
Fcp1-Pt   SAVPGAG IAQIIAFIGF LEIAVMKDIT G--GEFVGDF R------NNY LDFGWD---- -
Fcp-Ig    QLPYWLW IVMIIG-IGR AELFRIQKGW AKVNPETGKA DSALREGYEP GDLGFDPLGL A
Fcp-Pl    EVVPAAG LAQIVAFVGF LELPV-KDVT G-EGDFPGDF
```

MSR 3

```
                160         170         180         190         200        200
                |           |           |           |           |          |
iPcp-Ac   SKDEEGLER KLNSEIANGR LAMMAIIGMF FQDGIT-GSA YGDWANFTAS PLR
Fcp-Hc    ---DEKKDS KRATELNNGR AAQMGILALM VHEQID-NNP YIINSLLGSP VDFNAGF
Fcp1-Mp   AMSEETQAS KRATELNNGR AAQMGILALM VHEELN-NKP YVINDLVGAS YTFN
Fcp-Ls    AMSERTEAS K AIELNNGR AAQM
Fcp-Os    TFDEETKLS KRATELNNGR AAMMGILGLM VHEQLG-GSI PIVGEM
Fcp1-Pt   TFSEDKKLQ KRATELNQGR AAQMGILALM VHEQLG-VSI LP
Fcp-Ig    PSDPDEFRL MQEKELSHGR LAMIAAAGFL AQEAVS-GDT WGTYWGDATF
Fcp-Pl              KRNIELNQG AAM--ILALM VI
```

Figure 4.12

Amino acid alignment of the Chl $a+c$-binding proteins from chromophytic algae. Taxa include: Ac, *Amphidinium carterae*; Hc, *Heterosigma carterae*; Mp, *Macrocystis pyrifera*; Ls, *Laminaria* saccharina; Os, *Odontella sinensis*; Pt, *Phaeodactylum tricornutum*; Ig, *Isochrysis galbana* and Pl, *Pavlova lutherii*. Hatched boxes indicate putative membrane spanning regions (MSR 1-3). Boxed regions indicate sequence similarities between the FCP and *i*PCP sequences. Shaded regions indicate FCP similar regions, except Ig and Pl. Amino acids are numbered with reference to the *Heterosigma Fcp*1 sequence as in fig. 4.4.

111

Figure 4.13

Amino acid alignment of select Chl *a+b* and Chl *a+c* binding proteins from terrestrial plants, green algae, and chromophytes. Taxa include: Le, *Lycopersicon esculentum*; Cr, *Chlamydomonas reinhardtii*, Eg1, *Euglena gracilis*; Ms, *Mantoniella squamata*; Ac, *Amphidinium carterae*; Hc. *Heterosigma carterae*; Mp, *Macrocystis pyrifera* and Pt, *Phaeodactylum tricornutum*. Hatched boxes indicate putative membrane spanning regions (MSR 1-3). Boxes indicate similarities between Fcps, *i*Pcps and Cabs. Dashes (-) represent gaps introduced to increase similarity. Proposed Chl *a* (▲) and *b* (△) binding residues of the pea LHCII are indicated. Amino acids are numbered with reference to the *Lhcb*1-Le sequence, consistent with the numbering used in figure 4.14.

```
                                    10         20         30         40         50
                                    |          |          |          |          |
Lhca1-Le PSFLSSTKSK FAAAMPVSVG ATNSMSRFSM SADWMPGQPR ---PSYLDGS APGDFGFDPL GLGEVPA--N
Lhcb1-Le GNGRITMRKA VAKSA----- PSSSPWYGPD RVKYLGPFSG ES-PSYLTGE FPGDYGWDTA GLSADPE--T
Lhcb-Cr  LQVTCKATGK KTAAKAAAPK SSGVEFYGPN RAKWLGPYSE NATPAYLTGE FPGDYGWDTA GLSADPE--T
Lhcb-Eg1 MLATSGRKA- ----KAAPKS DNLSQWYGPD RAKWLGPLTG EV-PSYLTGE LPGDYGWDTA GLGSDPT--T
Lhcb-Ms          M ACIASSFVGS VAALKATKVQ AKSVSTVVKA DIYPEFGT-- YPG--GGESP IIPF-----G
iPcp-Ac                                                FESELGV AAPTGFWDPL GLAKDGSMKA
Fcp-Hc   MSLKL--AT FA-AALMGAS AFVAPNKMGV AKSSSALKMS ---FENEIGV QAPLGFWDPL GLLDEADQER
Fcp1-Mp  MKSAVMAVAC AAAPGFRGPS AFNGAALTTS AKACSAMKMS ---FESEIGA QAPLGFWDPL GLLADADQER
Fcp1-Pt        MKF AVFAFLLASA AAFAPAQ-QS ARTSVATNMA ---FENEIGA QQPLGYWDPL GLVADGDQEK
```

MSR 1

```
                   60         70         80         90        100        110
                    |          |          |          |          |          |
Lhca1-Le LERYKESELI HCRWAMLAVP GIIVPEALGL GN--WVKAQE WAAIPGGQAT YLGQPVPW-- ----------
Lhcb1-Le FAKNRELEVI HCRWAMLGAL GCVFPELLAR NGV-KFGEAV WFKAG----- ------SQIF SEGGLDYLGN
Lhcb-Cr  FKRYRELELI HARWAMLGAL GCQTPELLAK SGT-KFGEAV WFKAG----- ------AQIF SEGGLDYLGN
Lhcb-Eg1 LARYREAEVI HARWAMLGAL GVVTPELLAG NGVPFGEGAV WYKAG----- ------AQIF SADGLNYLGN
Lhcb-Ms  SEKNAEREVI HGRWAMLGVT GAWITENGTG IP-------- WFTAGTLCTP DDCTAVADKF P-GAVAPLAP
iPcp-Ac  FKRRRASEIK HGRIAMLATM GYITPELTGK FPG------- ---------- ------YLSP SMGLKYEDIP
Fcp-Hc   FDRLRTVEIK HGRISMLAIL GHLVTTAGVR LPG------- ---------- ------AYDL -AGDQFSSLP
Fcp1-Mp  FERLRYVEVK HGRIAMLAIA GHLTQQN-TR LPG------- ---------- ------MLSN SANLSFADMP
Fcp1-Pt  FDRLRYVEIK HGRICMLAVA GYLTQEAGIR LPG------- ---------- ------DIDY S-GTSFESIP
              ▲          ▲          ▲
```

MSR 2

```
                  120        130        140        150        160
                    |          |          |          |          |
Lhca1-Le ------GTLP TILAIEFLAI -AFVEHQR-S MEKDSEKKK- -------YPG GA--FDPLGY
Lhcb1-Le PSLVHAQSIL AIWACQVVLM -GAVEGYRIA GGPLGEVVDP L------YPG GS--FDPLGL
Lhcb-Cr  PSLVHAQNIV ATSAVQVILM -GLIEGYRVN GGPAGEGLDP L------YPG ES--FDPLGL
Lhcb-Eg1 PSLIHAQSVV LTFLSTLAIM -GAVEAYRYG GGVGDFGREL DTL----YPG GP--FDPLGL
Lhcb-Ms  EGSGY-PSFW AVLAIEVVLV -GLAEAYRTG LSDSPFEELT VGDV---SPG GR--FDSLGL
iPcp-Ac  NGLGAISKVP AVGWAQILAY AFYCEQSQDQ SEGSAGEAGD FGFKVLT--- ---------
Fcp-Hc   TGLKALSALP AAGVAQTIGF IGLIELGFAQ IKEELEAD-- CEARM----- DAAGWD----
Fcp1-Mp  NGVAALSKIP PAGLAQIFAF IGFLELAVMK NVEGS---FP GDFTLG-GNP FGASWD----
Fcp1-Pt  NGFAALSAVP GAGIAQITAF IGFLEIAVMK DITG--GEFV GDFR---NNY LDFGWD----
                   △          △
```

MSR 3

```
                  170        180        190        200        210        220        230
                    |          |          |          |          |          |          |
Lhca1-Le SKDPAKFEEL KVKEIKNGRL ALLAFVGFCV QQSAYPGTGP LENLATHLAD PWHNNIGDVI IPKGIFPN
Lhcb1-Le AEDPEAFAEL KVKEIKNGRL AMFSMFGFFV QAIVT-GKGP LENLADHLAD PVNNNAWAFA TNFVPGK
Lhcb-Cr  ADDPDTFAEL KVKEIKNGRL ASFSMFGFFV QAIVT-GKGP VQNLDDHLAN PTVNNAFAFA TKFTPSA
Lhcb-Eg1 ANDPDALAEL KVKELKNGRL AMVAMLGFYV QPLVT-KAGP VENLTFHLAD PSANNIFFTS GFA
Lhcb-Ms  AESGD-LEEL KIKELKHCRL SMFAWLGCIF QALAT-QEGP IANWQAHVAD PVHANVLTNA ASGFGFY
iPcp-Ac  SKDEEGLERK LNSEIANGRL AMMAIIGMFF QDGLT-GSAY GDWANFTASP LR
Fcp-Hc   ---DEKKDSK RAIELNNGRA AQMGILALMV HEQLD-NNPY IINSLLGSPV DFNAGF
Fcp1-Mp  AMSEETQASK RAIELNNGRA AQMGILALMV HEELN-NKPY VINDLVGASY TFN
Fcp1-Pt  TFSEDKKLQK RAIELNOGRA AQMGILALMV HEQLG-VSIL P            △
             ▲ ▲          ▲
```

(amino acid residues 44-51, Fig. 4.13). This primarily includes the WDPLGL motif of the FCPs to the (F/Y)DPLGL motif of the LHC I CABs or the (W/F)DTAGL motif of the LHC II CABs. In this case, the PL of the (F/Y)DPLGL motif is a signature sequence of the LHC I CABs which, interestingly, also occurs in the FCPs. In the CAB sequences, there is an additional FDPLGL motif in front of the third predicted MSR (residues 161-166; Fig.4.13), which is highly conserved and is part of a local 2-fold symmetry between the MSR1 and MSR3 (Kühlbrandt et al. 1994). In the FCP/iPCP sequences, there are no obviously conserved areas in front of the third putative MSR. However, the *Isochrysis* FCP sequence is an exception to this statement as it does have a FDPLGL motif in front of the third MSR, more closely resembling a CAB protein (Fig. 4.12).

Figures 4.14 and 4.15 are designed to demonstrate the location of the conserved amino acids with respect to the proposed structure of the LHC. A model of the pea sequence is shown in Figure 4.14 along with the identified chlorophyll molecules (represented by porphyrin rings) and the approximate location of the thylakoid membrane (hatched regions), though the latter has not been accurately determined. This diagram is labeled to show the amino acids conserved amongst all LHC types (CABs, FCPs, and *i*PCPs). The residues conserved between all known CAB, FCP and *i*PCP sequences are indicated with black circles; most of these are within MSR1 and MSR3. Significantly, many of the conserved residues are within the area of close contact between the two transmembrane α-helicies (MSR1 and MSR3). The area of close contact includes residues Ser 69 to Ala 76 and Gly 184 to Met 191 (Fig. 4.14), as defined by Kühlbrandt et al. (1994). Other conserved residues are thought to function as Chl *a* ligands in pea, which include the following residues: Glu 65, His 68, Gly 78 (backbone carbonyl), Glu 139, Glu 180, Asn 183 and Gln 197 (His in FCPs) (Fig. 4.13 and 4.14). All ligands thought to bind Chl *a* in pea (six in all) are conserved in the Chl *a* + *c*-binding proteins; presumably they would also be involved in Chl *a* binding.

A putative model of the *Heterosigma* FCP sequence is illustrated in Figure 4.15. The topology was predicted by aligning the FCP sequence with the CAB LHC II sequences and then

114

modeled by analogy to the pea LHC II 3-D structure. It also shows the location of the conserved (chemically similar) amino acids when the FCPs and the *i*PCPs are compared. The FCP/*i*PCP sequences are mainly conserved within MSR1 and MSR3. The similarity within the second putative MSR is significant but lower than when the first and third MSRs are compared. There is a lack of conserved residues in the lumen exposed portions of the protein and in the connector between MSR2 and MSR3. Residues thought to bind Chl *a*, by analogy to the pea structure, are well conserved (solid triangles, Fig. 4.13). The degree of sequence conservation in the membrane spanning regions is significant enough to suggest that the FCPs may have a similar structural topology, though the aqueous exposed areas are not conserved and may be structurally distinct.

Figure 4.14

Amino acid comparisons based on the structural models for the pea LHC II complex. Porphyrin rings represent the approximate location of chlorophyll molecules determined from the pea LHC II structure. The approximate location of the thylakoid membrane in each Figure is indicated by the hatched regions. The pea LHC II sequence is compared to both the CABs and the FCPs. The key to the conserved (similar) amino acids is given in the Figure. The Figure is a modification of Figure 4 from Kühlbrandt et al. (1994).

Figure 4.15

Model proposed for the *Heterosigma* FCP protein. Porphyrin rings represent the approximate location of putative chlorophyll molecules by analogy to the pea LHC II structure. The location of the chlorophylls in the FCP model are purely speculative and are included only to illustrate the conservation of the residues binding these particular molecules in the pea structure. The approximate location of the thylakoid membrane is indicated by the hatched regions. The *Heterosigma* FCP sequence is compared to the FCPs/iPCPs. The key to the conserved residues is indicated in the Figure.

117

4.3.4 Analysis of the FCP transit sequence

The N-terminus of the *Heterosigma* FCP was blocked; therefore, the precise cleavage site of the transit peptide is not known. However, comparison to the *Phaeodactylum* FCP cleavage site (Bhaya and Grossman 1991) would suggest that it occurs after Met 35 (Fig. 4.16), which is within a well conserved region. A comparison of the *Heterosigma* N-terminal region to that of other chromophyte transit sequences is shown in Figure 4.16.

Most of the chromophyte transit sequences have a basic amino acid within the first four residues (Fig. 4.16). This is followed immediately by a hydrophobic region (15-17 amino acids long) emphasized by the black section of Figure 4.16. The CAB transit peptides lack both a strongly hydrophobic region and a positively charged amino acid at the N-terminus. A proline residue followed by a positively charged amino acid usually occurs at or near the end of the hydrophobic section (Fig 4.16). This proline may function as a helix breaker should the hydrophobic region form an α-helix. After the hydrophobic region there is an increase in the number of hydrophilic residues, such as basic (Arg, Lys) and hydroxylated (Ser, Thr) amino acids. Few, if any, acidic residues (Glu, Asp) are present within the targeting sequences. A basic N-terminus, a hydrophobic region followed by a more polar section are all characteristics of a signal sequence (von Heijne 1990). However, these traits are not present in the *Isochrysis* FCP sequence (LaRoche et al. 1994). Though there is a hydrophobic region (Fig 4.16), a basic amino acid in the first four residues is not present. This sequence also has a number of hydroxylated amino acids at the beginning of the transit peptide (Fig 4.16), unlike the other chromophyte sequences.

Putative processing sites for the signal sequence-like regions are indicated in Figure 4.16 (open triangles) and are based on the (-3,-1)-rules for the prediction of signal sequence cleavage sites (von Heijne 1986). There are other processing sites that can be predicted using the (-3,-1)-rules though only one is shown. The section immediately following the hypothetical processing

118

site in the FCPs, and prior to the start of the mature polypeptide, is enriched in hydroxylated and basic amino acids (Fig 4.16). This amino acid composition is similar to the transit peptide of chloroplast localized proteins of terrestrial plants. Presumably, the remaining region would then target the polypeptide to the chloroplast envelope once it crossed the ER-like membrane. Such a bipartite transit sequence in the chromophytes has been suggested by Bhaya and Grossman (1991) and expanded on by Pancic and Strotmann (1993). In addition, the *Chroomonas* phycoerythrin transit sequence has two hydrophobic domains within it (Fig. 4.16). The first region is similar to a signal sequence and the second hydrophobic region resembles the transit sequence of a thylakoid lumen localized protein (Hiller et al. 1990).

Hc	MSLKLATFAAALMGASAFVAPNKMG	VAKSSSALKM	SF
Mp	MKSAVMAIACAAAPGLRGAEPFNGAALA	TSAKSSSAMKM	SF
Pt	MKTAVIASLIAGAAAFAPAKN	AARTSVATNM	AF
Os	MKLAIAALLAGSAAAFAPAQ	SGKASTALNM	AF
Os-a	MKFFCVAGLLASAAAFQAQPAAFTTYSPAVGGATSNVFSESSSPAHRNRRATIVM	DG
Ig	MMTLASLPSTAIAGLASAAPK	VQPRMAANDE	FA
C-pe	MFAKTLASLAVIGSAAAYVPMM	SMDMGRREVVQAGAAAAAVTPFISGAPAGA	AM

Figure 4.16

Analysis of the N-terminal transit peptide from some chromophyte nuclear encoded sequences. The sequences are from the FCPs of; Hc, *Heterosigma carterae*; Mp, *Macrocystis pyrifera*; Pt, *Phaeodactylum tricornutum*; Os, *Odontella sinensis*; and Ig, *Isochrysis galbana*. Other sequences are the the *Odontella sinensis*- γ-atpase (Os-a) and the *Chroomonas* (cryptophyte) phycoerythrin α-subunit (C-pe). Areas enclosed in black indicate hydrophobic regions. Shaded regions emphasize hydroxylated amino acids, serine and threonine. Basic amino acids (Arg, R; Lys, K) are underlined. Transit peptide cleavage sites are indicated by solid triangles (▼). Open triangles (▽) represent possible signal peptide cleavage sites based on the -1,-3 rule.

4.4 Discussion

4.4.1   Fcp cDNA structure and multigene families

The cloned *Fcp* cDNA (*Fcp*1) encodes one of the predominant polypeptides present in the thylakoids of *Heterosigma*, as confirmed by protein sequencing. The *Fcp* genes are nuclear encoded by a multigene family and the northern blot indicates they are highly expressed. Southern blot analyses of genomic DNA and cDNA sequencing indicate that there are at least 6-8 nearly identical copies of the *Fcp*1 gene in *Heterosigma*. In addition, there appear to be over 20 closely related *Fcp* gene sequences. This is comparable to the complexity observed in the tomato *Cab* gene family (table 1.2) where there are eight LHC II type I genes and greater than 13 genes encoding LHC II proteins in total (Green et al. 1991).

Similarly large LHC gene families have been reported in many other terrestrial plant taxa, such as petunia, where there are an estimated 16 Lhcb1 genes (Dunsmuir et al. 1983). Multi-Lhcb1 gene families are also present in moss (Long et al. 1989), green algae (Imbault et al. 1988; LaRoche et al. 1990) and in *Euglena gracilis* (Muchhal and Schwartzbach 1992). In the diatom *Phaeodactylum* there are at least six related members of the *Fcp* gene family (Bhaya and Grossman 1993). A similar number of *Fcp* cDNAs have been cloned and sequenced from the brown alga, *Macrocystis* (Apt et al. 1994). Moreover, the presence of multiple copies of the main LHC appears to be a general occurrence.

The gene copy number of the primary LHC II (Lhcb1) is quite variable between different terrestrial plant taxa. One mechanism for the generation of multiple gene copies is through unequal cross-over events which usually result in the creation of tandem repeats (Langridge 1991). When this occurs, subsequent unequal cross-over events can lead to increases or decreases in the size of the gene family. A high degree of similarity between the duplicated members of the *Fcp* or *Cab* gene families may be the result of either a recent gene duplication

event, if there is similarity in the surrounding non-coding sequences, or through concerted evolution (Tanksley and Pichersky 1988; Bhaya and Grossman 1993).

Large numbers of duplicated genes are usually assumed to be required for the generation of sufficient mRNA transcripts for the production of abundant proteins (Li 1983). As the FCPs are the most abundant proteins in the thylakoid membrane, this explanation seems reasonable. Alternatively, the presence of multiple gene copies may be less related to abundant protein production but more to the lack of negative selection against multiple copies on the genome. In this scenario, only a minimum number of *Cab* genes may be required and selected for.

I have named the more divergent *Fcp* sequence, *Fcp2*, to emphasize its distinction from the *Fcp1* sequences. The categorizing of the *Fcp2* sequence as being a unique gene 'type' from the *Fcp1* sequences requires more detailed examination of the different FCP complexes and their functions. However, as the immunological analysis in Chapter 2 suggested, there is quite an intricate antennal system in *Heterosigma*. This would suggest that many other *Fcp* gene 'types' exist. If the sequence divergence between the various *Cab* gene types is any indication, it is unlikely that Southern blots using either the *Fcp1* or *Fcp2* cDNA probes, under the previously described hybridization conditions, have detected the full extent of the *Fcp* gene family in *Heterosigma*. If distinct LHC I and LHC II antennae exist in *Heterosigma*, as is likely, then the sequence divergence between them may be too large to be easily detect by hybridization under the conditions used in this study. The differential cross-reactions of the FCPs with the two different antisera would also suggest this (Chapter 2). Bhaya and Grossman (1993) identified several *Fcp* genes that were present on two different genomic clones, which may be on the same chromosome. These sequences range in amino acid similarity from 86-99%. It would be interesting to examine the functional differences between these clones, if any. The *Fcp* gene family in *Phaeodactylum* does not appear to be as large as that in *Heterosigma*, based on reported Southern hybridizations (Bhaya and Grossman 1993).

The *Fcp* genes of *Heterosigma* are highly expressed as indicated by the strong hybridization signal on the northern blot. The cDNA preferentially utilizes 28 of the possible

codons in a manner similar to the *Phaeodactylum Fcp* (Grossman et al. 1990). The trends seen in the analysis of highly expressed yeast genes are generally consistent with the *Fcp* codon usage patterns (Bennetzen and Hall 1982). The codon usage in the *Heterosigma Fcp* cDNA shows a very strong bias for G + C in the third codon position (80%, Table 4.1 B) and this trend is evident in the other *Fcp* cDNAs. However, the bias is not as prominent for the diatom *Fcp* genes (*Phaeodactylum*, 64%; *Odontella*, 72%) or the *Isochrysis Fcp* sequence (67%). The *Macrocystis Fcp* cDNA shows extreme bias for G + C in the third position with approximately 91% occurrence for the five *Fcp* genes. The total G + C content for all codon positions of the *Macrocystis Fcp* cDNA is, however, similar to the that of *Heterosigma*. The analysis of *Cab* genes in angiosperms has shown a bias for G + C occurrence in the third codon of gymnosperms (Jansson and Gustafsson 1990) and monocots (Brinkmann et al. 1987) but not dicots.

A high degree of codon bias for a particular gene has been correlated with the relative concentrations of particular tRNA molecules within the cell (Ikemura 1982). Codon bias is also more pronounced in abundantly expressed proteins in bacteria (Sharp and Li 1986) and yeast (Bennetzen and Hall 1982). Together, the coincidence of tRNA pools with non-random codon preference in abundantly expressed proteins would be consistent with an increase in translational efficiency. For highly expressed genes, the use of rare codons may result in a decreased rate or premature termination of translation (Robinson et al. 1984) and be selected against. However, abundance of a particular protein did not appear to be correlated to the degree of codon bias in terrestrial plants (monocots) (Campbell and Gowri 1990) so the universality of such a proposal remains uncertain. Alternatively, the differences in G + C bias may be more related to the occurrence of different G/C ratios within certain regions of the eukaryotic genome, called isochores (for example, see Sharp 1991), as occurs in mammalian nuclear DNA.

The gene encoding the P-type ATPase from *Heterosigma akashiwo* (which is equivalent to *H. carterae*, see section 1.1) has just recently been published (Wada et al. 1994). It also shows codon usage bias, with 73% of the third codon positions being G or C. However, there is no bias for pyrimidines in the third codon position, as is the case with the *Heterosigma Fcp*1

cDNA. In addition, the degree of preference for the predominant codons in the ATPase gene is also not as distinct as in the *Fcp1* cDNA. The significance of the difference in bias between the two genes from *Heterosigma* is not known.

4.4.2 Structural aspects of sequence comparison

The similarity of the FCPs to the CABs was confirmed when the sequence of the first *Fcp* cDNA was determined (Grossman et al. 1990). In addition, there were three putative membrane spanning regions within the mature protein. As membrane spanning regions typically form α-helicies, the prediction of three transmembrane domains in the FCPs was consistent with the amount of measured α-helical structure (Hiller et al. 1987). When the *Heterosigma* FCP is compared to other FCPs or to the CABs, most of the conserved amino acids are within the membrane spanning, hydrophobic regions. This is particularly obvious within the first and third membrane spanning regions where there are probably considerable selective pressures against changes due to their importance in chlorophyll and carotenoid binding. The four amino acids that form the ionic bonding pairs between MSR 1 and MSR 3 in the pea LHC II structure are conserved in the Chl *a* + *c*-binding proteins (E 71—R 174; R 76—E169, Fig. 4.15). In addition to binding chlorophyll, these residues are thought to be the main protein stabilizing force within the lipid bilayer (Kühlbrandt et al. 1994). Moreover, within the area of close contact between these two α-helical membrane spanning regions—on the sides that face one another—there are a number of conserved smaller residues, probably selected for close packing (Green and Kühlbrandt 1995). In addition, one of the two Met residues (Met 73, Fig. 4.14), involved in the sterospecific binding of the internally located carotenoids in pea, is conserved in the Chl *a* + *c*-binding proteins. However, the position analogous to Met 188 in the pea structure is not conserved in the FCPs (or in all CABs) and its role may be fulfilled by a conserved Gln residue. Overall, this indicates that the structure and topology of the FCPs and the *i*PCPs are very similar to the CABs.

However, there are some noticeable differences between the two main LHC types (CABs and FCPs). First, compared to the CABs, the FCPs are shorter in the connectors joining the transmembrane regions, in the C-terminus and in the N-terminus, which accounts for their smaller size on polyacrylamide gels. The shorter connecting regions may be an indication that the FCPs helicies are more tightly packed together than those of the CABs (Green and Kühlbrandt 1995). Second, the two-fold symmetry that exists between the surface exposed regions just before the first and third MSRs in pea appears to be lacking in the FCPs since these areas are not conserved. What effect this may have on the organization of the FCPs is not known. Despite these differences, the overall structural relatedness clearly indicates that the different LHC types (CAB/FCPs etc.) are evolutionarily related and share a common ancestor. This also confirms the immunological work in chapter 2, and elsewhere (Caron et al. 1988; Passaquet et al. 1991; Hiller et al. 1993; Plumley et al. 1993; Durnford and Green 1994), that indicated the CABs and the FCPs are structurally related and share common antigenic epitopes.

Since the 3-D structure of the pea LHC II has been determined (Kühlbrandt et al. 1994), it is possible to use this information in locating putative pigment binding residues in other LHCs. Overall, the conservation of six of the seven putative Chl a-binding residues in the FCPs would suggest that these residues also function in the binding of chlorophyll. Presumably these residues would bind Chl a, but the possibility of some of these residues ligating Chl c or no chlorophyll at all can not be dismissed. The ligand of Chl $a7$ in pea is not known and the region under this Chl in the CABs (WFKAG; Fig. 4.14) is not conserved in the FCPs (Green and Kühlbrandt 1995). As this Chl is not bound by a side chain group, it is not possible to accurately assess whether it may be present in the FCPs.

Three Chl b ligands in the pea LHC II structure have been tentatively identified (Kühlbrandt et al, 1994). The alignment in Figure 4.13 suggests that two of these putative Chl b ligands, within MSR2 in pea (Gln 131 and Glu 139, indicated by open triangles), are conserved in the FCPs and iPCPs. However, a gap was inserted in the CAB sequence to maximize similarity to the other LHCs. In an α-helical structure, these residues would be on opposite

124

sides; therefore, it is unlikely that both would bind chlorophyll in the FCPs/iPCPs (Green and Kühlbrandt 1995).

The first of these conserved amino acids (E134, Fig. 4.12, 4.15) probably does not bind chlorophyll in the FCPs since most lack an Arg residue at position 137. In the pea LHC II, an Arg in a homologous position interacts with the conserved Glu to bind Chl $b5$ (Kühlbrandt et al. 1994). However, the presence of an Arg three positions away from the conserved Glu in the second MSR of the *Isochrysis* FCP, suggests that this residue may be a chlorophyll ligand in this case (see Fig. 4.12). Nonetheless, this Glu is very conserved in all FCPs and iPCPs and, if not binding chlorophyll, probably has another important function (Green and Kühlbrandt 1995).

Another amino acid with the potential to bind chlorophyll in the *Heterosigma* FCP is position 123 (V123 Fig. 4.12, 4.15). This could occur via a backbone carbonyl group because it would not have a H-bonding partner in the α-helix due to a conserved proline residue at position 119 in the FCPs (Fig 4.12, 4.15). Interestingly, this residue would be on the same side as the conserved Glu in the second MSR if an α-helix is formed (E 134, Fig. 4.12, 4.15). As well, the conserved Gln (Q125, Fig. 4.15) could provide a chlorophyll ligand, but not in addition to the former two because they would be on different faces of the α-helix. The FCPs probably lack a chlorophyll binding residue near the C-terminus, as occurs in the pea LHC II (position 211, Fig. 4.13), because this region is not conserved and it is absent in the *Phaeodactylum* FCP sequence.

An accurate determination of the Chl a/c ratio of the FCPs has not been done although estimates in the range of 1.4-5.6 have been reported (table 1.3). With an average Chl a/c ratio around 3 and assuming at least six bound Chl a molecules (since six of the residues thought to bind Chl a in the pea LHC II are conserved in the FCPs), there would be approximately two Chl c molecules per FCP polypeptide. In addition, anywhere from 5 to 12 fucoxanthin molecules may be present based on calculated molar ratios of Chl a, c and fucoxanthin (Friedman and Alberte 1984; Katoh et al. 1989). Although there are a few putative chlorophyll ligands in the second MSR of the FCPs, it is not possible to conclude if they would bind accessory chlorophylls (Chl c) as they are suspected of doing in the CABs (Chl b).

Other domains besides the membrane spanning regions may participate in the chlorophyll or carotenoid binding. The conserved domain in front of MSR1 in both the FCPs and CABs suggests that this region may be important for this purpose. In the pea LHC II structure, this region shields Chl $a4$ and lutein2 from the aqueous environment and may be involved in binding one of the centrally located lutein molecules (Kühlbrandt et al. 1994; Green and Kühlbrandt 1995). Presumably this region has an analogous function in the FCPs and may, in addition, provide a site for the binding of fucoxanthin. Similarly, in pea LHC II the position immediately preceding MSR3 is positioned above Chl $a1$ and lutein1 and is thought to be important in their binding (Kühlbrandt et al. 1994; Green and Kühlbrandt 1995). Although there is no obvious sequence conservation amongst the FCPs in this domain, it still may be important for both chlorophyll and carotenoid binding. It is likely that the polar groups at each end of fucoxanthin and the other carotenoids form hydrogen bonds to polar groups within the MSR connector regions. Similar interactions with the carboxyl group on Chl c, due to the lack of a phytol tail, could help bind or stabilize this molecule. Though not well conserved in the FCPs, the other domain linking MSR1 to MSR2 and the C-terminal section may have a role in either chlorophyll or carotenoid binding, as suspected for the pea LHC II structure (Kühlbrandt et al. 1994).

4.4.3 The transit sequence and protein import

The N-terminal region of the FCP protein from *Heterosigma* resembles a signal sequence that targets proteins to the ER in eukaryotes. The targeting of nuclear encoded, plastid localized proteins in chromophyte algae was first hypothesized to be mediated by a eukaryotic-like signal sequence by Sarah Gibbs (1979). The presence of a signal sequence-like transit peptide in plastid localized precursors was confirmed by Grossman et al. (1990) when the first nuclear encoded *Fcp* gene from a chromophyte was sequenced. The presequence of the *Heterosigma* FCP can be separated into several regions: a positively charged amino terminus (residues 1-4), a

hydrophobic section (residues 5-16), followed by a more polar region preceding the cleavage site. Though primary sequence conservation is not usually apparent, these are general characteristics of all signal peptides (von Heijne 1990).

The presence of a signal sequence would correlate well with the ultrastructure of the chromophyte plastid which has two additional membranes surrounding it (Gibbs 1970). The outermost membrane has ribosomes bound to the outer surface, like ER, and is commonly called the chloroplast ER (CER). The presence of a signal sequence offers an explanation for how nuclear encoded precursors may cross these two additional membranes. This was followed up by Bhaya and Grossman (1991) where cotranslational transport and processing of the FCP precursor was observed in an *in vitro* microsomal (ER) membrane system. In addition, terrestrial plant chloroplasts were not able to import the FCP precursor, illustrating the difference in transit sequence specificity. These observations lead to the proposal that the synthesis of chloroplast precursors occurred on the CER bound ribosomes and were cotranslationally transported through the membrane (Bhaya and Grossman 1991).

The presence of a putative bipartite transit sequence in the Odontella γ-ATPase subunit suggests that after the removal of the signal sequence, the remaining peptide directs the protein across the two membranes of the chloroplast envelope (Pancic and Strotmann 1993). The presequence of this protein is much larger than those of the FCP sequences shown in Figure 4.16. If the FCP presequence is bipartite, then the resulting transit peptide after the cleavage of the putative signal sequence would share some characteristics with the chloroplast transit peptides of plants (Bhaya and Grossman 1991; Apt et al. 1994). This similarity includes a high proportion of hydroxylated and basic amino acids. However, with the FCPs, this putative second leader sequence is predicted to be quite small compared to a typical plant transit peptide. It would be interesting to test various portions of the last half of the *Heterosigma* FCP presequence for transport competence into the pea *in vitro* chloroplast system.

In addition, it has been suggested that the transport of chloroplast precursors, following translocation through the outer membrane, occurs via membranous vesicles. These vesicles have

been observed between the two sets of surrounding chloroplast membranes in the chromophytes and are referred to as the periplastidal reticulum (Gibbs 1979). The periplastidal reticulum was hypothesized to pinch off from the CER and fuse with the outer chloroplast envelope, apparently depositing protein precursors into the lumen of the chloroplast envelope. If so, the bipartite sequence may still be important in directing the precursor across the inner chloroplast envelope.

An interesting exception is with the haptophyte FCP presequence. Though a hydrophobic region is present, the N-terminal region does not contain a basic amino acid which is a standard part of a signal sequence. This region is relatively rich in hydroxylated amino acids as compared to the other FCPs. It has been suggested that an unformylated initiation Met residue can compensate for the lack of a basic amino acid and that there is not an absolute requirement for a positive charge (von Heijne 1990). Otherwise, the *Isochrysis* FCP presequence is unlike that of other chromophytes. The significance of these differences, in terms of the transport mechanism, needs to be investigated.

CHAPTER 5

A phylogenetic analysis of the LHCs

5.1 Introduction

In this study I was interested in investigating the evolution of the two main LHC gene families: the *Cab*s and the *Fcp*s. This analysis involves three lines of investigation. (1) I was interested in analyzing the relationships amongst the different gene members of the *Cab* family. Since most of the genes from the tomato *Cab* gene family have been cloned, sequenced and well characterized (Green et al. 1991), it provided an opportunity to examine the evolutionary relationships amongst them and relate these trends to the proposed functions of the different protein complexes. (2) I wanted to examine the relationships between the known chromophyte *Fcp* and *iPcp* gene sequences; however, as there are few known *Fcp* sequences, this analysis will be limited. (3) I also wanted to examine the evolutionary relatedness of the FCPs to the CABs. Analysis of the relationships between the different *Cab* gene types to the *Fcp* genes may provide information as to when the divergence of the chromophyte antennae occurred. In addition, the usefulness of the LHCs in assessing organismal phylogeny will be discussed.

5.2 Methods

5.2.1 Protein Alignment

Phylogenetic analyses were carried out on the inferred protein sequences from a number of different light harvesting antennal proteins, from a diverse range of organisms. An original alignment was done using MACAW 2.02 (Schuler et al. 1990) with the 250 PAM matrix (Dayhoff et al. 1978) scale. Refinements to this alignment were made by hand using the Genetic Data Environment (GDE) software on a SUN workstation. The process of generating any sequence alignment carries with it the assumption of positional homology between the residues being compared. Highly divergent areas of the protein, which were impossible to align unambiguously, were omitted from the analysis. This was done to avoid the comparison of non-homologous regions of the protein which would violate the first assumption, positional homology, and could significantly alter the resulting tree topology (see Schlegel, 1991). Furthermore, regions of ambiguous similarity and multiple insertion/deletion events (indels) probably represent areas of 'multiple hits' and contain little, if any, phylogenetic signal (Swofford and Olsen 1990). Gaps were introduced in the alignments to maximize similarity. Those that were larger than one amino acid were treated as a single deletion event in the analysis.

The exact residues used in each analysis are indicated in Table 5.1 by residue number according to the labeling in Figure 5.1. The number of residues varies depending on the degree of sequence conservation between the taxa selected for the analysis. Amino acids were analyzed, rather than nucleotides, for a three reasons: (1) the third nucleotide of the codon tends to become randomized over large phylogenetic distances and is expected to have little phylogenetic signal, (2) the protein sequence is somewhat less sensitive to biases in the G + C content, and (3) it allowed the inclusion of LHC sequences that were determined only at the amino acid level (Swofford and Olsen 1990).

Table 5.1   Amino acid characters used in the phylogenetic analyses

dataset	characters*
Tomato CABs (Fig. 5.3)	48-71; 79-111;160-192; 219-223; 225-232, 241-286
Total CABs (Fig. 5.4)	as above
Green algal CABs (Fig. 5.5)	48-71; 79-111;160-192; 219-231, 241-301
FCP analysis (Fig. 5.6)	48-73; 79-111;140-188; 191-198; 241-275
Total LHC analysis (CABs/FCPs) (Fig. 5.7)	48-71; 79-111;159-194; 241-281

*as labeled in Figure 5.1

5.2.2 Phylogenetic analysis

Phylogenetic analyses were done with PHYLIP version 3.5 (Felsenstein 1992) using both parsimony (PROTPARS) and distance matrix (PROTDIST) algorithms. Both algorithms were used to determine if putative taxon relationships were consistent between each tree construction method. However, consistency between methods is not necessarily a good indication that the derived tree topology is the correct one (Felsenstein 1992). Comparing trees generated using the distance method should help in assessing which taxa may be rapidly evolving, based on the length of the branches in the tree. This should give some idea as to the likelihood of artifacts occurring in the parsimony tree since parsimony tends to fail under circumstances when the rate of change between different taxa is large (Schlegel 1991). In such cases, there is a tendency to group faster changing sequences together; a case referred to as long branches attracting (Felsenstein 1978). As well, distance methods tend to be sensitive to the number of taxa in the analysis, which can alter tree topology (Schlegel 1991).

Parsimony trees were done with the jumble option in effect, which randomizes the order in which each taxon is added to the analysis. This was repeated ten times for each tree since the

131

input order of the taxa can influence the final tree topology (Felsenstein 1992). Changes in amino acids were assigned a value corresponding to the number of minimal mutational steps required to cause such a change, based on the genetic code. Mutations resulting in a synonymous amino acid change are not included in the calculation. It is assumed that they occur much more frequently than non-synonymous amino acid changes and are phylogenetically unimportant (Felsenstein 1992). With the exception of omitting ambiguously aligned regions, which assigns a weight of zero to these areas, no external weighting was used. If more than one most parsimonious tree was found then a consensus tree was shown. Consensus trees include branch topologies that occur most frequently; branch relationships appearing in more than 50% of the trees of equal length (equally parsimonious) are definitely included (Felsenstein 1978).

Distance matrix analyses utilized the Dayhoff accepted point mutation (PAM) matrix (Dayhoff et al. 1978) in the calculation of distances between the different taxa. This is an empirical matrix which assigns a probability for the conversion of one amino acid to another. Gaps are treated as unknown amino acids and dropped from the calculation (Felsenstein 1992). From the distance matrix, the tree topology was determined using the neighbor-joining method (Saitou and Nei 1987). The neighbor-joining method determines tree topology by starting with a star-like tree and successively clustering taxa into neighbors. Neighbors are taxa that are connected by a single interior node and are clustered on the basis of a calculated minimum distance between them. The joining of neighbors continues until only one possible unrooted tree exists. The distance values between each taxon were randomly inputted using the jumble option.

An estimate of branch stability was assessed by a bootstrap analysis with 100 replicates. Bootstrapping is a process that randomly resamples the character sites (eg. amino acids at a specific position) until a dataset the same size as the original is obtained; the number of taxa in the analysis does not change. This is repeated any number of times, as determined by the user. Depending on the size of the dataset and the number of taxa, bootstrap trials can take a considerable amount of computer time. Because of this, trials are routinely limited to 100 replicates. The phylogenetic analysis is then done on each of the replicated datasets and the

results are expressed as the number of times, out of 100, a particular node on the resulting tree was supported (Felsenstein 1985). The bootstrap values are typically displayed directly on the branches of the tree they correspond to. It is an estimate of whether the tree is likely to change if the number of characters in the analysis had been larger. Bootstrap values should be considered as an indication of the consistency of the individual characters in determining a specific tree topology and not a probability that the relationships depicted are the true phylogeny. It is also important to realize that bootstrap analysis is not able to detect errors in phylogeny that are the result of systematic errors. These are errors that result when the evolutionary processes violate the assumptions of the phylogenetic method used (Swofford and Olsen 1990).

5.2.3 Terms and concepts

A phylogenetic tree is a type of dendrogram, ie., a branching diagram depicting hypothesized genealogical relationships of the taxa. The taxa are displayed at terminal nodes and the branches joining them are connected at internal nodes. The branching of the tree normally occurs in a bifurcating fashion. The trees in either method are constructed based on a specific criterion. This refers to the manner in which the evolutionary characters are weighted and assessed. Characters generally refer to a measurable observation of any trait pertaining to an organism or gene (taxon). In this study, a character refers to an amino acid residue located at a specific position within the protein. The identity of a character at a specific position (whether it is Ala, Leu, or Ser, etc.) is called the character state. A monophyletic group refers to two or more taxa that share the same ancestral taxon. Alternatively, groups are considered polyphyletic when the taxa within them are derived from two or more distinct ancestral genes or species. Paralogous sequences (eg. Lhca1 and Lhcb1) do not share the same evolutionary history as duplication and divergence of the gene occurred prior to the divergence of the organism (ie., a speciation event) (Schlegel, 1991). Genes of the same type (e.g. all Lhca4 sequences) are considered orthologous

133

because they diverged from a common ancestral sequence and should parallel the evolution of the organism (Schlegel, 1991).

The taxonomic positions of the organisms, from which the LHC sequences used in this analysis originated, are given in Table 5.2. The reader should also refer to Table 1.1, in the general introduction (Chapter 1), for further characteristics of the algal groups in question.

Table 5.2   Species and genes used in the phylogenetic analyses

Organism	Taxomonic position/gene	Reference
Lycopersicon esculentum	Angiosperm: dicot	
(tomato)	Lhcb1	Pichersky et al., 1985
	Lhcb2	Pichersky et al., 1987
	Lhcb3	Schwartz et al., 1991a
	Lhcb5	Pichersky st al., 1991
	Lhcb6a/b	Schwartz and Pichersky, 1990
	Lhca1	Hoffman et al., 1987
	Lhca2	Pichersky at al., 1987
	Lhca3	Pichersky et al., 1989
	Lhca4	Schwartz et al., 1991b
	psbS	Wallbraun et al., 1994
Spinacia oleracea (spinach)	Angiosperm: dicot	
	Lhcb6	Spangfort et al., 1991
	psbS	Kim et al., 1992
Arabidopsis thaliana	Angiosperm: dicot	
	Lhca1	Jensen et al., 1992
	Lhca2	Zhang et al., 1992
	Lhca3	Wang et al., 1993
	Lhca4	Zhang et al., 1991
	Lhcb4	Green and Pichersky, 1993
Lemna gibba (duckweed)	Angiosperm:monocot	
	Lhcb2	Karlin-Neumann et al., 1985
Gossypium hirsutum (cotton)	Angiosperm: dicot	
	Lhcb2	Sagliocci et al., 1992
Oryza sativa (rice)	Angiosperm: monocot	
	Lhcb1	Matsouka, 1990
Pinus sylvestris (Scots pine)	Gymnosperm	
	Lhcb1	Jansson and Gustafsson, 1990
	Lhcb2	Jansson and Gustafsson, 1990
	Lhcb5	Jansson and Gustafsson, 1992
	Lhca1	Jansson and Gustafsson, 1991
	Lhca2	Jansson and Gustafsson, 1991
	Lhca3	Jansson and Gustafsson, 1991
	Lhca4	Jansson and Gustafsson, 1992
Ginkgo biloba	Gymnosperm	
	Lhcb1	Chinn and Silverthorn, 1993
Polystichum munitum	Pteridophyta (fern)	Pichersky et al., 1990
Physcomitrella patens	Bryophyta (moss)	Long et al., 1989

Chlamydomonas reinhardtii	Chlorophyta	
	Lhcb	Imbault et al., 1988
	Lhca	Hwang and Herrin, 1993
Chlamydomonas moewusii	Chlorophyta	
	Lhcb	Larouche et al., 1991
Chlamydomonas stellata	Chlorophyta	
	Lhcb	Wolfe et al.,1993-unpubl.
	Lhca (2C.stellata)	pir S33466, S31393
Chlamydomonas eugametos	Chlorophyta	Gagné and Guertin, 1992
Dunaliella tertiolecta	Chlorophyta	
	Lhcb	LaRoche et al., 1990
Dunaliella salina	Chlorophyta	
	Lhcb	Long et al., 1989
Mantoniella squamata	Chlorophyta:	Rheil and Mörschel, 1993
	Prasinophyceae	
Euglena gracilis	Euglenophyta	Muchhal and Schwartzbach,
	Lhcb 1-4	1992
	Lhca (35 & 38)	Houlné and Schantz, 1988
Phaeodactylum tricornutum	Chromophyta: Diatom	Grossman et al., 1990
		Bhaya and Grossman, 1993
Odontella sinensis	Chromophyta: Diatom	Thelen & Pancic -gb 81054
Macrocystis pyrifera	Chromophyta: Brown alga	Apt et al., 1994
Heterosigma carterae	Chromophyta:	Durnford, D. this study
	Raphidophyta	
Isochrysis galbana	Haptophyta	LaRoche et al., 1994
Pavlova lutherii	Haptophyta	Hiller, R. unpublished*
Amphidinium carterae	Dinophyta: Dinoflagellate	Hiller et al., 1993 *

* Sequence determined at the protein level. However, recently the nucleic acid sequence of the gene encoding the dinoflagellate *i*PCP protein has been determined (gb Z47562, Z47563).

5.3 Results

5.3.1 Assessment of phylogentic signal

The datasets used in each analysis include the sequences shown in each Figure (Fig. 5.3-5.7) and the characters outlined in Table 5.1. The phylogenetic signal of the dataset was assessed by calculating the skewness of tree-length distributions from 10,000 randomly generated trees using the "random tree" option in the computer program PAUP (Phylogenetic Analysis Using Parsimony) (Swofford 1991). The random tree option randomly selects a tree topology and

135

calculates the length of that tree based on the dataset. It repeats this process 10 000 times and plots the tree lengths against the frequency of their occurrence. Random sequences produce nearly symmetrical distributions from a parsimony analysis of all possible tree lengths while those that contain a phylogenetic signal have a skewed distribution (Hillis et al. 1993). The characters in the FCP sequence dataset (used in Fig. 5.6) were manually randomized and the random tree distribution calculated (Fig. 5.2A) to compare it to the random tree distribution produced with the non-randomized sequences (Fig. 5.2D). As expected, the distribution was normal as compared to the very skewed distribution of the non-randomized dataset (compare Fig. 5.2A & B). This indicates that this is a useful method of assessing the potential phylogenetic signal in the datasets.

The tomato CAB (Fig. 5.2B) and FCP (Fig. 5.2D) datasets had a stongly skewed random tree distribution. The green algal dataset also showed a strongly skewed distribution, though to a lesser extent (Fig. 5.2C). This indicates that these datasets have not diverged beyond the point of being a potentially useful phylogenetic indicator. The last two datasets (Fig. 5.2E and F) give only a moderately skewed distribution. Though this is a conservative estimate, it indicates that the dataset may be approaching a limit of change where convergent or back mutations have become as frequent as divergent mutations (Meyer et al. 1986) and is approaching its limit as a phylogenetic indicator. However, PAUP does not assign weights to amino acid changes, nor does it reflect the genetic code, as PHYLIP does. Without weights, PAUP essentially calculates the similarity based on amino acid identity. This underestimates the true relatedness of the sequences and the random tree distributions should be considered a lower limit estimate of phylogenetic signal. In addition, since PAUP does not use a character weighting scheme, the random tree lengths in Figure 5.2 can not be compared directly to the length of the trees I will show.

136

Figure 5.1

Amino acid sequence alignment of selected sequences used in the phylogenetic analysis. Numbers on top of the alignment indicate character positions and are not an indication of the protein size. Regions of some proteins have been omitted to save space, and have been indicated by the presence of a *. The sequences are arranged by gene/protein type followed by the species it is from. Gene types are: Lhca and b, Light harvesting complexes associated with PS I or PS II, respectively, according to the nomenclature of Jansson et al. (1992); *i*PCP, *intrinisic* peridinin-chlorophyll protein; FCP, fucoxanthin-chlorophyll protein. The taxa are: Ac, *Amphidinium carterae*; At, *Arabidopsis thaliana*; C.eugam., *Chlamydomonas eugametos*; Cr, *Chlamydomonas reinhardtii*; Cs, *Chlamydomonas stellata*; Ds, *Dunaliella salina*; Eg, *Euglena gracilis*; Gb, *Ginkgo biloba*; Hc, *Heterosigma carterae*; Ig, *Isochrysis galbana*; Le, *Lycopersicon esculentum* (tomato); Lg, *Lemna gibba*; Mp, *Macrocystis pyrifera*, Ms, *Mantoniella squamata*; Os, *Odentella sinensis*; Pl, *Pavlova lutherii* (partial sequence); Pm, *Polystichum munitum*; Pp, *Physcomitrella patens*; Ps, *Pinus sylvestris*; Pt, *Phaeodactylum tricornutum*; So, *Spinacia oleracea*. Gaps in sequence are indicated by a -.

137

MSR 1

```
              1          11         21         31         41         51         61         71         81         91         101        111
              |          |          |          |          |          |          |          |          |          |          |          |
Lhcb6a-Le   *LLGAHVTAR  V--TTP---K  RFVVAAAAVA  PKKSWIPAVK  SGGNLVDPEW  LDGSLPGDFG  FDPLGLGKDP  A-------FL  KWYREAELIH  GRWAMAAVLG  IFVGQAWSGI  P----
psbS-So     *KSKAK-APK  KVEKPKLK-V  ED--------  ----------  ----------  ----------  ---GLFGTS   G-------GI  GFTKENELFV  GRVAMIGFAA  SLLGEGITGK  GIL--
Lhcb5-Le    *KKKAAAAPA  KAKAAAVSPA  DDELAKWYGP  DRRIFLPEGL  LDRSEI-PEY  LNGEVPGDYG  YDPFGLSKKP  E-------DF  AKYQAYELIH  ARWAMLGAAG  FIIPEAFNKF  GA---
Lhcb4-At    *GSGRFTAVF  GFGKKK-AAP  KKSAKKTVTT  DRPLWYPGAI  S------PDW  LDGSLVGDYG  FDPFGLGKPA  EYLQFDI*GI  QRFRECELIH  GRWAMLATLG  ALSVEWLTGV  T----
Lhca1-Le    *PSFLSSTKS  KFAAAMPVYV  GATNFMSRFS  MSADWMPGQP  R------PSY  LDGSAPGDFG  FDSLGLGEVP  A-------NL  ERYKESELIH  CRWAMLAVPG  IIVPEALGLG  N----
Lhca2-Le    *LGGRRLRVS  KYSTTPTARS  ATTVCVAADP  DRPLWFPGST  P------PPW  LDGSLPGDFG  FDPLGLASDP  E-------SL  RWNQQAELVH  CRWAMLGAAG  IFIPELLTKI  G----
Lhca4-Le    *LTGSSGKLN  REVSFRPSTS  SSYNSFKVEA  KKGQWLPGLA  S------PDY  LDGSLPGDNG  FDPLGIVEDP  E-------NL  KWFIQAELVN  GRWAMLGVAG  MLLPEVFTSI  G----
Lhca3-Le    *IGSRISQSV  TRKASFVVRA  ASTPPVKQGA  NRQLWFASKQ  S------LSY  LDGSLPGDFG  FDPLGL-SDP  EGTGGFI-EP  KMLAYGEVIN  GRFAMLGAAG  AIAPEILGKA  GLIP-
Lhcb1-Le    *GNGRITMRK  AVAKSA----  -PSSSPWYGP  DRVKYLGPFS  GES----PSY  LTGEFPGDYG  WDTAGLSADP  E-------TF  AKNRELEVIH  CRWAMLGALG  CVFPELLARN  GV---
Lhcb2-Le    *GEGRITMRR  TVKSAP----  ---QSIWYGE  DRPKYLGPFS  EQT----PSY  LTGEFPGDYG  WDTAGLSADP  E-------TF  AKNRELEVIH  CRWAMLGALG  CVFPEILSKN  GV---
Lhcb3-Le    *--NANPLRD  VVAMGS-ARF  TMSNDLWYGP  DRVKYLGPFS  AQT----PSY  LNGEFPGDYG  WDTAGLSADP  E-------AF  AKNRELEVIH  GRWAMLGALG  CIFPEVLEKW  VKV--
Lhcb-Gb     *GEGRITMRK  TASKKV---V  ASSGSPWYGP  DRVKYLGPFS  GEA----PSY  LTGEFPGDYG  WDTAGLSADP  E-------TF  AKNRELEVIH  SRWAMLGALG  CVFPELLSRN  GV---
Lhcb2-Lg    *SDGRFSMRR  TVKAVP----  ---QSIWYGA  DRPKFLGPFS  EQT----PSY  LTGEFPGDYG  WDTAGLSADP  E-------TF  AKNRELEVIH  SRWAMLGALG  CIFPELLSKN  GV---
Lhcb-Pp     *VEARVTMRR  TVSKSA----  -GSDTIWYGA  DRPKFLGPFS  GET----PSY  LNGEFAGDYG  WDTAGLSDDP  E-------TF  ARNRELEVIH  ARWAMLGALG  CLTPELLAKS  GV---
Lhcb1-Ps    *GEARVTMRK  ATTKKL--TA  SASTSPWYGP  DRVLYLGPFD  GEP----PSY  LTGEFPGDYG  WDTAGLSADP  E-------NF  AKNRELEVIH  CRWAMLGALG  CVFPELLARN  GV---
Lhcb-Pm     *GEARVQMMA  PKSKAP----  --SGGSIWYGS  DRPLYLGPFS  GSP----PSY  LSGEFPGDYG  WDTAGLSADP  E-------TF  AKNRELEVIH  SRWAMLGALG  CVTPELLAKN  GV---
Lhcb-Cr     *LQVTCKATG  KKTAAKAAAP  KSSGVEFYGP  NRAKWLGPYS  ENAT---PAX  LTGEFPGDYG  WDTAGLSADP  E-------TF  KRYRELELIH  ARWAMLGALG  CQTPELLAKS  GT---
Lhcb-Ds     *QGGRAGKST  KKGAKAVSKS  SSSANQFYGP  DATSGWDLQH  QH-----PRL  PTGEFPGDYG  WDTAGLSADP  E-------TF  KRYRELELIH  ARCGLLGALG  MVTPELLADE  DGIK-
Lhca-Cs     *--AALVANK  AALGFRAPAQ  ARRSVVARAE  QRASWFPGNP  A------PEY  LKGELAGDYG  FDPLGFGDEP  R-------RL  EMMVQAELVH  CRWAMLGVAG  ILLPEIGNKT  GMQSA
Lhca-Cr     *RTLSARTAA  PRGFSGRRVA  AVSNGSRVTM  KAGNWLPGSD  DA-----PAM  LPDDLPGNYG  FDPLSLGKEP  A-------SL  KRPTESEVIH  GRWAMLGVAG  SLAVELLGYG  N----
Lhca-Eg1    MLATSGRKA   -----KAAPK  SDNLSQWYGP  DRAKWLGPLT  GEV----PSY  LTGELPGDYG  WDTAGLGSDP  T-------TL  ARYREABVIH  ARWAMLGALG  VVTPELLAGN  GVP--
Lhca-Eg2                            MFASSGH     KDGLWFPNAE  P------PAL  LTGEYPADRG  FDPLNLAADP  D-------VY  ARMRVAEVFH  GRLAMLCTVG  CIVPELLGKG  A----
C.eugam.    *VRVQAAGKT  VSGSKTVSGG  KTVGSASAES  RRVAEVQAYL  ATL----PGC  GVBSGPFKGV  WDPLSLAATA  TVG-----DV  RWRWRESEITH GRVAMLAALG  FVVGEQLEDF  PA---
Lhcb-Ms     ----------  MACIASSFVG  SVAALKATKV  QAKSVSTVVK  ADIY---PEF  GT--YPG--G  GESPIIPF--  --------GS  EKNAEREVIH  GRWAMLGVTG  AWTTENGTGI  P----
iPcp-Ac     ----------  ----------  ----------  ----------  -------FEN  ERGVQDPVGF  FDPLGFTADG  SVE-----NF  KRLAQTEIKH  GRVAMLATMG  YITPEITGKL  G----
Fcp-Hc      MSLKL--A    TFA-AALMGA  SAFVAPNRMG  VAKSSSALKM  S------FEN  EIGVQAPLGF  WDPLGLLDEA  DQE-----RF  DRLRTVEIKH  GRISMLAILG  HLVTTAGVRL  PG---
Fcp1-Mp     *MKSAVMAVA  CAAAPGFRGP  SAFNGAALTT  SAKACSAMKM  S------FES  EIGAQAPLGF  WDPLGLLADA  DQE-----RF  ERLRVVEVKH  GRIAMLAIAG  HLTQON-TRL  PG---
Fcp1-Os              MK  LAIAALLAGS  AAAFAPAQS-  -GKASTALNM  A------FES  EIGAQQPLGF  FDPLGMLADA  DQE-----RF  DRLRYVEVKH  GRIAHVAFLG  QIVTRNGIHL  SG---
Fcp1-Pt             MK  FAVFAFLLAS  AAAFAPAQ-Q  SARTSVATNM  A------FEN  EIGAQQPLGY  WDPLGLVADG  DQE-----KF  DRLRYVEIKH  GRICMLAVAG  YLTQEAGIRL  PG---
Fcp-Ig            MMT  LASLPSTAIA  GLASAAPKVQ  -PRMAANDEF  A------Y-G  LPGGANILGE  FDPAGMLKGK  DKL-----EV  YRLRAETTH   GRVAMLASLG  FVVQEK-FH-
Fcp-Pl      ----------  ----------  ----------  ----------  --------    ----APLGF   YDPLGLLDNA  DEE-----YF  ERM-YVEL--  ----------  ----------  ------
```

138

MSR 2

```
        121        131        141        151        161        171        181        191        201        211

Lhcb6a-Le  --------WF EAGADPGAIA P--------- ---------- ---FSFGSL LGTQLLLM-G WVESKRWVDF FDNDSQSIDW ATPWSKTAEN FANFTGEQG- --
psbS-So    ---------- ---------- --SQLNLE   TGIPIYE--- ----AEPLLL FFILFTLL-G AIGALGDRGR FVDEPTTGLE KAV------- ---------- --
Lhcb5-Le   -NCGPEAVWF KTG------- ---ALLLDGN TLNYFGKNIP I---NLILAV VA-EVVLV-G GAEYYRIING LDLEDKL--- ---------- ---------- --
Lhcb4-At   --------WQ DAG------- ---KVELVD  GSSYLGQPLP F---SISTLI WI-EVLVI-G YIEFQR-NAE LDSEKRL--- ---------- ---------- --
Lhca1-Le   --WVKAQEWA AIPGGQATYL GQPVPW---- ---------- ----GTLPTI LAIEFLAI-A FVEHQR-SME KDSEKK---- ---------- ---------- --
Lhcb2-Le   --ILNTPSWY TAGEQEYF-- ---------- ---------- ---TDTTTL FIVELVLI-G WAEGRRWADI IKPGCVNTDP IFPNNKLTGT DVG------- --
Lhca4-Le   --ILNVPKWY DAGKSEYF-- ---------- ---------- ----ASSSTL FVIEFILF-H YVEIRRWQDI KNPGSVNQDP IFKNYSLPPN KCG------- --
Lhca3-Le   --QETALAWF QTG------- ----VIPPA  GTYNYW---- ---ADNYTL FVLEMALM-G FAEHRRFQDW AKPGSMGKQY FLGLEKGLGG SGDPA----- --
Lhcb1-Le   --KFGEAVWF KAG------- ----SQIFSE GGLDYLGNPS LVHAQSIIAI WACQVVLM-G AVEGYRIAGG PLGEVVDPL- ---------- ---------- --
Lhcb2-Le   --KFGEAVWF KAG------- ----SQIFSE GGLDYLGNPN LVHAQSILAI WACQVVLM-G FVEGYRVGGG PLGEGLDKI- ---------- ---------- --
Lhcb3-Le   --DFKEPVWF KAG------- ----SQIFSD GGLDYLGNPN LVHAQSILAV LGFQVVLM-G LVEGFRINGL PGVGEGNDL- ---------- ---------- --
Lhcb-Gb    --KFGEAVWF KAG------- ----SQIFSD GGLDYLGNPS LIHAQSIIAI WACQVILM-G AVEGYRIAGG PLGEITDPI- ---------- ---------- --
Lhcb2-Lg   --QFGEAVWF KAG------- ----AQIFSE GGLDYLGNPN LVHAQSIIAI WATQVVLM-G LIEGYRVGGG PLGEGLDPL- ---------- ---------- --
Lhcb-Pp    --KFGEAVWF KAG------- ----AQIFSE GGLDYLGNPS LVHAQSILAI WACQVVLM-G AVEGYRVAGG PLGEVTDPI- ---------- ---------- --
Lhcb1-Ps   --KFGEAVWF KAG------- ----AQIFSE GGLDYLGNPS LVHAQSILAI WACQVILM-G AVEGYRIAGG PLGEVTDPI- ---------- ---------- --
Lhcb-Pm    --KFGEAVWF KAG------- ----SQIFAE GGLDYLGNPS LVHAQSILAI WACQVILM-G AVEGYRVAGG PLGEVEDPI- ---------- ---------- --
Lhcb-Cr    --KFGEAVWF KAG------- ----AQIFSE GGLDYLGNPS LVHAQNIVAT SAVQVVLM-G LIEGYRVNGG PAGEGLDPL- ---------- ---------- --
Lhcb-Ds    --FGDAAIWF KAG------- ----AAIFQD GGLNYLGNPS LIHAQNIVAT LAVQVVLM-G LVEGYRVNGG PAGEGLDPL- ---------- ---------- --
Lhca-Cs    APGLAFPHWW EAG------- ----KV-VAE SGP------- ---WSMGSL TFTMPLLM-G WAEHKRLYDF VKPGSQAVGW AWAPLGLGSI TGVEEPLVAK ENG
Lhca-Cr    --WYDAPLWA VNG------- ------GKATWFG IEVPFDLNAL LAREFVAM-A AAEGQRGDEG QRGDAGGVV- ---------- ---------- --
Lhcb-Eg1   --FGEGAVWY KAG------- ----AQIFSA DGLNYLGNPS LIHAQSVVLT FLSTLAIM-G AVEAYRYGGG VGDFGRELDT L--------- ---------- --
Lhca-Eg2   -----WF    EAGDS----- ---------- ---------- ---VDGLKL GFITMAIA-A PTEYWRGQGG FNWQKGE-LD RS-------- ---------- --
C.eugam.   -----FF    N--------- -------FD  GHI--TGQAI KQFDQVQQGF WEPLLIAI-G LAESYRVSLG WATPTGTGFN NLKDE----- ---------- --
Lhcb-Ms    --------WF TAGTLCTPDD CTAVADKFP- GAVAPLAPEG SGY-PSFNAV LAIEVVLV-G LABAYRTGLS DSPFEELTVG DV-------- ---------- --
iPcp-Ac    ---------- ---------- --YLSPST   GVKYDDIPNG LGAISKVPAA GWGQMIAYAA FSELSQDQSA STPAAEGAFG FKVLT----- ---------- --
Fcp-Hc     ---------- ---------- --AYDL-A   GDQPSSLPTG LKALSALPAA GVAQTIGFIG LIELGFAQIK EELE--ADCE ARM------- ---------- --
Fcp1-Mp    ---------- ---------- --MLSNSA   NLSFADMPNG VAALSKIPPA GLAQIFAFIG FLELAVMKNV BGS---FFGD FTLG------ ---------- --
Fcp-Os     ---------- ---------- --NIDYAGNS FDSFPNGWAA ISGPDAIFQA GLLQIVAFVG ILELAVMKDV TGEG-EFPGD FR-------- ---------- --
Fcp1-Pt    ---------- ---------- --DIDYS-   GTSFESIPNG FAALSAVPGA GIAQIIAFIG FLEIAVMKDI TGG--EFVGD FR-------- ---------- --
Fcp1-Ig    ---------- ---------- --PLFSGDN  GPAIEQIP-- ----QLPYWL WIVMTIG-IG RAELFRIQKG WAKVNPETGK ADSALREG-- ---------- --
Fcp-Pl     ---------- ---------- ---------- ---------- ---LEVVPAA GLAQIVAFVG FLELPV-KDV TGEG-DFPGD F--------- ---------- --
```

MSR 3

```
          221     231            241        251        261        271        281        291        301

Lhcb6a-Le YPGGK-F FDPLALAGTLNNGVY VPDTEKLERL KLAEIKHSRL AMLAMLIFYF EAGQ--GKTP LGALGL
psbS-So   IPPGKDV RSALGL--------- KTKGPLFGPT KSNELFVGRL AQLGFAFSLI GEIIT-GKGA LAQLNI-ETG VPINEIEPLV LLNVVFFFIA AINPGTGKFI TDDEEED
Lhcb5-Le  HPGGP-- FDPLGL--------- AKPDQAAIL  KVKEIKNGRL AMFSMLGFPI QAYVT-GQGP VENLAAHLSD PFGNNLLTVI GGASERVPTL
Lhcb4-At  YPGGK-F FDPLGL--------- AADPEKTAQL QLAEIKHARL AMVAFLGFAV QAAAT-GKGP LNNWATHLSD PLHTTIIDTF SSS
Lhca1-Le  YPGGA-- FDPLGY--------- SKDPAKFEEL KVKEIKNGRL ALLAIVGFCV QQSAYLGTGP LENLATHLAD PWHNNIGDVI IPKGIFPN
Lhca2-Le  YPGGL-W FDPLGWG-------- SGSPAKIKEL RTKEIKNGRL AMLAVMGAWF QHIYT-GTGP IDNLFAHLAD PGHATIFAAF SPK
Lhca4-Le  YPGGI-- FNPLNF--------- ----APTEEA KEKELANGRL AMLAFLGFIV QHNVT-GKGP FDNLLQHLSD PWHNTIIQTL SN
Lhca3-Le  YPGGPL- FNPLGF--------- GKDEKSMKEL KLKEIKNGRL AMLAILGYFI QALVT-GVGP YQNLLDHLAD PVNNNVLTSL KFH
Lhcb1-Le  YPGGS-- FDPLGL--------- AEDPEAFAEL KVKEIKNGRL AMFSMFGFFV QAIVT-GKGP LENLADHLAD PVNNNAWAFA TNFVPGK
Lhcb2-Le  YPGGA-- FDPLGL--------- ADDPEAFAEL KVKEIKNGRL AMFSMFGFFV QAIVT-GKGP IENLSDHIND PVANNAWAYA TNFVPGK
Lhcb3-Le  YPGGQ-Y FDPLGL--------- ADDPTTFAEL KVKEIKNGRL AMFSMFGFFV QAIVT-GKGP LENLLDHLDN PVANNAWVYA TKFVPGA
Lhcb-Gb   YPGGS-- FDPLGL--------- ADDPDAFAEL KVKELKNGRL AMFSMFGFFV QAIVT-GKGP IENLADHLAD PVNNNAWAFA TNFVPGK
Lhcb2-Lg  YPGGA-- FDPLGL--------- ADDPEAFAEL KVKEIKNGRL AMFSMFGFFV QAIVT-GKGP IENLADHIAD PVANNAWAFA TNFVPGK
Lhcb-Pp   YPGGS-- FDPLGL--------- ADDPDTFAEL KVKEIKNGRL AMFSMFGFFV QAIVT-GKGP LENLNDHLAD PVANNAWAYA PTSPPGTR
Lhcb1-Ps  YPGGS-- FDPLGL--------- AEDPEAFAEL KVKEIKNGRL AMFSMFGFFV QAIVT-GKGP IENLADHLAD PVNNNAWAYA TNFVPGK
Lhcb-Pm   YPGGS-- FDPLGL--------- ADDPEAFAEL KVKELKNGRL AMFSMFGFFV QAIVT-GKGP IENLSDHLAD PAVNNAWAYA TNFTPGK
Lhcb-Cr   YPGES-- FDPLGL--------- ADDPDTFAEL KVKEIKNGRL ASFSMFGFFV QAIVT-GKGP VQNLDDHLAN PTVNNAFAFA TKFTPSA
Lhcb-Ds   YPGEA-- FDPLGL--------- ADDPDTFAEL KVKEIKNGRL AMFACLGFFV QAIVT-GSGP IENLTDHLAN PAERNAFAYA TKFTPQ
Lhca-Cs   YPGGI-- FDFPQM--------- SKNEAKYKWY KQSEVTNGRL AMTAMFGFVS QHIAYPGKGP IDNLVDHVSN PTKVTFATNG VSVPHFIEYH
Lhca-Cr   YPGGA-- FDPLGF--------- AKDSSKSGEL KLKEIKNGRL AMVAFLGFVA QHAAT-GKGP IAALGEHLAN PWGANFATNG ISVPFF
Lhcb-Eg1  YPGGP-- FDPLGL--------- ANDPDALAEL KVKELKNGRL AMVAMLGFYV QPLVT-KAGP VENLTFHLAD PSANNIFFTS GFA
Lhca-Eg2  YPG---- FDPLNL--------- -----TTDYT RAAEVKNGRL ALTAVAGLTA QYLAT-GESP LANLSAHLAN PIGANITTNL A
C.eugam.  YDLGNLY FDPLGLK------- PEDPEELREL QTKELANNGRL AMIAIAGFVL QEV---AEPG VB-IFQHLFF TIEKDIVEEI DIIEKDLGLP PSFPVPTLPN LSS
Lhcb-Ms   SPFGR-- FDSLGL--------- AESGD-LEEL KIKELKHCRL SMFAWLGCIF QALAT-QEGP IANWQAHVAD PVHANVLTNA ASGFGFY
iPcp-Ac   ------- ------------- SSDPAELEKK LSAPIANGPL AMMAIIGMFF QDGLT-GSAR GDWASYTASP L
Fcp-Hc    ---DAAG WD------------ ---DEKKDSK RAIELNNGRA AQMGILALMV HEQLD-NNPY IINSLLGSPV DFNAGF
Fcp1-Mp   GNPFGAS WD------------ AMSEETQASK RAIELNNGRA AQMGILALMV HEELN-NKPY VINDLVGASY TFN
Fcp-Os    NGALDFG WD------------ TFDEETKLSK RAIELNNGRA AMMGILGLMV HEQLG-GSIP IVGEM
Fcp1-Pt   NNYLDFG WD------------ TFSEDKKLQK RAIELNQGRA AQMGILALMV HEQLG-VSIL P
Fcp-Ig    YEPGDLG FDPLGLA------- PSDPDEFRLM QEKELSHGRL AMIAAAGFLA QEAVS-GDTW GTYWGDATF
Fcp-Pl    ------- ------------- -------K   RNIELNQG-A A-M-ILALMV ------
```

140

Figure 5.2

The random tree distributions of the datasets used in the phylogenetic analyses (figures 5.3-5.7). For each dataset, the lengths of 10 000 randomly generated trees were calculated and plotted against their frequency of occurrence. (A) The order of the amino acids for each taxa in the FCP dataset (used in Fig. 5.6) were randomized manually and the distributions of the randomly generated tree lengths were calculated. The other graphs are the random tree distributions of the datasets used in the analysis of the following sequences: (B) the tomato CAB proteins (Fig. 5.3), (C) the green algal sequences (Fig. 5.5), (D) the FCP sequences (Fig. 5.6), (E) the CAB and FCP sequences (total LHC) sequences (Fig. 5.7 A/B), and (F) the CAB sequences (Fig. 5.4 A/B). The mean (⇓) and the median (↓) of each distribution is indicated on the graph.
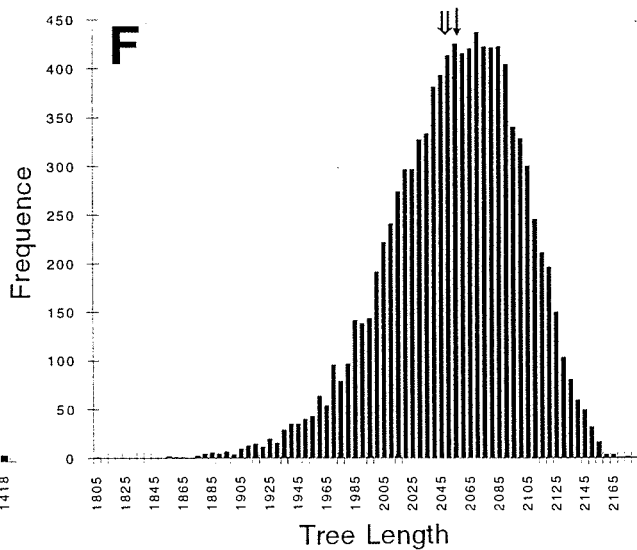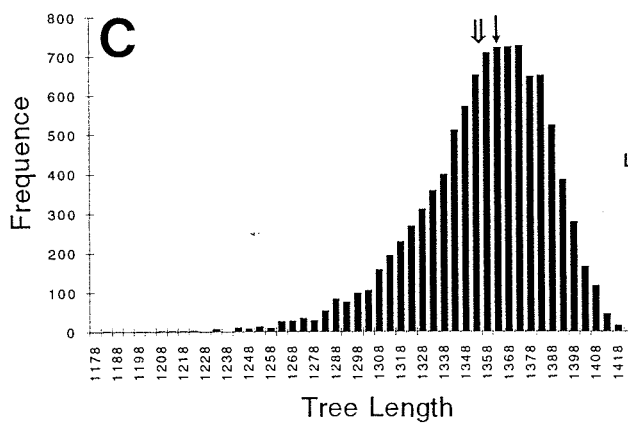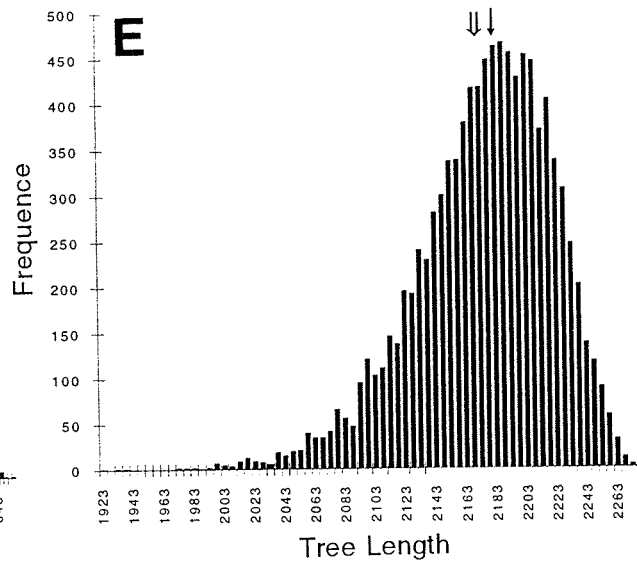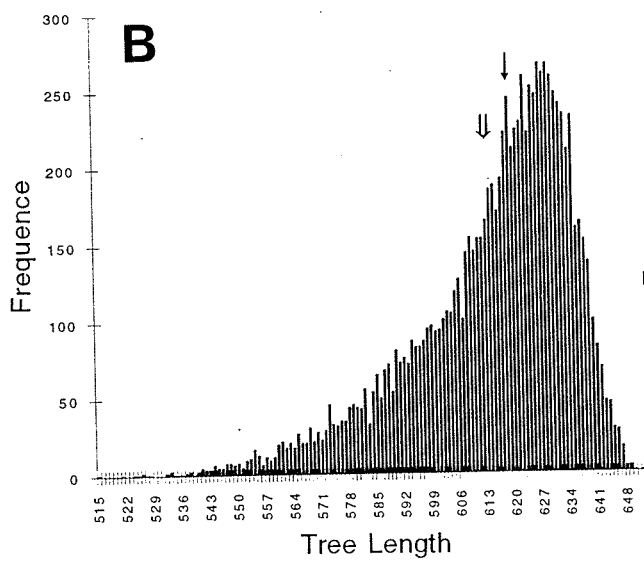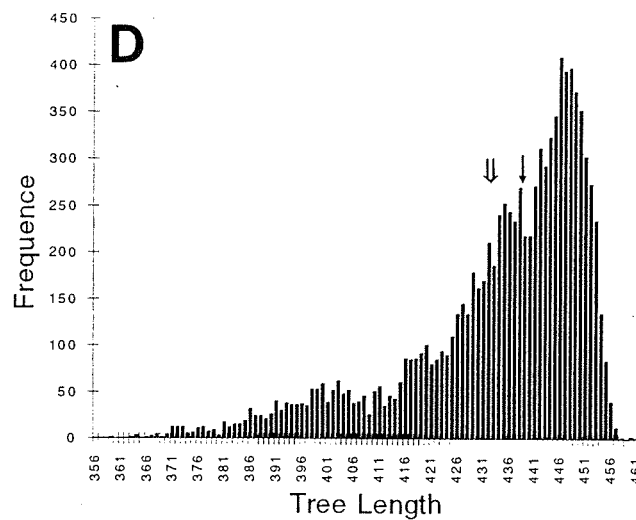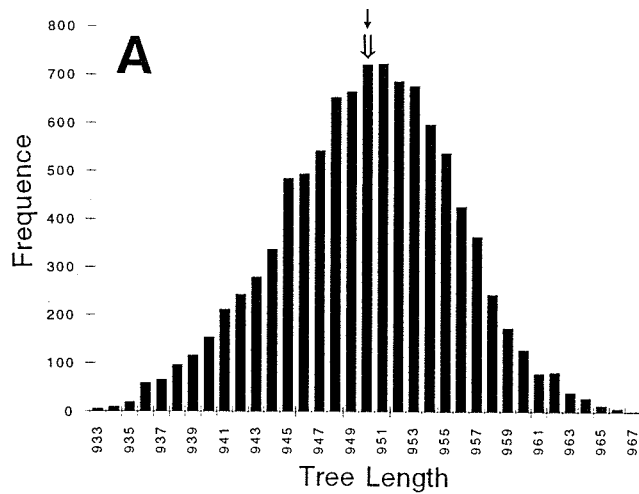
5.3.2 Phylogeny of the tomato Chlorophyll a + b-binding protein family:

Most of the genes encoding Chl *a* + b-binding proteins from tomato have been cloned and sequenced. This provides an opportunity to examine the intraspecies relationships between the different *Cab* gene members of this multigene family (Table 1.2). The phylogenetic trees in Figure 5.3 are based on 147 amino acid residues, representing approximately 67% of the mature protein. The amino acids used in the analysis are indicated in Table 5.1. These characters make up portions of the mature protein that are part of the membrane spanning regions (MSR 1-3) and include the stroma exposed areas in front of them.

The distance (Fig. 5.3, A) and the parsimony (Fig. 5.3, B) methods were fairly consistent in their predicted relationships between the different tomato *Cab* genes. The clustering of the LHC II sequences on a branch with Lhcb5 (CP26) is robust and consistent between analyses. This relationship is also supported by high bootstrap values. Within this branch, Lhcb5 and Lhcb3 form the deepest nodes with Lhcb2 and Lhcb1 consistently clustering together (Fig. 5.3, A & B).

The Lhcb4 gene from *Arabidopsis* was used in the alignment because the corresponding gene from tomato has not been cloned. It was expected that the relationship of tomato Lhcb4 (CP29) would not deviate significantly from what is observed with the *Arabidopsis* gene, as will become obvious in the next section. In both trees, the Lhcb4 gene has a closer evolutionary relationship to the Lhcb6a/b (CP24) gene than it does to the other Lhcb sequences.

The Lhca (LHC I) sequences are quite divergent and their depicted relationships are not well supported by bootstrap analysis. The Lhca sequences typically have a 45% identity to Lhcb1 and are 50-60% identical to each other. There is moderate support for the closer association of Lhca2, 3 and 4 in both trees. This is to the exclusion of Lhca1, which frequently forms a monophyletic group with Lhcb4 and Lhcb6 in the distance tree (Fig. 5.3A). The clustering of Lhcb6 and Lhcb4 is considered robust because the branch is well supported by bootstrap replicates (89 and 83%) and is consistent between methods.
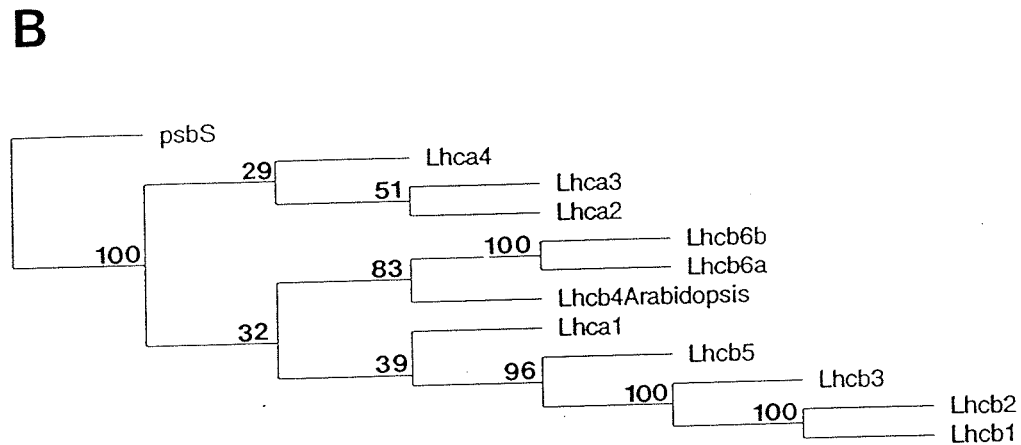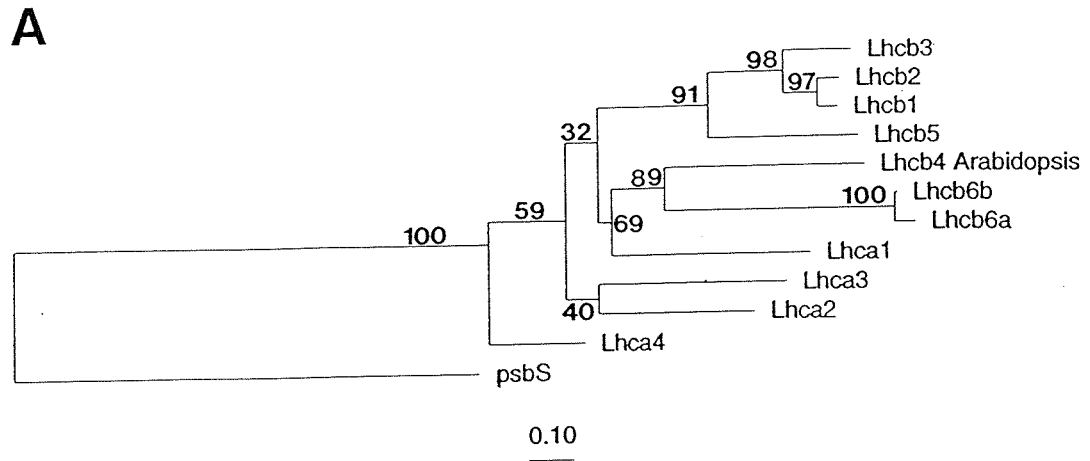
A

B

Figure 5.3

    Phylogenetic analyses of the CAB proteins from tomato (*Lycopersicon esculentum*). (A) Distance matrix analysis using the neighbor-joining tree construction method. (B) Parsimony analysis using an identical dataset as in (A) and as described in the results section. As labeled, the Lhcb4 gene was from *Arabidopsis thaliana*. Gene names listed in the Figure correspond to the following protein complexes: Lhcb1→3 (LHC II types I→III), Lhca1→4 (LHC I types I→IV), Lhcb4 (CP29), Lhcb5 (CP26), Lhcb6a/b (CP24), psbS (the 22 kDa protein). Further explanations and characterizations of the complexes are given in Table 1.2. Bootstrap values calculated from 100 replicates are given at the appropriate nodes.

144

Changing the outgroup from PsbS (CP22) to Lhcb6 (CP24) did not significantly alter the relationships presented in Figure 5.3. The relationship of the Lhcb4/6 branch, with respect to the main Lhca branch (Lhca2-4), could be altered depending on the alignment used and number of characters selected. However, no significant change in tree topology was observed when the amino acid changes were weighted according to their chemical similarity. Little change was also observed when a tree was constructed with the same distance data but using a different algorithm (eg. Fitch). Despite this, rates of divergence are relatively large in the Lhca portion of the tree as indicated by the long branch lengths in Figure 5.3A. Bootstrap values are also low, suggesting the characters used are inconsistent in their phylogeny prediction.

5.3.3 Phylogeny of the Chlorophyll a + b-binding protein family:

This analysis shows the relationships between the diverse CAB proteins from different taxa. In this case, I intentionally included LHC I and LHC II sequences to assess their relatedness to one another and to make inferences about the evolution of this gene family. The first analysis used 146 amino acid positions (characters), defined in Figure 5.1. Figure 5.4B is a consensus of four most parsimonious trees with a length of 1469.

There is an apparent separation of the LHC II proteins (Lhcb1, 2, 3 and 5) and the LHC I proteins (Lhca1-4) in the distance (Fig. 5.4A) and the parsimony trees (Fig. 5.4B). The main exception to this is the presence of Lhcb4 and Lhcb6 within the LHC I branch, which was the case in the previous analysis (Fig. 5.3). The *Euglena* Lhca sequences are also outside this main division. However, the separation of LHC I and LHC II is not well supported by the bootstrap values in either tree. Other branch variation, indicated by low bootstrap values, was within the established branches of the tree and did not alter the separation of the Lhca and Lhcb lineages (ie. within the Lhcb1-2 branches).
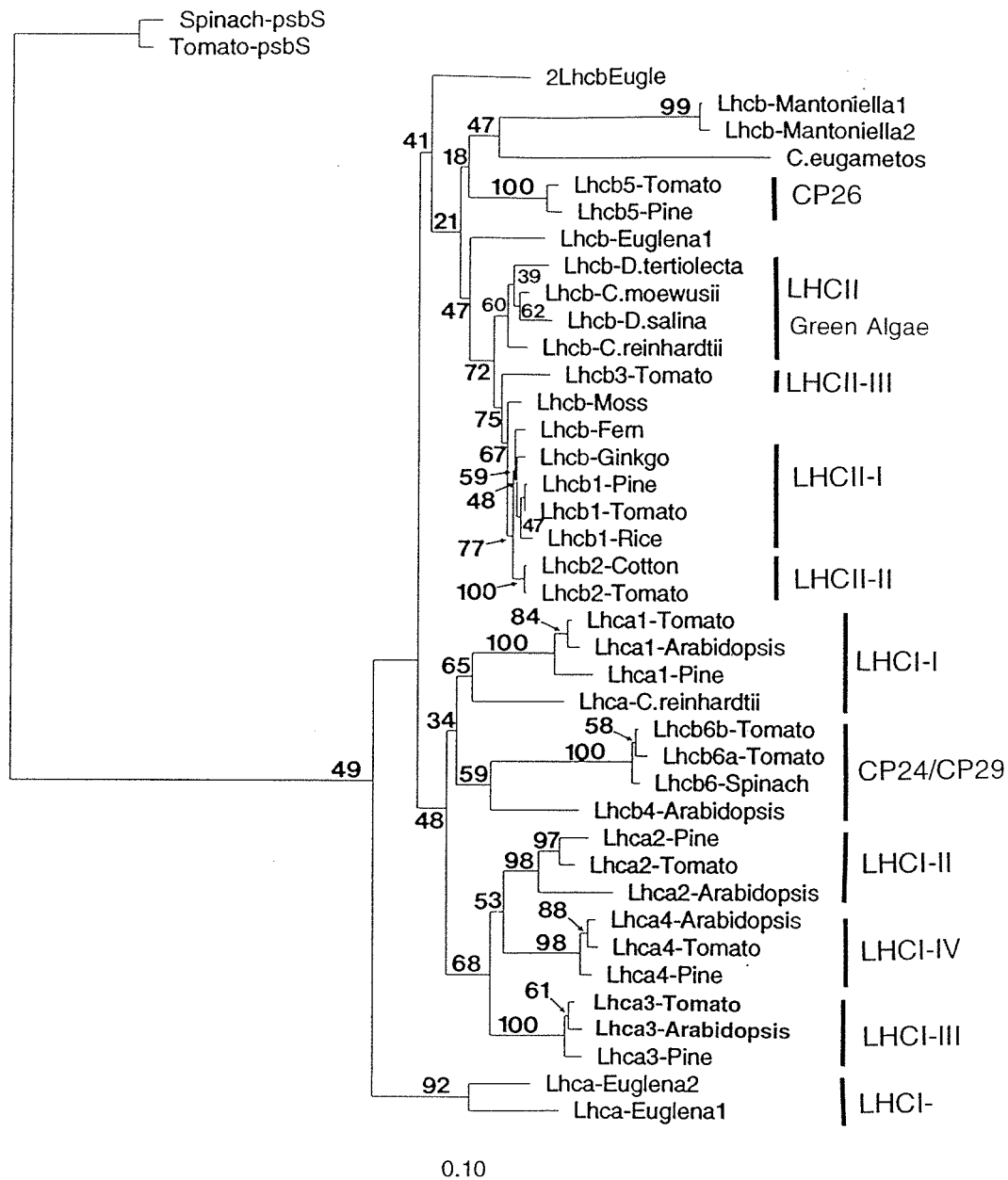
Figure 5.4A

Distance matrix analysis of the LHC proteins from select Chlorophyll $a + b$ containing taxa, including both PS I associated (Lhca) and PS II associated (Lhcb) antennal complexes. The tree was constructed from the distance matrix using the neighbor-joining method described in the results section. Sequences are labeled according to their gene names and the identification of the corresponding complexes are indicated in the figure legend of Figure 5.3 and in table 1.2. Refer to Table 5.2 for references and full species names. Bootstrap values calculated from 100 replicates are given at the appropriate nodes.
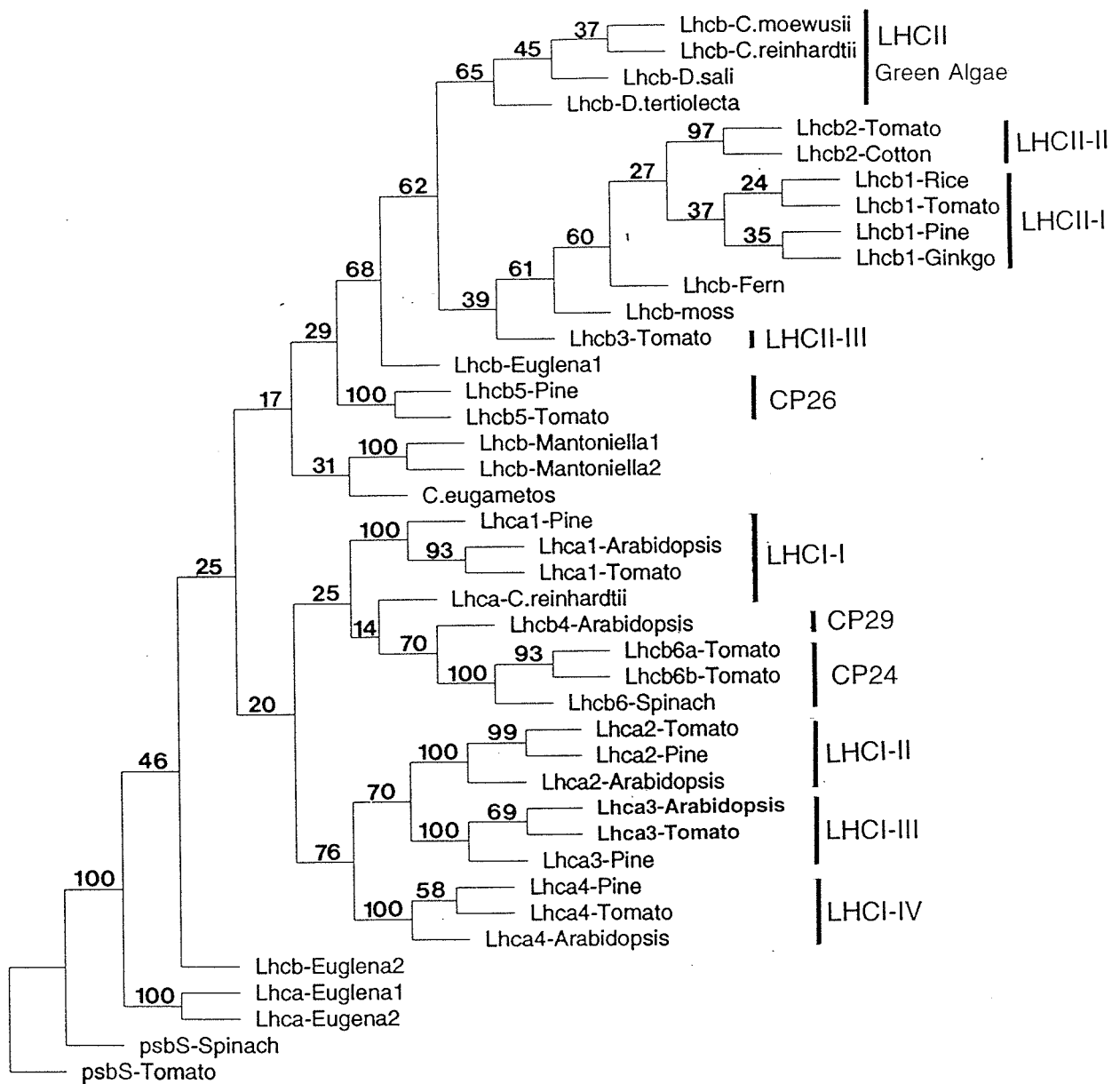
Figure 5.4B

Parsimony analysis of the LHC proteins from select Chlorophyll *a* + *b* containing taxa, including both PS I associated (Lhca) and PS II associated (Lhcb) antennal complexes. Analysis was done using an identical dataset as in (A), as described in the results section. Refer to Table 5.2 for references and full species names. Bootstrap values calculated from 100 replicates are given at the appropriate nodes.

The distance and parsimony trees were consistent in the clustering of the same CAB types (e.g. Lhca4-pine with Lhca4-tomato), and this was true for all known angiosperm and gymnosperm taxa. This suggests that the major LHC *Cab* gene lineages diverged before the angiosperm / gymnosperm separation. The association of the chlorophyte Lhcb sequences, however, does not follow this pattern. In both trees, they form a monophyletic branch separate from the Lhcb1, Lhcb2, and Lhcb3 sequences of the terrestrial plants. The separation of the Lhcb1/2 sequences from the green algal Lhcb sequences is supported (bootstrap value 75 or 62, Fig. 5.4), though the separation of Lhcb3 is not as well supported in the parsimony tree (Fig. 5.4B). This suggests that Lhcb1-3 diverged after the chlorophyte and terrestrial plant lineages had separated.

The fern Lhcb sequence clusters with the Lhcb1 clade in the distance tree but before the Lhcb1/2 separation in the parsimony tree. The bootstrap values coinciding with these branches are also low. Because of this, the relationship is considered unstable and conclusions regarding the divergence of the fern sequence, before or after the Lhcb1/2 functional separation, are inconclusive (Fig. 5.4 A/B). Divergence of the moss Lhcb sequence before the Lhcb1/2 divergence is moderately supported by the bootstrap analysis (Fig. 3, A/B), though this is under the assumption that the moss sequence represents the main LHC II antennal protein and is not a paralogous gene. Some caution must be exercised in the analysis of the Lhcb relationships between the fern, moss, and chlorophyte CABs as a limited number of sequences have been characterized; these are often assumed to be homologous to the type 1 LHC II (Lhcb1) sequence of the angiosperms but they could just as well be paralogous sequences.

Both trees show that the Lhcb3 sequence has separated at an earlier point than the Lhcb1 and 2 gene types. Although this is a consistent relationship, it is not well supported (bootstrap value 67 or 61, Fig. 5.4A/B). This is suggestive of a divergence of this CAB type around, or just after, the chlorophyte/terrestrial plant separation. Lhcb5 diverged early in the *Cab* gene evolution, before the separation of the chlorophytes. The precise position of Lhcb5, however, is not well supported by bootstrap analysis. The close association of Lhca types 2, 3 and 4 are

supported in both trees (Fig. 5.4). The clustering of Lhca1, Lhcb4, and Lhcb6 is consistent between methods but is not substantiated with the bootstrap trials. The relative relationships of the *Euglena* (Euglenophyte) and *Mantoniella* (Prasinophyceae) LHCs were not well resolved on these trees.

A second analysis was done to resolve the relationships amongst the green algae and selected terrestrial plant sequences (Fig. 5.5, A/B). This analysis included 20 taxa and a larger number of amino acids (170), making up approximately 75% of the mature protein (Fig. 5.2A). The outgroup was the *Chlamydomonas eugametos* light responsive gene that has similarity to the CAB proteins (Gagné and Guertin 1992). This was used as the outgroup because it is one of the most divergent green algal sequences. Figure 5.5 A and A' are both unrooted trees and are identical except that A' is displayed in the radial format.

The distance tree (Fig. 5.5, A) shows a topology that strongly supports the association of the green algal LHC II sequences with three out of four *Euglena* LHC II sequences. However, there is a fair degree of divergence between these sequences, as indicated by the length of the branches. The Euglena2 sequence is more distantly related than the others but it is still associated with the LHC II branch. The *Euglena* sequences were isolated by heterologous hybridization with a plant LHC II probe (Muchhal and Schwartzbach 1992). Their identification as LHC II sequences was by analogy to the terrestrial plant system. The LHC I sequences from the green algae and euglenoids tend to form a monophyletic assemblage separate from the LHC II group but the branches are long and not strongly supported by the bootstrap analysis. The *Euglena*-Lhca sequences were cloned by immunoscreening a cDNA library with a LHC II polyclonal antibody. However, sequence comparisons and immunoprecipitation with a barley LHC I monoclonal antibody provided some evidence that they are LHC I sequences (Houlné and Schantz 1988). The identification of the *C. reinhardtii* Lhca sequence as a PS I associated antenna was confirmed by N-terminal sequencing of a PS I associated polypeptide (p22) (Bassi et al. 1992; Hwang and Herrin 1993). The characterization of the Lhca gene from *C. stellata* has not been published and little information on its function or organization is known. However, it

149

Figure 5.5

Phylogenetic analysis of the LHC proteins from select green algae and terrestrial plant CAB proteins, including both PS I associated (Lhca) and PS II associated (Lhcb) antennal complexes. The tomato sequence is a Lhcb1 and the others are Lhcb types unless otherwise indicated. (A) Distance matrix analysis using the neighbor-joining tree construction method. The inset (A') is identical to (A) except it is displayed in a radial format to facilitate interpretation. (B) Parsimony analysis using an identical dataset as in (A), as described in the results section. Refer to Table 5.2 for references and full species names. Bootstrap values calculated from 100 replicates are given at the appropriate nodes.
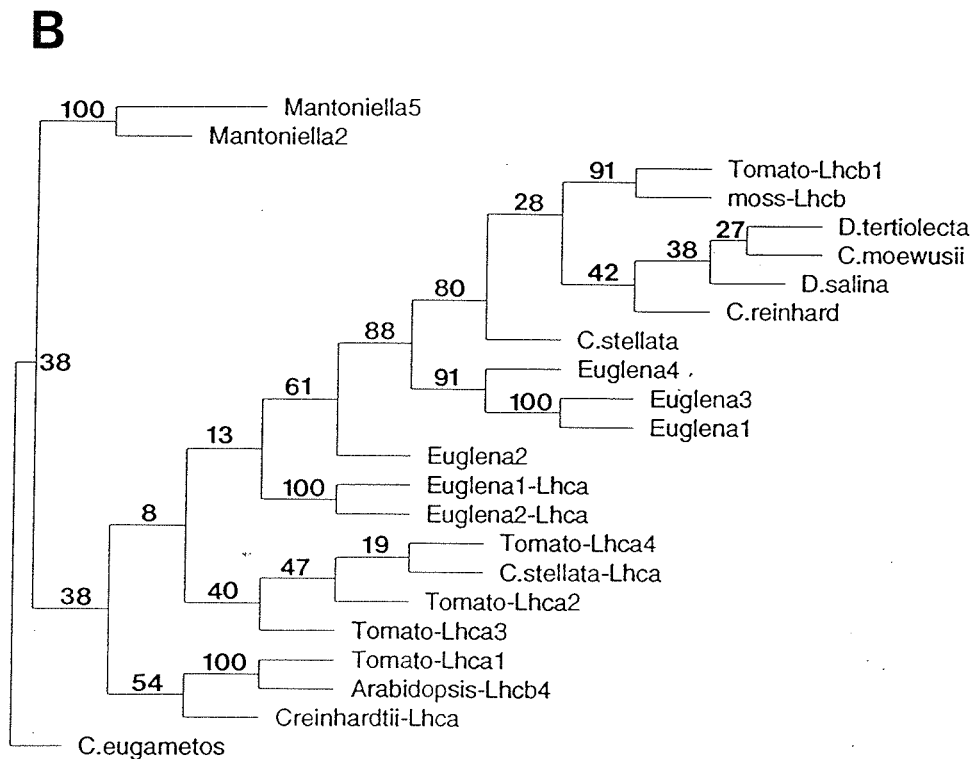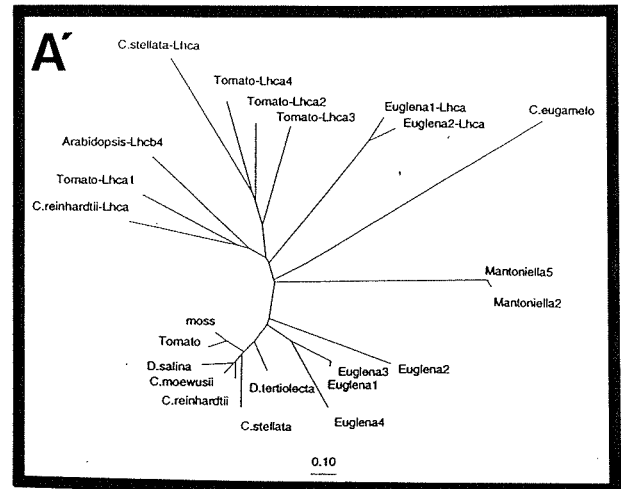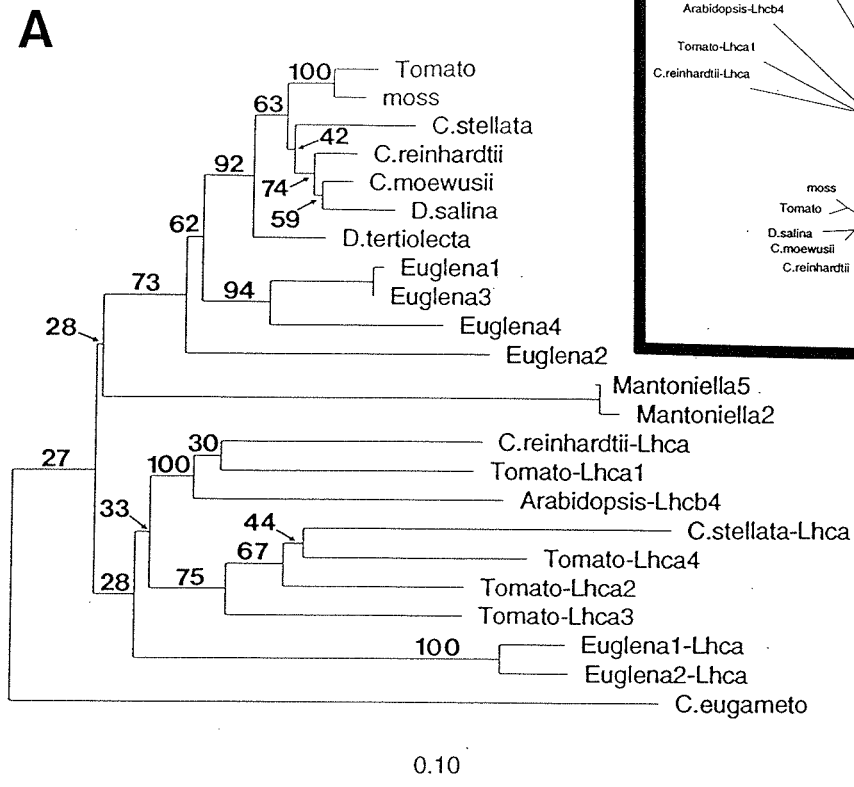
does share a degree of similarity to the N-terminal sequence of a *C. reinhardtii* LHC I protein, p15.1 (Bassi et al. 1992).

The *Mantoniella* Lhcb sequences form a branch at the base of the LHC II cluster in the distance tree, as they did in Figure 5.4. However, a large divergence between the *Mantoniella* LHCs and the Lhca and Lhcb sequences are indicated by the length of the branches separating them. This is clearly evident in the radial tree (Fig. 5.5 A'). The position of the *Mantoniella* sequence is not supported in a bootstrap trial (28%) and this low value may reflect a fair amount of noise within the sequence. This noise may be the result of non-divergent mutations (parallel, non-divergent changes) occurring at a significant rate.

Parsimony analysis of this dataset gave two trees with a length of 1356. The differences between the two were within the green algal LHC II branch and no significant deviations from the consensus tree (Fig. 5.5 B) were observed . The overall topology of the parsimony tree concurs with the distance tree. The main difference between the two methods is in the placement of the *Mantoniella* LHC branch. With parsimony, the *Mantoniella* sequence separates before the algal LHC II and LHC I divergence. I am skeptical about this topology because of the long branches connecting the *Mantoniella* and *C. eugametos* sequences (Fig. 5.5 A'). The presence of these very divergent sequences can result in treeing artifacts with a parsimony analysis (Felsenstein 1978), which is probably the case here. Though the distance tree is probably a more accurate reflection of the true gene phylogeny, I am unable to conclude whether the *Mantoniella* LHC is more closely related to either LHC I or LHC II type sequences.

In both trees (Fig 5.5), the two *Chlamydomonas* Lhca sequences cluster with a specific terrestrial plant Lhca sequence and not with each other (Fig. 5.4 and 5.5). The bootstrap values are low for these branches so the exact association of the two sequences can not be concluded. Nevertheless, this offers preliminary evidence that the divergence of the LHC I sequences occurred early in the evolutionary history; before the green algal/terrestrial plant separation.

5.3.4 Phylogeny of the chlorophyll a + c-binding proteins

This analysis examines the relationships of the chlorophyll $a + c$ binding proteins (FCPs, *i*PCPs) from several groups of non-green algae. The dataset consists of 9 sequences (from six taxa) and includes 153 amino acids ($\approx 90\%$ of the mature protein) in the analysis, indicated in Table 5.1. The protein sequences are from two diatoms (*Phaeodactylum* and *Odontella*), a brown alga (*Macrocystis*), a raphidophyte (*Heterosigma*) and a haptophyte (*Isochrysis*). All of these algae contain fucoxanthin, Chl *a* and Chl *c*. The other protein sequence is from the dinoflagellate, *Amphidinium*, which contains the predominant carotenoid peridinin, instead of fucoxanthin. The *Amphidinium* *i*PCP sequence was determined directly by amino acid sequencing, while the FCP protein sequences were inferred from the gene sequence.

Both the distance (Fig. 5.6 A) and parsimony (Fig. 5.6 B) trees showed a clustering of the FCP sequences separate from the *i*PCP (*intrinisic* peridinin-chlorophyll protein) of the dinoflagellate, *Amphidinium*. This is with the exception of the *Isochrysis* sequence which was distant from the other sequences, even though it has a fucoxanthin binding LHC (FCP). The distance analysis did not resolve the relationships amongst the FCP sequences of the brown alga, diatoms or the raphidophyte.

The parsimony analysis (Fig. 5.6 B) gave only one most parsimonious tree with a length of 490. The pennate (*Phaeodactylum*) and centric (*Odontella*) diatoms form a monophyletic group in agreement with traditional morphological characters (Round and Crawford 1990), though the length of the branches separating them (Fig. 5.6 A) indicate an early divergence. The *Heterosigma* sequence is also grouped on this branch, which is separated from the brown algal clade. These relationships are supported by high bootstrap values and the branch is considered robust. The *Isochrysis* sequence is very divergent and is at the base of the tree. Interestingly, the *Heterosigma* sequence groups with the brown algae when the *Amphidinium* sequence is removed from both the parsimony and distance analyses (not shown).

153

A

0.10

B

C
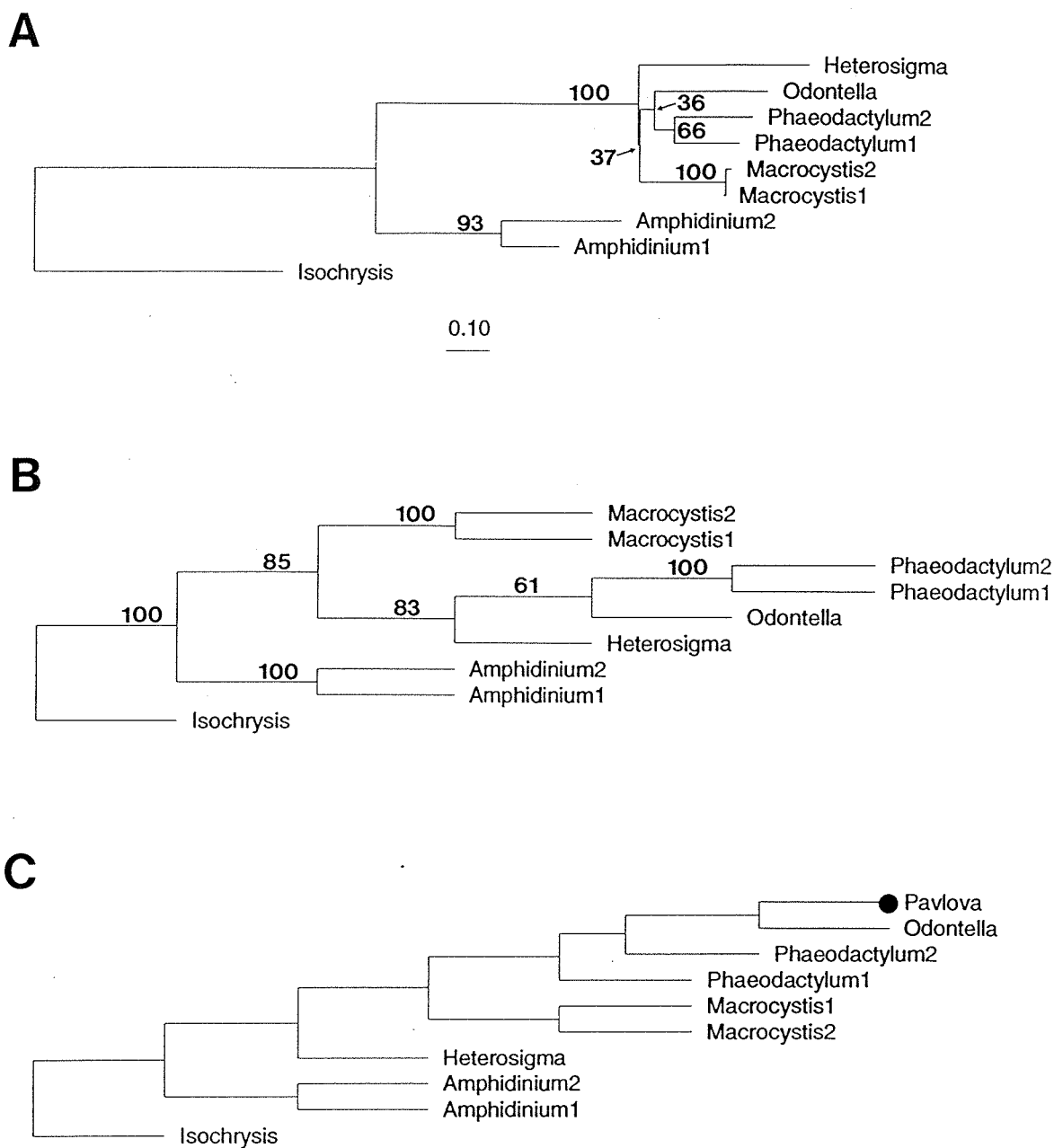
Figure 5.6

Phylogenetic analysis of the FCP proteins from various Chl $a + c$ containing organisms. (A) Distance matrix analysis using the neighbor-joining tree construction method. (B) Parsimony analysis using an identical dataset as in (A) and described in the results section. (C) Parsimony analysis using a limited number of characters (87) in order to incorporate the partial sequence information from *Pavlova lutherii* (●). Refer to Table 5.2 for references and full species names. Bootstrap values calculated from 100 replicates are given at the appropriate nodes.

The very distant placement of the *Isochrysis* sequence was odd in light of the partial

sequence information from *Pavlova*, another haptophyte (Roger Hiller, unpubl.). I ran a

parsimony tree with the partial *Pavlova* sequence included in the alignment (Fig. 5.6 C ●).

Because of its inclusion, only 87 amino acids (≈49% of mature protein) were used in the analysis

(Fig. 5.1). There are some obvious differences in topology between Figure 5.6 B and 5.6 C. The

main point is the close affinity of the *Pavlova* partial sequence to the main FCP cluster, which is

separate from the *Isochrysis* sequence. Though these are preliminary data, it seems unusual that

two taxonomically related species show such a remarkable difference. The most obvious

explanation is that the *Isochrysis* sequence is paralogous and not the predominant FCP in the

organism.
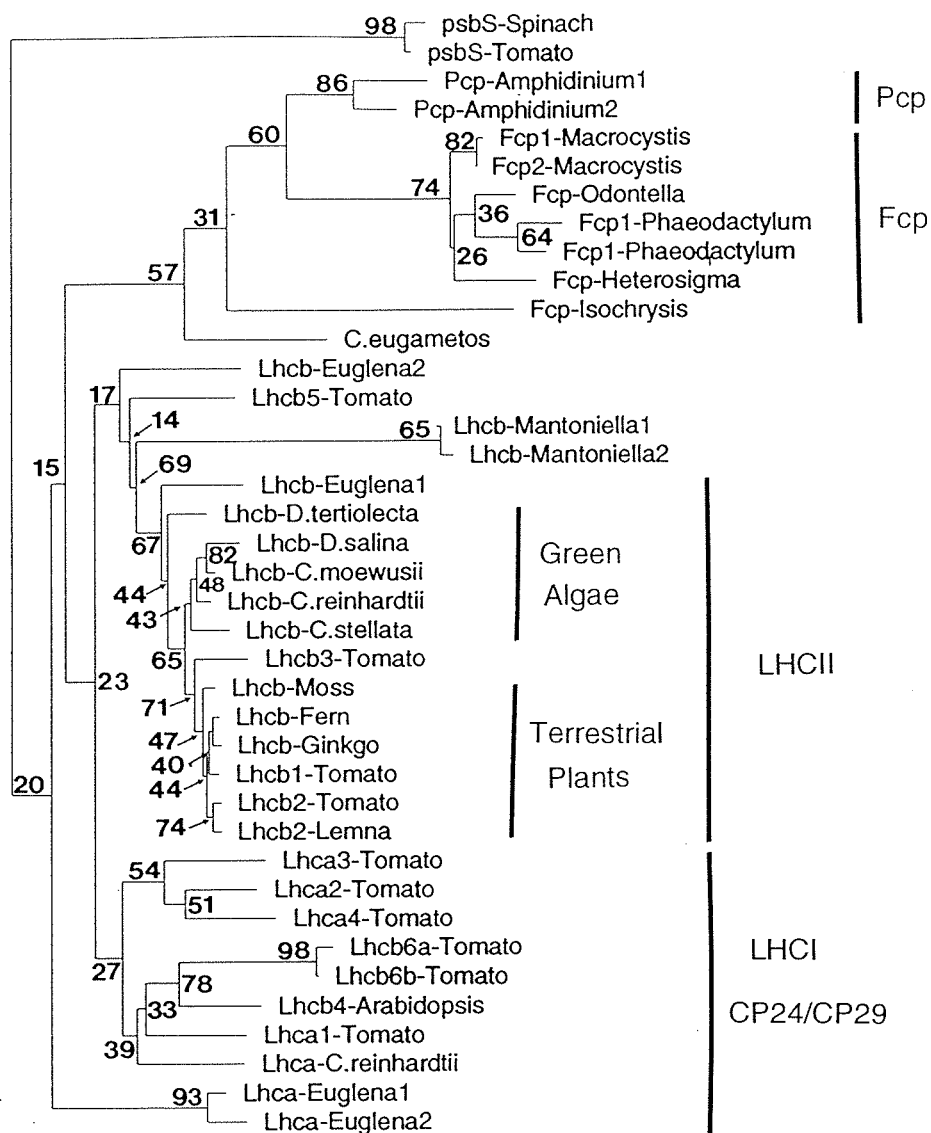

5.3.5 Evolution of the LHC family of proteins

This analysis looks at the evolutionary relationships between the CABs, FCPs and the

*i*PCPs; the main intrinisic light-harvesting antennae. A dataset of 39 taxa and 135 amino acids

(≈61-79% of mature polypeptides) was constructed. The exact characters used in the analysis are

given in Table 5.1. The taxa included were from the main LHC proteins of the non-green algae

(FCP/*i*PCPs), the green algae and the terrestrial plants (CABs). In addition, representatives from

other LHC types, such as LHC I (Lhca1-4), CP24 (Lhcb6), CP26 (Lhcb5) and CP29 (Lhcb4),

were included.

With this dataset, the interpretation can vary depending on whether parsimony or distance

methods are used. In the distance tree there are three major groups (Fig. 5.7 A, A'): (1) the

chlorophyll *a* + *c* proteins (FCP, *i*PCP) that group with the *C. eugametos* (*CAB*) sequence, (2) the

LHC II (Lhcb) sequences, and (3) the LHC I (Lhca) sequences that group with Lhcb6 (CP24)

and Lhcb4 (CP29). The *Euglena* LHC I (Lhca-Euglena1/2) sequences form the earliest branch

in the tree. In this tree it appears that Chl *a* + *b* sequences from LHC I and II are more closely

related to each other than either is to the Chl *a* + *c* lineage. This is clearly evident in the radial

155

Figure 5.7A

　　Distance matrix analysis of all LHC proteins from select Chl a + b-containing and Chl a + c-containing taxa, including both PS I associated (*Lhca*) and PS II associated (*Lhcb*) antennal complexes. The tree was constructed from the distance matrix using the neighbor-joining method as described in the results. Inset (A') is a radial display of tree (A) and is otherwise identical. Refer to Table 5.2 for full species names and references. Bootstrap values calculated from 100 replicates are given at the appropriate nodes.

A

psbS-Spinach
psbS-Tomato
Pcp-Amphidinium1
Pcp-Amphidinium2 } Pcp
Fcp1-Macrocystis
Fcp2-Macrocystis
Fcp-Odontella
Fcp1-Phaeodactylum
Fcp1-Phaeodactylum } Fcp
Fcp-Heterosigma
Fcp-Isochrysis
C.eugametos
Lhcb-Euglena2
Lhcb5-Tomato
Lhcb-Mantoniella1
Lhcb-Mantoniella2
Lhcb-Euglena1
Lhcb-D.tertiolecta
Lhcb-D.salina
Lhcb-C.moewusii
Lhcb-C.reinhardtii
Lhcb-C.stellata } Green Algae
Lhcb3-Tomato
Lhcb-Moss
Lhcb-Fern
Lhcb-Ginkgo
Lhcb1-Tomato
Lhcb2-Tomato
Lhcb2-Lemna } Terrestrial Plants
LHCII
Lhca3-Tomato
Lhca2-Tomato
Lhca4-Tomato
Lhcb6a-Tomato
Lhcb6b-Tomato
Lhcb4-Arabidopsis
Lhca1-Tomato
Lhca-C.reinhardtii
Lhca-Euglena1
Lhca-Euglena2
LHCI
CP24/CP29

0.10

A'

Isochrysis
Phaeodactylum   Heterosigma
Odontella
Macrocystis
1Amphidini
3Amphidini
C.eugametos
Mantoniella
Lhcb-Euglena2
Lhcb5
Lhcb-Euglena1
psbS-Spinach
psbS-Tomato
Lhca-Euglena1
D.tert.
D.salina
C.reinhard
Lhcb3
Lhca-Euglena2
Lhcb1
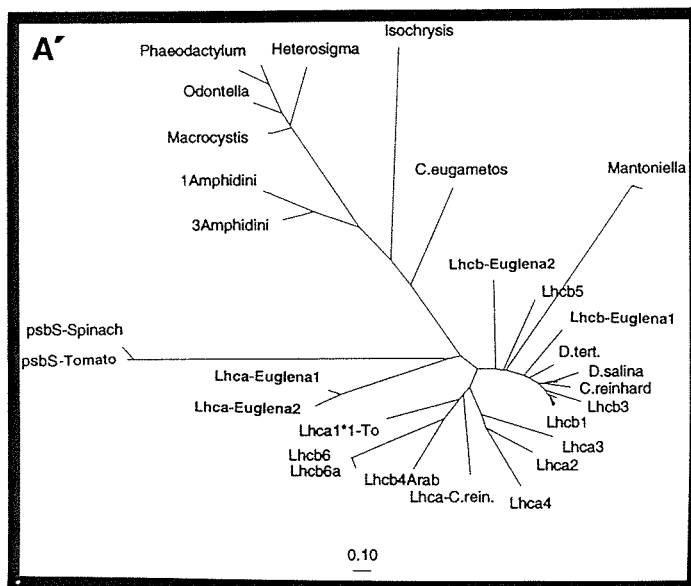Lhcb6
Lhcb6a
Lhca3
Lhcb4Arab
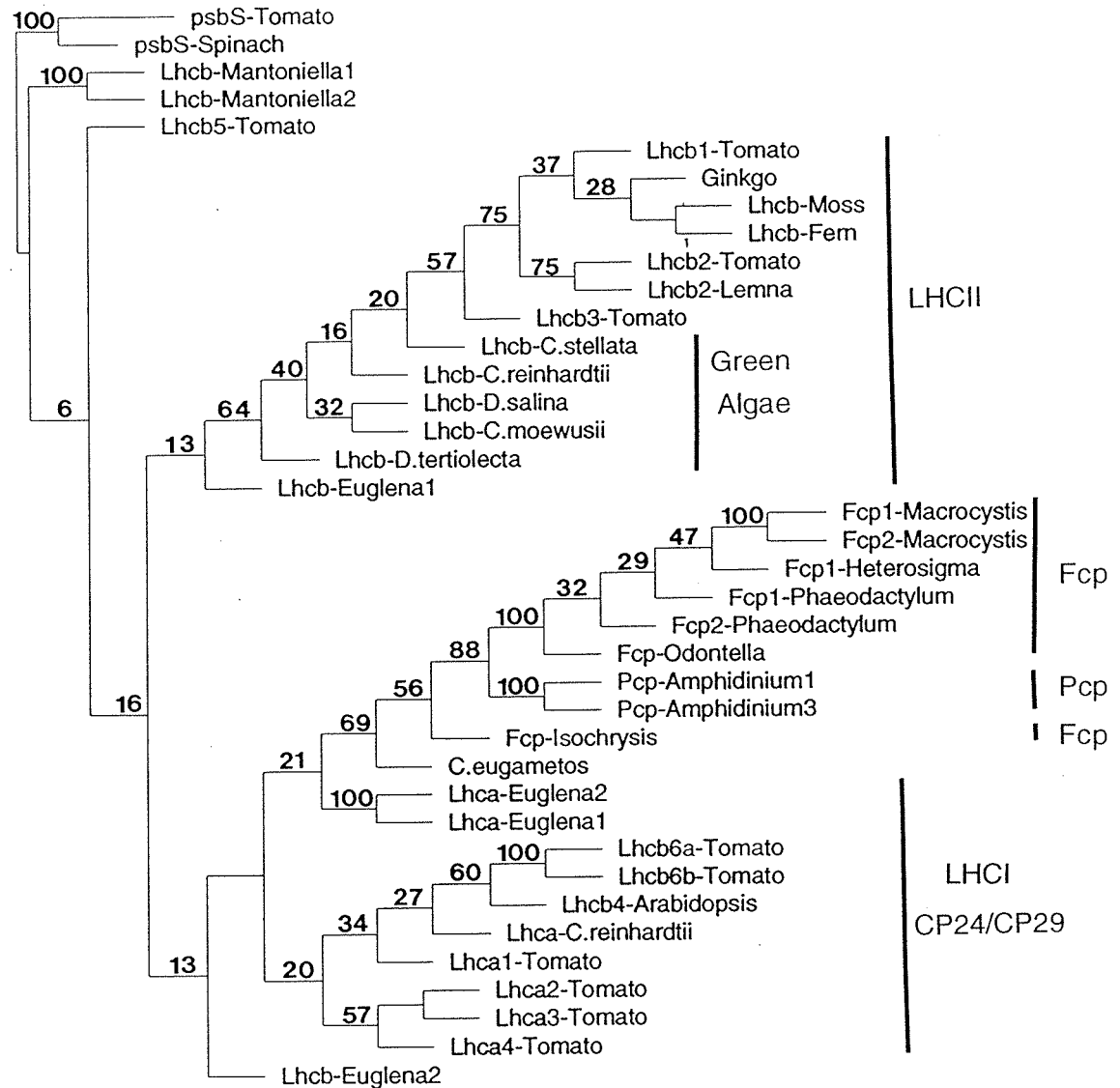Lhca2
Lhca-C.rein.   Lhca4
0.10

157

Figure 5.7B

Parsimony analysis of all LHC proteins from select Chl *a* + *b*-containing and Chl *a* + *c*-containing taxa including both PS I associated (*Lhca*) and PS II associated (*Lhcb*) antennal complexes. Analysis was done using an identical dataset as in (A), as described in the results section. The tomato PsbS protein is used as the outgroup. Refer to Table 5.2 for references and full species names. Bootstrap values calculated from 100 replicates are given at the appropriate nodes.

tree (Fig. 5.7 A'). Because of the low bootstrap values on the main branches leading to these branches, however, interpretations have to be made with caution since the resolution in this region of the tree is poor.

With parsimony, six equally parsimonious trees were found with a length of 1792. The consensus of these six trees is shown in Figure 5.7 B. Most of the variations were within the internal nodes of the main branches and were considered insignificant. This analysis produced a tree topology where the Chl $a + c$ sequences and the LHC I (Lhca) lineage formed a monophyletic group; this relationship was not supported by bootstrap replicates. As the branch lengths are very different between the FCPs and the CABs, I am skeptical of this association and suspect it may be the result of treeing artifacts that are common to parsimony; particularly in cases where there are unequal rates between the lineages. The low bootstrap values separating the two lineages (16%) makes any firm conclusions regarding the relationships of the CABs and FCPs unsupported. Relationships within the major lineages are generally the same as with distance analysis.

5.4 Discussion

5.4.1 CAB protein evolution

The trees suggest that the two main LHC II types (Lhcb1-2) form a monophyletic clade in agreement with the analysis of Jansson (1994). The Lhcb1 and Lhcb2 polypetides make up the main LHC II complex and are generally 90% identical in the mature polypeptide (excluding the N-terminal region) so their close association on the tree is not surprising. Together, Lhcb1 and Lhcb2 make up the main peripheral antennae of PS II. Its principal function is the capture of light followed by the transfer of the excitation energy to the core reaction center of PS II. It has also been implicated in the mediation of thylakoid appression, the regulation of energy distribution between the photosystems and in photoprotection, all of which have been previously reviewed (Anderson and Andersson 1988; Bassi et al. 1990; Jansson 1994). LHC II is thought to exist as a trimeric unit (Kühlbrandt and Wang 1991) composed of Lhcb1 and Lhcb2 polypeptides at a ratio of approximately 2:1 (Jansson 1994). Both Lhcb1 and Lhcb2 have phosphorylatable threonine residues in the amino terminus of the mature polypeptide that are thought to be responsible for the state transition observed in thylakoids (Mullet 1983). The close evolutionary relatedness of these two complexes is reflected in their functional similarities. However, the Lhcb2 complex is enriched in a peripheral subpopulation of LHC II, has different phosphorylation kinetics and appears later in development (Larsson et al. 1987). These differences are thought to be important in the adaptation to different light intensities. It is likely that the appearance of the Lhcb1 / Lhcb2 lineages was due to a fine tuning of the light adaptation response.

The more distantly related Lhcb3 (LHC II type III) sequence is a minor antennal component that is about 80% identical to the Lhcb1 and Lhcb2 complexes. This protein has a shorter N-terminus and is often found in the LHC II complex (Green et al. 1992b). Lhcb3 may function as a linker between the bulk trimeric LHC II complex and the PS II core. This is

160

suggested because of its close association with the core complex of PS II in barley Chl *b*-less

mutants (Harrison and Melis 1992). The more proximal location of Lhcb3 to PS II, with respect

to Lhcb1/2, would accommodate the hypothesis of an earlier evolutionary history of this protein,

as depicted in the tomato CAB tree (Fig. 5.3). The addition of the peripherally located Lhcb1 and

Lhcb2 complexes could occur without much change in the preexisting inner antenna

organization.

The Lhcb5 (CP26) sequence is consistently found at the base of the Lhcb clade,

suggesting it has evolved prior to the peripheral LHC II complex. CP26 remains associated with

the PS II core when the main peripheral LHC II antennae are removed (Camm and Green 1989)

and is considered an inner antenna. It is likely that the peripheral antennal proteins of LHC II

arose from the divergence of a gene encoding a inner antennal, CP26-like, protein. Duplication

and divergence of the *Lhcb5* gene would have to be accompanied by a change in size and

chlorophyll binding capabilities. The Chl *a/b* ratio of LHC II is 1.3 compared to approximately

3.3 for CP26 (van Amerongen et al. 1994; Table 1.2). This suggests that changes in the amounts

or relative proportion of Chl *a* and *b* bound by each polypeptide has changed in the course of

evolution (Green and Kühlbrandt 1995). Evolution of Lhcb1/2 at a later point would also suggest

an increased level of regulation by phosphorylation as CP26 is not reversibly phosphorylated.

An interesting finding is the close association of Lhcb6 (CP24) and Lhcb4 (CP29) in the

trees, separate from Lhcb5 (CP26). This is somewhat surprising as the CP29 and CP26 pigment-

protein complexes often copurify (Green 1988) and both are associated with the core complex of

PS II (Camm and Green 1989). Their presence in both grana (α) and stromal (β) localized PS II

centers (Allen and Staehelin 1992) suggests that they are part of the basic PS II unit. Because of

the biochemical similarities, the complexes are also known as CP29 type I (CP26) and CP29

type II (Cp29) (Pichersky et al. 1991). The distant relationship between Lhcb4 and Lhcb5 has

been suggested through direct sequence comparison (Morishige and Thornber 1992; Green and

Pichersky 1993) and by phylogeny construction (Jansson 1994). Lhcb6 and Lhcb4 consistently

group together, though the long branches joining them (Fig. 5.3 A) suggest they diverged from

one another long ago. The minor CP24 complex binds little chlorophyll and, like CP29, is enriched in PS II fractions (Dunahay and Staehelin 1986).

The Lhca1 sequence (LHC I type I) is closely related to the PS II inner core antennae, CP29 and CP24. Although consistent between methods, this relationship is only moderately supported by bootstrap replicates in the distance trees (Fig. 5.3/4 A). A clustering of the CP29 (Lhcb4) and LHC I (Lhca1) sequences was also supported in a recent dendrogram (Jansson 1994). PS I antennae, Lhca2-4, appear to be more related to one another than either is to Lhca1 or Lhcb6/4. However, all the Lhca sequences are quite divergent from one another as indicated by the long branch lengths separating them on the distance trees.

The relationship amongst the different LHC I antennae are unstable and not strongly supported by bootstrap replicates in most cases. Parsimony analysis consistently places Lhca2 and Lhca3 as the closest relatives within the LHC I lineage; this is supported by bootstrap replicates in Figure 5.3B. Distance methods, however, consistently place Lhca2 and Lhca4 as the closest relatives, although this is not strongly supported by bootstrap trials in any of the trees shown. The latter relationship has been observed in another dendrogram that is based on a distance method (Jansson 1994). However, no indication of reliability or significance of this branching order was given. It is tempting to conclude a direct evolutionary relationship between Lhca2 and Lhca3 as they are the polypeptides making up the LHC I-680 complex. In tomato, the *Lhca2* and *Lhca3* genes are linked and have been mapped to chromosome 10, suggesting a closer affinity due to a more recent duplication event (Pichersky et al. 1989). On the other hand, the genes encoding Lhca1 and Lhca4, which make up the LHC I-730 pigment-protein complex, are on different chromosomes in tomato (Pichersky et al. 1987; Schwartz et al. 1991) and despite the apparent functional association, there does not seem to be any direct evolutionary relationship between the two complexes.

5.4.2 Evolutionary relationships amongst the Cab gene family

The analysis of the *Cab* gene family from a diverse array of organisms reinforces the idea that two particular CAB types (eg. Lhcb1) from different organisms are more similar to one another than either is to different CAB types (eg. Lhca4) from the same organism. Comparison of Lhcb1 and Lhcb2 sequences previously revealed that the same CAB type encodes a nearly identical polypeptide with similar numbers of introns (Chitnis and Thornber 1988). The distinction between Lhcb1 and Lhcb2 was also made by detecting signature amino acids (Jansson and Gustafsson 1990), which are conserved residues specific to a particular CAB type. In addition, a phylogenetic analysis of angiosperm Lhcb1 and Lhcb2 sequences (Demmin et al. 1989; Matsuoka 1990) also revealed this trend.

Sequencing of a fern (*Polystichum munitum*) (Pichersky et al. 1990) and moss Lhcb (*Physcomitrella patens*) (Long et al. 1989) provided useful markers in estimates of the main LHC II divergence. The analysis of the fern sequence was unable to clearly resolve the branching order. However, the moss Lhcb sequence is consistently at a branch before the fern and Lhcb2 separations, indicating that the functional separation of the Lhcb1 and Lhcb2 complexes occurred after the bryophyte lineage separated. This is a tentative conclusion until more representatives from bryophytes are sequenced, in order to rule out the possibility of misleading tree topology as a result of comparing paralogous genes. Overall, the duplication and separation of the Lhcb2 (LHC II type II) sequences probably occurred after the bryophyte lineage separated and at about the same time as the lineage leading to the pteridophytes (ferns) diverged. This is in agreement with earlier predictions by Pichersky et al. (1990). It is apparent that the divergence of all the major CAB types occurred early. This definitely occurred before the angiosperm/gymnosperm separation, as the same CAB types from pine consistently grouped with the same CAB types from the angiosperms. It is not possible to draw further conclusions as to the earliest separation of most the CABs because there are too few sequences known from the other taxonomically distinct groups such as the Chlorophyta and the Euglenophyta.

There is a distinct separation of the green algal LHC II sequences from the terrestrial plant Lhcb1-3 sequences suggesting that the minor Lhcb3 antenna appeared just after the green algal

lineage diverged. This may indicate a fundamental difference in the regulation (Sukenik et al. 1987) and organization of the peripheral LHC II antennae in green algae. The lack of Lhcb1 and Lhcb2 type antennal proteins in green algae suggests the Lhcb1/2 duplication and functional divergence occurred in the lineage leading to the terrestrial plants after the green algal lineages had separated.

The identification of an LHC I sequence from the green alga *Chlamydomonas reinhardtii* was confirmed by peptide sequencing of a PS I associated protein (Hwang and Herrin 1993). This analysis confirms its identity as a Lhca1 type, as this sequence tends to cluster with the angiosperm Lhca1 sequences; however, the relationship is not supported (bootstrap value 30 or 54, Fig. 5.5). In this case, the *Chlamydomonas* sequence is more closely related to tomato Lhca1 than to the Lhca sequence from *Chlamydomonas stellata*. Should this relationship hold as more LHC I sequences from green algae are determined, it would suggest that the LHC I sequences had diverged before the chlorophyte/land plant separation. This brings up the interesting suggestion that the different LHC I complexes evolved prior to the establishment of the different homologous LHC II complexes (Hwang and Herrin 1993), which will be discussed in the following sections.

The *Euglena* sequences for LHC I and LHC II are very divergent, as compared to their green algal and terrestrial plant counter parts, indicating that this alga separated from the green algal lineage very early. Of the four complete LHC II proteins, three are more related to one another (*Euglena* sequences 1, 3 and 4) and form a branch just before the green algal/ land plant Lhcb1/2 sequences. The respective LHC II and LHC I proteins from the green algae and *Euglena* group together, although the branch lengths in the LHC I lineage are very long on the distance trees. This suggests that LHC I and LHC II polypeptides had functionally separated prior to the appearance of the euglenoid chloroplast. It also shows a close direct link between the green algal and *Euglena* chloroplasts, previously hypothesized on the basis of the presence of chlorophyll *a* and *b*.

5.4.3   Evolution of the Chl a + b and Chl a + c gene families:

The relative position of the Chl a + c-binding proteins in relation to the Chl a + b-binding proteins of LHC I and LHC II is not clear from an examination of either the parsimony or distance trees. The distance trees (Fig. 5.7 A, A') suggest that LHC I sequences are more closely related to the CAB LHC II sequences whereas the parsimony tree (Fig. 5.7 B) supports the idea that the FCPs are more related to the CAB LHC I sequences; neither tree is supported by bootstrap replicates. I am inclined to believe that the distance tree is a more accurate reflection of the true relationships because the dataset has a fair bit of noise and the branches between some taxa are very long. Under these conditions, the parsimony tree may not be reliable (Stewart 1993). However, these proteins have diverged to such an extent that the resolution of such distant events may not be possible.

Of importance is the time of divergence of LHC I and LHC II CAB complexes and the separation of the lineage leading to the Chl a + c-binding proteins. It would be interesting to know if the FCP complex had evolved from ancestral PS II associated or PS I associated antennae. It has been suggested that the FCPs evolved from a LHC I or CP24-like ancestor and that the LHC II complexes of higher plants evolved after the divergence of the chromophytes and the chlorophytes (LaRoche et al. 1994). I agree that the major peripheral LHC II CAB sequences evolved after the divergence of the chlorophytes and chromophytes. However, there is insufficient (convincing) evidence, at the moment, to suggest a closer relationship of the FCPs to the Lhca-CABs. In fact, the distance trees indicate that the CAB LHC II and LHC I sequences are more closely related to one another than either is to the FCPs. This would suggest that the FCPs diverged from the ancestral LHC before their was a separation of the LHC I and LHC II type genes.

It is not possible to make a firm conclusion regarding the ancestry of the FCPs except to say that they diverged from the CABs very early. The long branches (large divergences) between these taxa, the possibility of comparing paralogous genes, and the limited information on other

members of the FCP family make any conclusions tentative. A clearer picture of the FCP/CAB relationships will develop when LHC I sequences from the chromophytes are characterized.

5.4.4 Evaluation of species relationships based on the LHC protein trees:

The usefulness of the nuclear encoded CAB proteins as an index of phylogenetic relationships is limited and will depend on several factors: the distance between the organisms studied, the evolutionary questions asked, and the availability of the appropriate sequence information. A significant obstacle in the utilization of CAB proteins for phylogeny is their small size. The relatively small number of useful characters raises questions about the reliability or significance of the topology observed. This "uncertainty" is often reflected by low bootstrap values. One also has to be cautious in the construction and interpretation of phylogenies based on a single gene/protein, which could lead to erroneous trees (Cao et al. 1994). As most of the conserved residues are thought to be membrane spanning, there is probably a functional constraint on these regions for being hydrophobic. If so, the more distant relationships may more readily reach a point where non-divergent (homoplasmic) mutations occur at a significant rate, which can mask true evolutionary relationships (Meyer et al. 1986). Another problem in using the CAB proteins for phylogeny, which is a common concern with nuclear encoded proteins, is the possibility of comparing paralogous genes. Since the CABs and FCPs are encoded by a multigene family, care must be taken to assure that the sequences used have shared the same evolutionary pathway. This is not always possible to judge due to the fragmentary nature of the sequence information. In most cases, there is only one LHC polypeptide identified from a particular organism and is assumed that it is analogous to Lhcb1 of the terrestrial plants. However, this assumption is dubious because there is insufficient characterization of the LHCs from anything other than select terrestrial plants and *Chlamydomonas*. With the Chl *a* + *c*-containing organisms, the cloned sequences typically encode the most abundant LHC in the organism. However, there is little direct evidence for a preferential association with either

photosystem. Without structural/functional information it is difficult to judge which sequences may be orthlogous or paralogous. Nonetheless, the observed gene relationships give some important clues as to taxon relationships in combination with morphological, biochemical and other molecular sequence studies; some of these cases are mentioned below.

This study, and others (Muchhal and Schwartzbach 1992; Jansson 1994), clearly show that the light harvesting proteins from *Euglena* are homologous to those of the green algae and land plants. The presence of three membranes around the chloroplast and the possession of Chl's *a* + *b* suggests that the euglenoid chloroplast was acquired secondarily; evolving from a symbiotic green algae (Gibbs 1978). Phylogenetic analyses of *psb*A (D1), *rbc*L/S (Rubisco large and small subunit), *tuf*A (Morden et al. 1992), chloroplast 5S rRNA (Somerville et al. 1992), and *psa*B (PS I core complex) (Assali and Loiseaux-de Goër 1992) also provides evidence for a close relationship between the green algal and euglenoid chloroplasts.

Euglena contains the xanthophylls diadinoxanthin and diatoxanthin and stores a β-1,3-glucan, paramylon, in the cytoplasm. These characteristics more closely resemble the chromophytes rather than the green algae. As well, an analysis of the chloroplast encoded SSU rRNA points toward a closer affinity of the *Euglena* chloroplast with those of the chromophytes (Douglas and Turner 1991b; Giovannoni et al. 1993). However, there may be problems involving biased base composition in these studies that may have caused this association (see discussion by Lockhart et al. 1994).

A couple of statements can be made concerning the branch topology for the *Mantoniella* sequences (Fig. 5.4, 5.5). First, the early branching of the *Mantoniella* sequences is in agreement with the early divergence of the Prasinophyceae from the green algal lineage based on morphological characteristics, such as the synthesis of a Chl *c*-like pigment (Mg-2,4 D) and the presence of scales on the cell body and flagella (Melkonian 1990), and from rRNA phylogenetic analysis (Steinkötter et al. 1994). Second, based on the tree topology the acquisition of a chloroplast (via a secondary endosymbiosis with a green alga), or genes, by a phagotrophic host leading to the euglenophytes would have occurred after the separation of the prasinophytes from

the other green algal lineages. An alternative interpretation that could explain the early branching of *Mantoniella* is that the LHC evolved from a paralogous member of the LHC family that was different from the gene leading to the chlorophyte LHC II gene lineage.

The relationships between the fucoxanthin-containing algae are not well resolved and there are too few complete sequences to make it interesting. However, the fucoxanthin-containing chromophytes form a distinct group separate from the peridinin-containing dinoflagellate, *Amphidinium*. This is in agreement with the traditional view of a distant relationship between the dinoflagellates and the other chromophytes. This was based on morphological characters, such as differences in the xanthophyll content, presence of a unique soluble LHC complex, the presence of only three membranes around the chloroplast, the apparent lack of histones, and persistently condensed chromosomes (Taylor, 1990). In addition, Phylogenetic analysis of nuclear rRNA consistently shows a deep divergence between the dinoflagellates and the other chromophytes (Bhattacharya et al. 1990; Hendricks et al. 1991; Cavalier-Smith et al. 1994b). Though the *i*PCPs are definitely FCP-related, whether this arose as the result of a divergence from the chromophyte line or from an independent evolution of the chloroplast, can not be resolved with this data (but see Cavalier-Smith, 1994).

The positions of the both haptophytes depicted in the FCP / *i*PCP trees (Fig. 5.6) are not consistent with standard taxonomic position of this group. The first is the early branching of *Isochrysis* before the dinoflagellate, and the other chromophytes. This would not be an accurate reflection of the organismal relationships as a number of studies indicate that the haptophytes form a sister group to the heterokont/oomycete lineage (Andersen 1991; Bhattacharya et al. 1992; Cavalier-Smith 1994; Medlin et al. 1994; Cavalier-Smith et al. 1994b). A likely explanation for the odd position of *Isochrysis* is that the FCP sequence is a paralogous gene, resulting in an erroneous tree topology. The sequence was isolated by immuno-screening a cDNA library with a FCP specific antibody that could have detected a product paralogous to the other sequences since most of the antennal proteins are immunologically related. Furthermore, there is evidence that there is only a single copy of this *Isochrysis Fcp* gene (LaRoche et al. 1994), though in terrestrial

168

plants and all other known chromophytes, the main antennal protein is encoded by a multigene family (Green et al., 1991; Bhaya et al., 1993; Apt et al., 1994; Chapter 4). Further characterization of the haptophyte family of antennae will have to be done before this can be resolved.

The second unusual relationship is the very close association of *Pavlova* with the diatoms when the tree is constructed with the available *Pavlova* FCP protein sequence data (87 characters). The *Pavlova* polypeptide binds both Chl's *a*, *c* and fucoxanthin, and is biochemically and immunologically quite similar to other FCPs (Fawley et al. 1987; Hiller et al. 1988). its current tree position is not expected. However, being a haptophyte, I would have expected it to form a deep branch at the base of the FCP lineage if it is a true reflection of organismal phylogeny. Nevertheless, the dataset is small and the species relationships may change when the complete sequence is determined.

5.4.5   Light-harvesting protein evolution: pathways and evolutionary sources

The FCP and CAB proteins are clearly homologous and were derived from a single ancestral gene, though the trees show there was probably an early separation of the FCP and CAB lineages. I propose that the antennal proteins associated with PS I were some of the first proteins that acquired the function of light-harvesting, probably from one of a photoprotective role. It would have been from this complex that the CABs and FCPs diverged at separate times. This is suggested for a few reasons: first, the LHC I proteins in the terrestrial plants originated from very early duplications as indicated by the large divergence between them, as compared to the smaller divergence between the LHC II antennal proteins. Second, the different LHC I proteins from the green algae seem to have a greater affinity for specific LHC I types of the terrestrial plants, rather than to each other. Furthermore, the green algal LHC II proteins are clearly separated from those of the terrestrial plants. This suggest that the LHC I genes had diverged into the different types before the separation of the green algae and land plants and prior to the duplication and

169

divergence of the LHC II-related genes (Hwang and Herrin 1993). Third, the presence of a CAB/FCP related LHC I complex in the red algae (Chapter 3; Wolfe et al. 1994) along with PBSs, also suggests that the LHC I antennal proteins originated prior to the membrane intrinsic LHC II protein complex. Though none have been reported, it remains to be seen whether there is an intrinisic PS II associated antennae in red algae. As well, the sequence of the red algal LHC I antennae needs to be determined to get a better idea of possible relationships.

If the presence of a suitable LHC I associated antennae complex had been established, then the loss of PBS, due to light or nutrient stresses, may have necessitated the adaptation of an LHC I-related antennae to associate with PS II. This seems plausible as some representatives of the PS II associated inner antennae of land plants (CP24 and CP29) are evolutionarily closer to the LHC I proteins than to LHC II proteins. The fact that some inner antennae of PS II are LHC I-related suggests that as additional complexes were recruited, they were added on and became more peripherally located. In addition, the LHC I-680 (Lhca2 & 3) (Knoetzel et al. 1992) and the LHC II complexes have similar fluorescence emission maxima (680 nm) making a functional transition from a PS I association to a PS II association seem plausible. Alternatively, a separate LHC I-related complex may have been stressed induced for a photoprotective role in times of PBS degradation.

Since the chloroplast is generally thought to have evolved from a cyanobacterium, the CAB/FCP related LHCs must have become the main antennae after the establishment of the chloroplast. In order to explain the present day pigmentation of the algae, there must have been at least two independent losses of phycobilisomes from the ancestral organisms; once leading to the green algae and at least one other time leading to the chromophyte plastid. Moreover, there would have to a gain in the ability to synthesize Chl b and Chl c in the green algae and chromophytes, respectively.

It is probable that the phycobilisomes were replaced by a LHC I-related complex that was induced for photoprotection during times of stress. The presence of an inducible LHC I-related system could act to protect the photosystem in the event of a loss or reduction of the PBS due to

either high light or nutrient deprivation (Bryant 1992). The PBSs are efficient but metabolically expensive, requiring about ten times more amino acids per chromophore (Bryant 1992), so an initial loss of this antenna due to a nutrient deficiency (Grossman et al. 1993) or other stress related events seems reasonable. In time and with sufficient modifications, this system could eventually replace the PBS system of the cyanobacterial or red algal endosymbiont.

Such a senario seems more likely since there are a number of LHC related complexes that are induced during either a light or nutrient stress. These include the early light-inducible proteins (ELIPs) in the terrestrial plants (Grimm et al. 1989; Adamska and Kloppstech 1994), the carotenoid biosynthesis-related (Cbr) proteins in *Dunaliella* (Levy et al. 1992) and the high light-inducible proteins (HLIPs) in cyanobacteria (Dolganov et al. 1994). These proteins have sequence similarities to the CAB proteins in the hydrophobic domains, primarily in the first and third transmembrane regions. Though a number of putative chlorophyll *a* ligands are well conserved in the ELIPS and HLIPS (Green and Kühlbrandt 1995), it has not been conclusively determined whether they bind any chlorophyll or carotenoids.

Though unrelated to the CAB proteins, there is an iron stress inducible protein (*isi*A) in cyanobacteria that is homologous to CP43 (*psb*C) (Laudenbach and Straus 1988) and binds chlorophyll (Burnap et al. 1993). It has been postulated to act as a chlorophyll reserve (Burnap et al. 1993) or as a antennal replacement in the absence of PBS (Pakrasi et al. 1985). The similarity of the *isi*A gene to the LHCs of the prochlorophytes (Hiller & Larkum; LaRouche & Partensky; van der Staay & Green; unpubl.) demonstrates the potential of stress induced proteins in the creation of novel antennal systems.

There are two potential cyanobacterial molecular sources from which the eukaryotic LHCs may have evolved after the evolution of the chloroplast; the *psb*S gene and the HLIPs. Though the *psb*S gene product has been immunologically detected in a cyanobacterium (Nilsson et al. 1990), two other groups have failed to do so (Kim et al. 1994; Vermaas, pers. comm.). At the moment it has only been cloned from tomato and spinach so its presence in cyanobacteria is uncertain. Nonetheless, the *psb*S protein binds chlorophyll and is predicted to span the

171

membrane four times (Kim et al. 1992; Funk et al. 1994). If present in the cyanobacteria, a C-terminal deletion could give a LHC precursor with three transmembrane helicies (Green and Pichersky 1994).

The HLIPs are another potential source for the evolution of the LHCs as they are known to occur in cyanobacteria (Dolganov et al. 1994) and they have homologues in red algae (Reith, unpubl.; gb X62578) and the Glaucophyta (Stirewalt and Bryant, unpubl.) (see Green and Kühlbrandt 1995). The HLIPs are only 72 amino acids long, yet there is sequence similarity to the first or third membrane spanning region of the ELIPs and CABs. It has been proposed that these proteins function as homodimers and were the evolutionary source of the eukaryotic LHCs. One can envisage a series of gene duplications and fusions that could give a LHC that spanned the membrane three times and adopted the role of light-harvesting (Dolganov et al. 1994; Green and Kühlbrandt 1995).

With the available sequence information it is not possible to make any firm conclusions regarding the ancestry of the FCPs. This will have to wait until more diverse FCP family members have been identified and sequenced. The sequence of the red algal LHC proteins will also be an important piece of information. I would predict that they will be more closely related to the FCPs rather than to the CABs.

CHAPTER 6

Summary

This dissertation demonstrates that the light-harvesting antennae of *Heterosigma carterae* form an intricate system comparable to the complexity of the LHC antennae seen in the terrestrial plants. *Heterosigma* possesses up to 12 differently migrating polypeptides that cross react to different extents with CAB and FCP specific antisera. There are four prominent LHCs in *Heterosigma* with apparent molecular masses of 18-21 kDa. The gene encoding one of these antennae, the 19.5 kDa polypeptide, was cloned and sequenced. Based on Southern hybridization and cDNA sequencing, there are approximately 6-8 gene copies encoding this polypeptide. Overall, there are probably over 20 related genes encoding the FCPs in *Heterosigma*.

The FCPs are structurally related to the CABs as determined by the immunological cross-reaction data and through direct sequence comparisons. This implies they are evolutionarily related and had evolved from the same ancestral gene. Both the CABs and the FCPs have three putative membrane spanning regions. In the pea LHC II two of the three membrane spanning regions (MSR1 and 3) interact to bind carotenoids and chlorophyll. The corresponding regions in the FCPs are very conserved. Some of the highly conserved amino acids in the FCPs are thought to bind chlorophyll and/or are important in helix-helix interactions that can help to stabilize the complex. These striking similarities indicate that the FCPs and the CABs are structurally quite similar.

The LHCs from the diverse algal taxa that utilize fucoxanthin as an accessory chlorophyll (the FCPs) are related and form a natural monophyletic group. The intrinisic peridinin-chlorophyll protein complexes (*i*PCPs) from the dinoflagellates are definitely more closely related to the FCPs than they are to the CABs and they form a sister group to the FCP clade. It seems that the acquisition of xanthophylls for a primary role in light harvesting, rather than solely

a photoprotective one, occurred early in the evolution of these LHCs. The use of different primary xanthophylls in the distinct antennae complexes (fucoxanthin, peridinin, vaucherioxanthin, etc.) was then the result of divergence following the separation of the main algal taxa. This was probably related to the light-environment the algal group experiences in the marine habitat.

The presence of CAB/FCP-related LHCs in the red algae provides a link between the antennal systems of the three major groups of photosynthetic organisms. In addition, the lack of such immunologically related LHCs in the cyanobacteria and prochlorophytes suggests that the membrane intrinisic LHCs originated following the endosymbiotic origin of the chloroplast that gave rise to the first true photosynthetic eukaryote. Such a scenario implies a monophyletic origin for the chloroplast as it is unlikely that related proteins could evolve independently in different lineages.

This work could continue in a number of directions. It would be interesting to further characterize the chlorophyll protein complexes in *Heterosigma* in addition to the sequencing of the genes encoding them. This would give an indication of the gene family complexity and how much diversity exists between the different members. This will allow the determination of the nature of the divergence that is responsible for the differential immunological cross-reactivity seen in Chapter 2. This information would be useful if antennae gene characterization studies from representatives of other major algal taxa were also being done. This comparison would be very useful in determining if specific gene types are conserved between the diverse algal groups similar to the conservation of the *Cab* gene types between the angiosperms and gymnosperms. Should specific gene types exist it would make an interesting gene evolution study; it would allow one to determine at what point certain gene duplications had occurred in relation to the phylogeny of the algae being compared. This would also complement the work currently being done with the *Cab* gene family and may help to determine which specific *Cab* gene types are more closely related to the *Fcp* genes.

The analyses of the gene family complexity have to be done in conjunction with structural and functional studies on the pigment-protein complexes. Such functional/structural studies

include whether the complex is specifically associated with either PS I or PS II, more accurate estimates of the pigment content of the different antennae complexes and the regulatory role of the divergent FCPs in the adaptation to different light regimes. These studies will allow for a better understanding of how the antennal systems of divergent organisms are organized and give an idea of the similarities and differences in the ways they adapt to harvest light. The correlation of the function of a specific complex to the estimated time of appearance will allow for the assessment of the evolution of the photosynthetic antennal systems in different organisms.

REFERENCES

Aebersold, R., J. Leavitt, L.E. Hood and S.B.H. Kent (1987). Internal amino acid sequence analysis of proteins separated by one- or two- dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc. Natl. Acad. Sci. USA* 184: 6970-6974.

Adamska, I. and K. Kloppstech (1994). Low temperature increases the abundance of early light-inducible transcript under stress conditions. *J. Biol. Chem.* 269 (48): 30221-30226.

Aebersold, R., G.W. Pipes, R.E.H. Wettenhall, H. Nilco and L.E. Hood (1990). 187: 56-65.

Alberte, R.S., A.L. Friedman, D.L. Gustafson, M.S. Rudnick and H. Lyman (1981). Light-harvesting systems of brown algae and diatoms. Isolation and characterization of the chlorophyll a/c and chlorophyll a/fucoxanthin pigment-protein complexes. *Biochim. Biophys. Acta* 635: 304-316.

Alfonso, M., G. Montoya, R. Cases, R. Rodríguez and R. Picorel (1994). Core antenna complexes, CP43 and CP47, of higher plant photosystem II. Spectral properties, pigment stoichiometry, and amino acid composition. *Biochemistry* 33: 10494-10500.

Allen, J.F. (1992). Protein phosphorylation in regulation of photosynthesis. *Biochim. Biophys. Acta* 1098: 275-335.

Allen, K.D. and L.A. Staehelin (1991). Resolution of 16 to 20 Chlorophyll-Protein Complexes Using a Low Ionic Strength Native Green Gel System. *Anal. Biochem.* 194: 214-222.

Allen, K.D. and L.A. Staehelin (1992). Biochemical characterization of photosystem II antenna polypeptides in grana and stroma membranes of spinach. *Plant Physiol.* 100: 1517-1526.

Andersen, R.A. (1991). The cytoskeleton of chromophyte algae. *Protoplasma* 164: 143-159.

Anderson, J.M. (1985). Chlorophyll-protein complexes of a marine green alga, *Codium* species (Siphonales). *Biochim. Biophys. Acta* 806: 145-153.

Anderson, J.M. and B. Andersson (1988). The dynamic photosynthetic membrane and regulation of solar energy conversion. *Trends Biochem. Sci.* 13: 351-355.

Anderson, J.M., P.K. Evans and D.J. Goodchild (1987). Immunological cross-reactivity between the light-harvesting chlorophyll *a/b*-proteins of a marine green alga and spinach. *Physiol. Plantarum.* 70: 597-602.

Apt, K.E., D. Bhaya and A.R. Grossman (1994). Characterization of genes encoding the light-harvesting proteins in diatoms: biogenesis of the fucoxanthin chlorophyll *a/c* protein complex. *J Applied Phycol.* 6: 225-230.

Apt, K.E., N.E. Hoffman and A.R. Grossman (1993). The γ-subunit of R-phycoerythrin and its possible mode of transport into the plastid of red algae. *J. Biol. Chem.* 268(22): 16208-16215.

Arsalane, W., B. Rousseau and J.C. Thomas (1992). Isolation and Characterization of Native Pigment-Protein Complexes From Two Eustigmatophyceae. *J. Phycol.* 28: 32-36.

Assali, N.E. and S. Loiseaux-de Goër (1992). Sequence and phylogeny of the psaB gene of *Pylaiella littoralis* (Phaeophyta). *J. Phycol.* 28: 209-213.

Assali, N.E., R. Mache and L.-d. Goër (1990). Evidence for a composite phylogenetic origin of the plastid genome of the brown alga Pylaiella littoralis (L.) Kjellum. *Plant Mol. Biol.* 15(307-315):

Awramik, S.M. (1992). The oldest records of photosynthesis. *Photosyn. Res.* 33: 75-89.

Barbato, R., F. Rigoni, M.T. Giardi and G.M. Giacometti (1989). The minor antenna complexes of an oxygen evolving photosystem II preparation: purification and stoichiometry. *FEBS lett.* 251(1/2): 147-154.

Barrett, J. and J.M. Anderson (1980). The P-700-chlorophyll a-protein complex and two major light-harvesting complexes of *Acrocarpia paniculata* and other brown seaweeds. *Biochim. Biophys. Acta* 590: 309-323.

Bassi, R., B. Pineau, P. Dainese and J. Marquardt (1993). Carotenoid-binding proteins of photosystem II. *Eur. J. Biochem.* 212: 297-303.

Bassi, R., F. Rigoni and G.M. Giacometti (1990). Chlorophyll binding proteins with antenna function in higher plants and green algae. *Photochem. Photobiol.* 52(6): 1187-1206.

Bassi, R. and D. Simpson (1987). Chlorophyll-protein complexes of barley photosystem I. *Eur. J. Biochem.* 163: 221-230.

Bassi, R., S.Y. Soen, G. Frank, H. Zuber and J.D. Rochaix (1992). Characterization of chlorophyll *a/b* proteins of photosystem I from *Chlamydomonas reinhardtii. J. Biol. Chem.* 267(36): 25714-25712.

Bengis, C. and N. Nelson (1975). Purification and properties of the Photosystem I reaction center from chloroplasts. *J. Biol. Chem.* 250: 2783-2788.

Bennetzen, J.L. and B.D. Hall (1982). Codon selection in yeast. *J. Biol. Chem.* 257(6): 3026-3031.

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

Bhattacharya, D., H.J. Elwood, L.J. Goff and M.L. Sogin (1990). Phylogeny of *Gracilaria lemaneiformis* (Rhodophyta) based on sequence analysis of its small subunit ribosomal RNA coding region. *J. Phycol.* 26(181-186):

Bhattacharya, D., L. Medlin, P.O. Wainwright, E.V. Ariztia, C. Bibeau, S.K. Stickel and M.L. Sogin (1992). Algae containing chlorophylls *a + c* are polyphyletic: molecular evolutionary analysis of the Chromophyta. *Evolution* 46(6): 1801-1817.

Bhaya, D. and A. Grossman (1991). Targeting proteins to diatom plastids involves transport through an endoplasmic reticulum. *Mol. Gen. Genet.* 229: 400-404.

Bhaya, D. and A.R. Grossman (1993). Characterization of gene clusters encoding the fucoxanthin chlorophyll proteins of the diatom *Phaeodactylum tricornutum. Nucl. Acid Res.* 21(19): 4458-4466.

Bjorland, T. and S. Liaaen-Jensen (1989). Distribution patterns of carotenoids in relation to chromophyte phylogeny and systematics. *The Chromophyte Algae: Problems and Perspectives*. Oxford, Clarendon Press. 37-60.

Blankenship, R.E. (1992). Origin and early evolution of photosynthesis. *Photosyn, Res.* 33: 91-111.

Boczar, B.A., T.P. Delaney and R.A. Cattolico (1989b). The gene for the ribulose-1,5-bisphosphate carboxylase small subunit protein of the marine chromophyte *Olisthodiscus luteus* is similar to that of a chemoautotrophic bacterium. *Proc. Natl. Acad. Sci. USA* 86: 4996-4999.

Boczar, B.A. and B.B. Prezelin (1987). Chlorophyll-protein complexes from the red-tide dinoflagellate, *Gonyaulax polyedra* Stein. *Plant Physiol.* 83: 805-812.

Boczar, B.A. and B.B. Prezelin (1989). Organization and Comparison of Chlorophyll-Protein Complexes from Two Fucoxanthin-Containing Algae: *Nitzschia closterium* (Bacillariophyceae) and *Isochrysis galbana* (Prymnesiophyceae). *Plant Cell Physiol.* 30(7): 1047-1056.

Boczar, B.A., B.B. Prezelin, J.P. Markwell and J.P. Thornber (1980). A chlorophyll c-containing pigment-protein complex from the marine dinoflagellate, *Glenodinium* SP. *FEBS lett.* 120(2): 243-247.

Brinkmann, H., P. Martinez, F. Quigley, W. Martin and R. Cerff (1987). Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. *J. Mol. Evol.* 26: 320-328.

Brown, J.S. (1988). Photosynthetic pigment organization in diatoms (Bacillariophyceae). *J. Phycol.* 24: 96-102.

Bryant, D.A. (1992). Puzzles of chloroplast ancestry. *Current Biology* 2(5): 240-242.

Büchel, C. and C. Wilhelm (1993). Isolation and characterization of a photosystem I-associated antenna (LHCI) and a photosystem I-core complex from the chlorophyll c-containing alga *Pleurochloris meiringensis* (Xanthophyceae). *J. Photochem. Photobiol. B: Biol.* 20: 87-93.

Büchel, C., C. Wilhelm, N. Hauswirth and A. Wild (1992). Evidence for a lateral heterogeneity by patch-work like areas enriched with photosystem I complexes in the three thylakoid lamellae of Pleurochloris meiringensis (Xanthophyceae). *Crypt. Bot.* 2: 375-386.

Bullerjahn, G.S., T.C. Jensen, D.M. Sherman and L.A. Sherman (1990). Immunological characterization of the *Prochlorothrix hollandica* and *Prochloron* sp. chlorophyll a/b antenna proteins. *FEMS Micobiol. letts.* 67: 99-106.

Bullerjahn, G.S., H.C.P. Matthijs, L.R. Mur and L.A. Sherman (1987). Chlorophyll-protein composition of the thylakoid membrane from *Prochlorothrix hollandica*, a prokaryote containing chlorophyll b. *Eur. J. Biochem.* 168: 295-300.

Burnap, R.L., T. Troyan and L.A. Sherman (1993). The highly abundant chlorophyll-protein complex of iron-deficient *Synechococcus* sp. PCC7942 (CP43') is encoded by the isiA gene. *Plant Physiol.* 103: 893-902.

Campbell, W.H. and G. Gowri (1990). Codon usage in higher plants, gren algae, and cyanobacteria. *Plant Physiol.* 92: 1-11.

Camm, E.L. and B.R. Green (1989). The chlorophyll ab complex, CP29, is associated with the photosystem II reaction centre core. *Biochim. Biophys. Acta* 974: 180-184.

Cao, Y., J. Adachi, A. Janke, S. Pääbo and M. Hasegawa (1994). Phylogenetic relationships among Eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. *J. Mol. Evol.* 39: 519-527.

Caron, L. and J. Brown (1987). Chlorophyll-carotenoid protein complexes from the diatom, *Phaeodactylum tricornutum*: Spectropotometric, pigment and polypeptide analyses. *Plant Cell Physiol.* 28(5): 775-785.

Caron, L., R. Remy and C. Berkaloff (1988). Polypeptide composition of light-harvesting complexes from some brown algae and diatoms. *FEBS lett.* 229(1): 11-15.

Cattolico, R.A., J.C. Boothroyd and S.P. Gibbs (1976). Synchronous growth and plastid replication in the naturally wall-less alga *Olisthodiscus luteus*. *Plant Physiol.* 57: 497-503.

Cavalier-Smith (1987). Glaucophyceae and the origin of plants. *Evol. Trends Plants* 2: 75-78.

Cavalier-Smith, T. (1982). The origins of plastids. *Biol. J. Linn. Soc.* 17: 289-306.

Cavalier-Smith, T. (1993). The origin, losses and gains of chloroplasts. *Origins of Plastids*. New York, Chapman and Hall. 291-348.

Cavalier-Smith, T. (1994). Origin and relationships of haptophyta. *The Haptophyte Algae*. Oxford, Clarendon press. 413-435.

Cavalier-Smith, T., M.T.E.P. Allsopp and E.E. Chao (1994b). Chimeric conundra: are nucleomorphs and chromists monophyletic or polyphyletic? *Proc. Natl. Acad. Sci. USA* 91: 11368-11372.

Chitnis, P.R., D. Purvis and N. Nelson (1991). Molecular cloning and targetted mutagenesis of the gene *psaF* encoding subunit III of photosystem I from the cyanobacteria *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* 266(30): 20146-20151.

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

Chromczynski, P. and N. Sacchi (1987). Single-step method of RNA isolation by acid guanidinium tiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162: 156-159.

Chrystal, J. and A.W.D. Larkum (1987). Pigment-protein complexes and light harvesting in Eustigmatophyte alage. *Progress in Photosynthesisi Research*. Dordrecht, Martinus Nijhoff Publishers. 189-192.

Chu, Z.X. and J.M. Anderson (1985). Isolation and characterization of a siphonaxanthin-chlorophyll *a/b*-protein complex of photosystem I from a *Codium* species (Siphonales). *Biochim. Biophys. Acta* 806: 154-160.

Church, G.M. and W. Gilbert (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* 81: 1991-1995.

Cramer, W.A., S.E. Martinez, D. Huang, G.-S. Tae, R.M. Everly, J.B. Heymann, R.H. Cheng, T.S. Baker and J.L. Smith (1994). Structural aspects of the cytochrome *b6f* complex; structure of the lumen-side domain of cytochrome *f*. *J. Bioenerg. Biomembr.* 26(1): 31-47.

Cunningham, F.X. and J.A. Schiff (1986). Chlorophyll-protein complexes from *Euglena gracilis* and mutants defective in chlorophyll *b*: polypeptide composition. *Plant Physiol.* 80: 231-238.

Cunningham, F.X. and J.A. Schiff (1986b). Chlorophyll-protein complexes from *Euglena gracilis* and mutants deficient in chlorophyll I: pigment composition. *Plant Physiol.* 80: 223-230.

Dayhoff, M.O., R.M. Schwartz and B.C. Orcutt (1978). A model of evolutionary change in proteins. *Atlas of protein sequence and structure*. Washington, D.C., National Biomedical Research Foundation. 345-352.

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

Demmin, D.S., Y.C. Stockinger and L.L. Walling (1989). Phylogenetic relationships between the chlorophyll a/b binding protein (CAB) multigene family: an intra- and interspecies study. *J. Mol. Evol.* 29: 266-279.

Dolganov, N.A.M., D. Bhaya and A.R. Grossman (1994). Cyanobacterial protein with similaroty to the chlorophyll a/b binding proteins of higher plants: evolution and regulation. *Proc. Natl. Acad. Sci. USA* In Press:

Douady, D., B. Rousseau and L. Caron (1994). Fucoxanthin-chlorophyll *a/c* light-harvesting complexes of *Laminaria saccharina*: partial amino acid sequences and arrangement in thylakoid membranes. *Biochem.* 33(11): 3165-3170.

Douglas, S.E. (1994b). Chloroplast origins and evolution. *The molecular biology of cyanobacteria*. Boston, Kluwer Academic Publishers.

Douglas, S.E. and D.G. Durnford (1989). The small subunit of ribulose-1,5-bisphosphate carboxylase is plastid-encoded in the chlorophyll *c*-containing alga *Cryptomonas Φ*. *Plant Mol. Biol.* 13: 13-20.

Douglas, S.E., D.G. Durnford and C.W. Morden (1990). Nucleotide sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from *cryptomonas Φ*: Evidence supporting the polyphyletic origin of chloroplasts. *J. Phycol.* 26: 500-508.

Douglas, S.E. and C.A. Murphy (1994). Structural, transcriptional, and phylogenetic analysis of the atpB gene cluster from the plastid of *Cryptomonas Φ* (Cryptophyceae). *J. Phycol.* 30: 329-340.

Douglas, S.E., C.A. Murphy, D.A. Spencer and M.W. Gray (1991). Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* 350: 148-151.

Douglas, S.E. and S. Turner (1991b). Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor. *J. Mol. Evol.* 33: 267-273.

Dreyfuss, B.W. and J.P. Thornber (1994). Assembly of the light-harvesting complexes (LHCs) of photosystem II. *Plant Physiol.* 106: 829-839.

Dunahay, T.G. and L.A. Staehelin (1986). Isolation and characterization of a new minor chlorophyll a/b-protein complex (CP24) from spinach. *Plant Physiol.* 80: 429-434.

180

Dunsmuir, P., S.M. Smith and J. Bedbrook (1983). The major chlorophyll *a/b* binding protein of petunia is composed of several polypeptides encoded by a number of distinct genes. *J. Mol. Appl. Genet.* 2: 285-300.

Durnford, D.G. and B.R.G. Green (1994). Charcacterization of the light harvesting proteins of the chromophytic alga, *Olisthodiscus luteus* (*Heterosigma carterae*). *Biochim. Biophys. Acta* 1184: 118-123.

Duval, J.C., H. Jupin and C. Berkaloff (1983). Photosynthetic properties of plastids isolated from macrophytic brown seaweeds. *Physiol. Vég.* 21(6): 1145-1157.

Erickson, J.M., M. Rahire and J.D. Rochaix (1984). *Chlamydomonas reinhardtii* gene for the 32,000 mol. wt. protein of photosystem II contains four large introns and is located entirely within the chloroplast inverted repeat. *EMBO* 3(12): 2753-2762.

Fawley, M.W. and A.R. Grossman (1986). Polypeptides of a light-harvesting complex of the diatom *Phaeodactylum tricornutum* are synthesized in the cytoplasm of the cell as precursors. *Plant Physiol.* 81: 149-155.

Fawley, M.W., S.J. Morton, K.D. Stewart and K.R. Mattox (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.

Fawley, M.W., K.D. Stewart and K.R. Mattox (1986b). The novel light-harvesting pigment-protein complex of *Mantoniella squamata* (Chlorophyta): Phylogenetic implications. *J. Mol. Evol.* 23: 168-176.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27: 401-410.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39(4): 783-791.

Felsenstein, J. (1992). PHYLIP (phylogeny inference package). Seattle, WA., University of Washington.

Friedman, A.L. and R.S. Alberte (1984). A diatom light-harvesting pigment-protein complex. purification and isolation. *Plant Physiol.* 76: 483-489.

Frohman, M.A. (1990). RACE:rapid amplification of cDNA ends. *PCR Protocols*. San Diego, Academic Press. 28-38.

Funk, C., W.P. Schröder, B.R. Green, G. Renger and B. Andersson (1994). The intrinisic 22 kDa protein is a chlorophyll-binding subunit of photosystem II. *FEBS Lett.* 342: 261-266.

Gagné, G. and M. Guertin (1992). The early genetic response to light in the green unicellular alga *Chlamydomonas eugametos* grown under light/dark cycles involves genes that represent direct responses to light and photosynthesis. *Plant Mol. Biol.* 18: 429-445.

Gantt, E. (1981). Phycobilisomes. *Annual review of plant physiology*. Palo Alto, Ca., Annual Reviews Inc. 327-347.

Gantt, E. (1990). Pigmentation and photoacclimation. *Biology of the red algae*. New York, Cambridge University Press. 203-219.

Garbary, D.J. and P.W. Gabrielson (1990). Taxonomy and evolution. *Biology of the Red Algae.* Cambridge, Cambridge University Press. 477-498.

Ghanotakis, D.F. and C.F. Yocum (1990). Photosystem II and the oxygen-evolving complex. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 41: 255-276.

Gibbs, P.B. and J. Biggins (1991). *In vivo* and *in vitro* protein phosporylation studies on Ochromonas danica, an alga with a chlorophyll *a/c*/fucoxanthin binding protein. *Plant Physiol.* 97: 388-395.

Gibbs, P.B. and J. Biggins (1991). Thylakoid organization in the Chromophytic alga *Ochromonas dancia.* Isolation and characterization of a new pigment-protein complex. *Plant Physiol.* 97: 381-387.

Gibbs, S.P. (1970). The comparative ultrastructure of the algal chloroplast. *Ann. N.Y. Acad. Sci.* 175: 454-473.

Gibbs, S.P. (1978). The chloroplasts of *Euglena* may have evolved from a symbiotic green algae. *Can. J. Bot.* 56: 2883-2889.

Gibbs, S.P. (1979). The route of entry of cytoplasmically synthesized proteins of algae possessing chloroplast ER. *J. Cell Sci.* 35: 253-266.

Gibbs, S.P. (1981). The chloroplast endoplasmic reticulum: structure, function and evolutionary significance. *Inter. Rev. Cytology* 72: 49-99.

Gillott, M.A. and S.P. Gibbs (1980). The crytomonad nucleomorph: its ultrastructure and evolutionary significance. *J. Phycol.* 16: 558-568.

Giovannoni, S.J., N. Wood and V. Huss (1993). Molecular phylogeny of oxygenic cells and organelles based on small-subunit ribosomal RNA sequences. *Origins of Plastids.* New York, Chapman and Hall. 159-168.

Glazer, A.N. (1989). Light guides: directional energy transfer in photosynthetic antenna. *J. Biol. Chem.* 264(1): 1-4.

Golbeck, J.H. (1992). Structure and function of photosystem I. *Annu. Rev. Plant. Physiol. Plant Mol. Biol.* 43: 293-324.

Golbeck, J.H. and D.A. Bryant (1991). Photosystem I. *Current Topics in Bioenergetics.* San Diego, Academic Press. 83-177.

Golden, S.S., C.W. Morden and K.L. Greer (1993). Comparison of sequences and organization of photosynthesis genes among the prochlorophyte *Prochlorthrix hollandica*, cyanobacteria, and chloroplasts. *Origins of plastids. Symbiogenesis, prochlorophytes, and the origins of chloroplasts.* New York, Chapman and Hall. 141-158.

Govind, N.S., S.J. Roman, R. Iglesias-Prieto, R.K. Trench, E.L. Triplett and B.B. Prézelin (1990). An analysis of the light-harvesting peridinin-chlorophyll a-proteins from dinoflagellates by immunoblotting techniques. *Proc. R. Soc. Lond. B* 240: 187-195.

Gray, M.W. and W.F. Doolittle (1982). Has the endosymbiont hypothesis been proven? *Microbiol. Rev.* 46(1): 1-42.

Green, B.R. (1988). The chlorophyll-protein complexes of higher plant photosynthetic membranes. *Photosyn. Res.* 15: 3-32.

Green, B.R., D.G. Durnford, R. Aebersold and E. Pichersky (1992). Evolution of structure and function in the chl a/b and chl a/c antenna protein family. *Research in Photosynthesis*. New York, Kluwer Academic Publishers. 195-201.

Green, B.R. and W. Kühlbrandt (1995). Sequence conservation of light-harvesting and stress-response proteins in relation to the three-dimensional molecular structure of LHCII. *Photosyn. Res.* In press:

Green, B.R. and E. Pichersky (1993). Nucleotide sequence of an *Arabidopsis thaliana* Lhcb4 gene. *Plant Physiol.* 103: 1451-1452.

Green, B.R. and E. Pichersky (1994). Hypothesis for the evolution of three-helix chl a/b and chl a/c light-harvesting antenna proteins from two-helix and four-helix ancestors. *Photosyn. Res.* : in press.

Green, B.R., E. Pichersky and K. Kloppstech (1991). Chlorophyll a/b-binding proteins: an extended family. *Trends Biochem. Sci.* 16: 181-186.

Green, B.R., D. Shen, R. Aebersold and E. Pichersky (1992b). Identification of the polypeptides of the major light-harvesting complex of photosystem II (LHCII) with their genes in tomato. *FEBS letts* 305(1): 18-22.

Greenwood, A.D., H.B. Griffiths and U.J. Santore (1977). Chloroplasts and cell compartments in Cryptophyceae. *Br. Phycol. J.* 12: 119.

Grimm, B., E. Kruse and K. Kloppstech (1989). Transiently expressed early light-inducible thylakoid proteins share transmembrane domains with light-harvesting chlorophyll binding proteins. *Plant Mol. Biol.* 13: 583-593.

Grossman, A., A. Manodori and D. Snyder (1990). Light-harvesting proteins of diatoms: their relationship to the chlorophyll *a/b* binding proteins of higher plants and their mode of transport into plastids. *Mol. Gen. Genet.* 224: 91-100.

Grossman, A.R., M.R. Schaefer, G.G. Chiang and J.L. Collier (1993). The phycobilisome, a light-harvesting complex responsive to environmental conditions. *Microbiol. Rev.* 57(3): 725-749.

Gugliemelli, L.A. (1984). Isolation and characterization of the pigment-protein particles from the light-harvesting complex of *Phaeodactyum tricornutum*. *Biochim. Biophys. Acta* 766: 45-50.

Hansson, Ö. and T. Wydrzynski (1990). Current perceptions of photosystem II. *Photosyn. Res.* 23: 131-162.

Hara, Y. and M. Chihara (1987). Morpology, ultrastructure and taxonomy of the Raphidophycean alga *Heterosigma akashiwo*. *Bot. Mag. Tokyo* 100: 151-163.

Hara, Y., I. Inouye and M. Chihara (1985). Morphology and ultrastructure of *Olisthodiscus luteus* (Raphidophyceae) with special reference to the taxonomy. *Bot. Mag. Tokyo* 98: 251-262.

Harrison, M.A. and A. Melis (1992). Organization and stability of polypeptides associated with the chlorophyll *a-b* light-harvesting complex of photosystem-II. *Plant Cell Physiol.* 33(5): 627-637.

Hauska, G. (1986). Composition and structure of cytochrome bc_1 and b_6f complexes. *Photosynthesis III: photosynthetic membranes and light-harvesting systems*. New York, Springer-Verlag. 496-507.

Haworth, P., J.L. Watson and C.J. Arntzen (1983). The detection, isolation and characterization of a light-harvesting complex which is specifically associated with photosystem I. *Biochim. Biophys. Acta* 724: 151-158.

Hendricks, L., R. De Baere, Y. Van de Peer, P.J. Neefs, A. Goris and R. De Wachter (1991). The evolutionary Position of the rhodophyte *Porphyra umbilicalis* and the Basidiomycete *Leucosporidium scottii* among other eukaryotes as deduced from the complete sequences of small ribosomal subunit RNA. *J. Mol. Evol.* 32: 167-177.

Hibberd, D.J. and R.E. Norris (1984). Cytology and ultrastructure of *Chlorarachnion reptans* (chlorarachniophyta divisio nova, Chlorarachniophyceae classis nova). *J. phycol.* 20: 310-330.

Hill, R. and R. Bendel (1960). Function of the two cytochrome components in chloroplasts: a working hypothesis. *Nature* 186: 136-137.

Hiller, R.G., J.M. Anderson and A.W.D. Larkum (1991). The chlorophyll-protein-complexes of algae. *Chlorophylls*. Boca Raton, CRC Press. 529-547.

Hiller, R.G., A.-M. Bardin and E. Nabedryk (1987). The secondary structure content of a pigment-protein complex from the thylakoids of two chromophytic algae. *Biochim. Biophys. Acta* 894: 365-369.

Hiller, R.G. and A.W.D. Larkum (1981). Chlorophyll-proteins of the red alga *Griffithsia monilis*. *Photosynthesis III. Structure and molecular organisation of the photosynthetic apparatus*. Philadelphia, Balaban International Science Scervices. 387-396.

Hiller, R.G. and A.W.D. Larkum (1985). The chlorophyll-protein complexes of *Prochloron* sp. (Prochlorophyta). *Biochim. Biophys. Acta* 806: 107-115.

Hiller, R.G., A.W.D. Larkum and P.M. Wrench (1988). Chlorophyll proteins of the prymnesiophyte *Pavlova lutherii* (Droop) comb. nov.: identification of the major light-harvesting complex. *Biochim. Biophys. Acta* 932: 223-231.

Hiller, R.G., P.M. Wrench, A.P. Gooley, G. Shoebridge and J. Breton (1993). The major intrinsic light-harvesting protein of *Amphidinium*: characterization and relation to other light-harvesting proteins. *Photochem. and Photobiol.* 57(1): 125-131.

Hillis, D.M., M.W. Allard and M.M. Miyamoto (1993). Analysis of DNA sequence data: phylogenetic inference. *Methods in Enzymolgy*. New York, Academic Press. 456-487.

Hoffman, N.E., E. Pichersky, V.S. Malik, C. Castresana, K. Ko, S.C. Darr and A.R. Cashmore (1987). A cDNA clone encoding a photosystem I protein with homolgy to photosystem II chlorophyll a/b-binding polypeptides. *Proc. Natl. Acad. Sci.* 84: 8844-8848.

Holt, A.S. (1966). Recently characterized chlorophyll. *The Chlorophylls*. New York, Academic Press. 111-118.

Holton, T.A. and M.W. Graham (1991). A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucl. Acids Res.* 19(5): 1156.

Houlné, G. and R. Schantz (1987). Molecular analysis of the transcripts encoding the light-harvesting chlorophyll a/b protein in *Euglena gracilis*: Unusual size of the mRNA. *Curr. Genet.* 12: 611-616.

Houlné, G. and R. Schantz (1988). Characterization of cDNA sequences for LHCI apoproteins in *Euglena gracilis*: the mRNA encodes a large precursor containing several consecutive divergent polypeptides. *Mol. Gen. Genet.* 243: 479-486.

Hsu, B.-D. and J.-Y. Lee (1987). Orientation of pigments and pigment-protein complexes in the diatom *Cylindrotheca fusiformis*. A linear-dichroism study. *Biochim. Biophys. Acta* 893: 572-577.

Hulburt, E.M. (1965). Flagellates from brackish waters in the vicinity of Woods Hole, Massachusetts. *J. Phycol.* 1: 87-94.

Hwang, S. and D.L. Herrin (1993). Characterization of a cDNA encoding the 20-KDa photosystem I light-harvesting polypeptide of *Chlamydomonas reinhardtii. Curr. Genet.* 23: 512-517.

Iglesias-Prieto, R., N.S. Govind and R.K. Trench (1993). Isolation and characterization of three membrane-bound chlorophyll-protein complexes from four dinoflagellate species. *Phil. Trans. R. Soc. Lond. B* 340: 381-392.

Ikemura, T. (1982). Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. *J. Mol. Biol.* 158: 573-597.

Imbault, P., C. Wittemer, U. Johanningmeier, J.D. Jacobs and S.H. Howell (1988). Structure of the *Chlamydomonas reinhardtii* cabII-! gene encoding a chlorophyl-a/b-binding protein. *Gene* 73: 397-407.

Ingram, K. and R.G. Hiller (1983). Isolation and characterization of a major chlorophyll a/c_2 light-harvesting protein from a *Chroomonas* species (Cryptophyceae). *Biochim. Biophys Acta* 722: 310-319.

Ish-Shalom, D. and I. Ohad (1983). Organization of chlorophyll-protein complexes of photosystem I in *Chlamydomonas reinhardii. Biochim. Biophys. Acta* 722: 498-507.

Jain, R., R.H. Gomer and J.J. Murtagh (1992). Increasing specificity from the PCR-RACE technique. *BioTechniques* 12(1): 58-59.

Jansson, S. (1994). The light-harvesting chlorophyll a/b-binding proteins. *Biochim. Biophys. Acta* 1184: 1-19.

Jansson, S. and P. Gustafsson (1990). Type I and type II genes for the chlorophyll *a/b*-binding protein in the gymnosperm *Pinus sylvestris* (Scots pine): cDNA cloning and sequence analysis. *Plant Mol. Biol.* 14: 287-296.

Jansson, S. and P. Gustafsson (1991). Evolutionary conservation of the chlorophyll a/b-binding proteins:cDNAs encoding type I, II and III LHC I polypeptides from the gymnosperm Scots pine. *Mol. Gen. Genet.* 229: 67-76.

Jansson, S., E. Pichersky, R. Bassi, B.R. Green, M. Ikeuchi, A. Melis, D.J. Simpson, M. Spangfort, L.A. Staehelin and J.P. Thornber (1992). A nomenclsture for the genes encoding the chlorophyll *a/b*-binding proteins of higher plants. *Plant Mol. Biol. Reporter* 10(3): 242-253.

Jeffery, S.W. (1989). Chlorophyll c pigments and their distribution in the chromophyte algae. *The Chromophyte Algae: Problems and Perspectives.* Oxford, Clarendon Press. 13-36.

Jeffrey, S.W. and G.F. Humphrey (1975). New spectrophotometric equations for determining chlorophylls a, b, c_1 and c_2 in higher plants, algae and natural phytoplankton. *Biochem. Physiol. Pflanzen* 167: 191-194.

Karabin, G.D., M. Farley and R.B. Hallick (1984). Chloroplast gene for Mr 32000 polypeptide of photosystem II in *Euglena gracilis* is interrupted by four introns with conserved boundary sequences. *Nucl. Acid Res.* 12(14): 5801-5812.

Katoh, T., M. Mimuro and S. Takaichi (1989). Light-harvesting particles isolated from a brown alga, *Dictyota dichotoma*. A supramolecular assembly of fucoxanthin-chlorophyll-protein complexes. *Biochim. Biophys. Acta* 976: 233-240.

Kies, L. and B.P. Kremer (1990). Phylum Glaucocystophyta. *Handbook of Protoctista.* Boston, Jones and Bartlett. 152-166.

Kim, S., E. Pichersky and C.F. Yocum (1994). Topological studies of spinach 22 KDa protein of photosystem II. *Biochim. Biophys. Acta* 1188: 339-348.

Kim, S., P. Sandusky, N.R. Bowlby, R. Aebersold, B.R.G. Green, S. Vlahakis, C.F. Yocum and E. Pichersky (1992). Characterization of a spinach *psb*S cDNA encoding the 22 kDa protein of photosystem II. *FEBS* 314(1): 67-71.

Knoetzel, J. and L. Rensing (1990). Characterization of the photosynthetic apparatus from the marine dinoflagellate *Gonyaulax polyedra* I. Pigment and polypeptide composition of the pigment-protein complexes. *J. Plant Physiol.* 136: 271-279.

Knoetzel, J. and D. Simpson (1991). Expression and organization of antenna proteins in the light- and temperature-sensitive barley mutant chlorina-[104]. *Planta* 185: 111-123.

Knoetzel, J., I. Svendsen and D.J. Simpson (1992). Identidication of the photosystem I antenna polypeptides in barley: isolation of three pigment-binding antenna complexes. *Eur. J. Biochem.* 206: 209-215.

Kowallik, K.V. (1993). Origin and evolution of plastids from chlorophyll-a+c-containing algae:suggested ancestral relationships to red and green algal plastids. *Origins of plastids.* New York, Chapman and Hall. 223-263.

Krauss, N., W. Hinrichs, I. Witt, P. Fromme, W. Pritzkow, Z. Dauter, C. Betzel, K.S. Wilson, H.T. Witt and W. Saenger (1993). Three-dimensional structure of system I of photosynthesis at 6A resolution. *Nature* 361: 326-331.

Kühlbrandt, W. and D.N. Wang (1991). Three-dimensional structure of plant light-harvesting complex determined by electron crystallography. *Nature* 350: 130-134.

Kühlbrandt, W., D.N. Wang and Y. Fujiyoshi (1994). Atomic model of plant light-harvesting complex by electron crystallography. *Nature* 367: 614-621.

Kyte, J. and R.F. Doolittle (1982). A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* 157: 105-132.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophabe T4. *Nature* 227: 680-685.

Lam, E., W. Ortiz and R. Malkin (1984). Chlorophyll a/b proteins of photosystem I. *FEBS letters* 168(1): 10-14.

Lam, E., W. Ortiz, S. Mayfield and R. Malkin (1984b). Isolation and characterization of a light-harvesting chlorophyll *a/b* protein complex associated with photosystem I. *Plant Physiol.* 74: 650-655.

Langridge, J. (1991). Molecular genetics and comparative evolution. Taunton, England, Research Studies Press LTD.

Larkum, A.W.D. and J. Barrett (1983). Light-harvesting processes in algae. *Advances in Botanical Research.* New York, Academic Press. 1-219.

Larkum, A.W.D., C. Scaramuzzi, G.C. Cox, R.G. Hiller and A.G. Turner (1994). Light-harvesting chlorophyll c-like pigment in *Prochloron. Proc. Natl. Acad. Sci. USA* 91: 679-683.

LaRoche, J., J. Bennett and P.G. Falkowski (1990). Characterization of a cDNA encoding for the 28.5-KDa LHCII apoprotein from the unicellular marine chlorophyte, *Dunaliella tertiolecta. Gene* 95: 165-171.

LaRoche, J., D. Henry, K. Wyman, A. Sukenik and P. Falkowski (1994). Cloning and nucleotide sequence of a cDNA encoding a major fucoxanthin-, chlorophyll *a/c*-containing protein from the chrysophyte *Isochrysis galbana*: implications for evolution of the cab gene family. *Plant Mol. Biol.* 25: 355-368.

Larouche, L., C. Tremblay, C. Simard and G. Bellemare (1991). Characterization of a cDNA encoding a PSII-associated chlorophyll a/b-binding protein (CAB) from *Chlamydomonas moewusii* fitting into neither type I or type II. *Curr. Genet.* 19: 285-288.

Larsson, U.K., J.M. Anderson and B. Andersson (1987). Variations in the relative content of the peripheral and inner light-harvesting chlorophyll *a/b*-protein complex (LHC II) subpopulations during thylakoid light adaption and development. *Biochim. Biophys. Acta* 894: 69-75.

Larsson, U.K., C. Sundby and B. Andersson (1987b). Characterization of two different subpopulations of spinach light-harvesting chlorophyll *a/b*-protein complex (LHCII): polypeptide composition, phosphorylation pattern and association with photosystem II. *Biochim. Biophys. Acta* 894: 59-68.

Laudenbach, D.E. and N.A. Straus (1988). Characterization of a cyanobacterial iron stress-induced gene similar to psbC. *J. Bacteriol.* 170(11): 5018-5026.

Lee, R.E. (1989). Phycology. New York, Cambridge University Press.

Legocki, R.P. and D.P.S. Verma (1981). Multiple immunoreplica technique: screening for specific proteins with a series of different antibodies using one polyacrylamide gel. *Anal. Biochem.* 111: 385-392.

Levy, H., I. Gokhman and A. Zamir (1992). Regulation and light-harvesting complex II association of a *Dunaliella* protein homologous to early light-induced proteins in higher plants. *J. Biol. Chem.* 267(26): 18831-18836.

Lewin, R.A. (1975). Extraordinary pigment composition of a prokaryotic algae. *Nature* 256: 735-737.

Li, W.-H. (1983). Evolution of duplicate genes and pseudogenes. *Evolution of genes and proteins*. Sunderland, Mass., Sinauer Associates INC. 14-37.

Liaud, M.F., C. Valentine, W. Martin, F.Y. Bouget, B. Kloareg and R. Cerff (1994). The evolutionary origin of red algae as deduced from the nuclear genes encoding cytosolic and chloroplast glyceraldehyde-3-phosphate dehydrogenases from *Chondus crispus*. *J. Mol. Evol.* 38: 319-327.

Lichtlé, C., J.C. Duval and Y. Lemoin (1987). Comparative biochemical, functional and ultrastructural studies of photosystem particles from a Cryptophyceae: *Cryptomonas rufescens*; isolation of an active phycoerytrin particle. *Biochim. Biophys. Acta* 894: 76-90.

Lichtlé, C., H. Jupin and J.C. Duval (1980). Energy transfers from photosystem II to photosystem I in *Cryptomonas rufescens* (Cryptophyceae). *Biochim. Biophys. Acta* 591: 104-112.

Lichtlé, C., R.M.L. McKay and S.P. Gibbs (1992). Immunogold localization of photosystem I and photosystem II light-harvesting complexes in cryptomonad thylakoids. *Biol. Cell* 74: 187-194.

Lichtlé, C., A. Spilar and S.P. Gibbs (1992b). Immunogold localization of light-harvesting and photosystem I complexes in the thylakoids of *Fucus serratus* (Phaeophyceae). *Protoplasma* 166: 99-106.

Livne, A., D. Katcoff, Y.Z. Yacobi and A. Sukenik (1992). Pigment-protein complexes of *Nannochloropsis* sp. (Eustigmatophyceae): an alga lacking chlorophylls *b* and *c*. *Research in Photosynthesis*. Boston, Kluwer Academic Publishers. 203-206.

Lockhart, P.J., D. Penny, M.D. Hendy and A.W.D. Larkum (1993). Is *Prochlorothrix hollandica* the best choice as a prokaryotic model for higher plant Chl *a/b* photosynthesis? *Photosyn. Res.* 37: 61-68.

Lockhart, P.J., M.A. Steel, M.D. Hendy and D. Penny (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* 11(4): 605-612.

Long, Z., S.U. Wang and N. Nelson (1989). Cloning and nucleotide sequence 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.

Ludwig, M. and S.P. Gibbs (1987). Are the nucleomorphs of cryptomonads and *Chlorarachnion* the vestigial nuclei of eukaryotic endosymbionts? *Ann. NY Acad. Sci.* 501: 198-211.

Magruder, W.H. (1984). Specialized appendages on spermatia from the red alga *Aglaothamnion neglectum* (Ceramiales,Ceramiaceae) specifically bind with trichogynes. *J. Phycol.* 20: 436-440.

Maier, U.G., C.J.B. Hofmann, S. Eschbach, J. wolters and G.L. Igloi (1991). Demonstration of nucleomorph-encoded eukaryotic small subunit ribosomal RNA in cryptomonads. *Mol. Gen. Genet.* 230: 155-160.

Malkin, R. (1992). Cytochrome bc_1 and b_6f complexes of photosynthetic membranes. *Photosyn. Res.* 33: 121-136.

Margulis, L. and R. Obar (1985). Heliobacterium and the origin of chrysoplasts. *Biosystems* 17: 317-325.

Marquardt, J. and A. Ried (1992). Fractionation of thylakoid membranes from *Porphyridium purpureum* using the detergent N-lauryl-β-iminodipropionate. *Planta* 187: 372-380.

Marsho, T.V. and B. Kok (1972). *Methods in Enzymology*. New York, Academic Press. 515-522.

Martin, W., C.C. Somerville and S. Loiseaux-de Goër (1992). Molecular phylogenies of plastid origins and algal evolution. *J. Mol. Evol.* 35: 385-404.

Matsuoka, M. (1990). Classification and characterization of cDNA that encodes the light-harvesting chlorphyll a/b binding protein of photosystem II of rice. *Plant Cell Physiol.* 31(4): 519-526.

Mawson, B.T., P.J. Morrissey, A. Gomez and A. Melis (1994). Thylakoid membrane development and differentiation: Assembly of the chlorophyll *a-b* light-harvesting complex and evidence for the origin of Mr=19, 17.5 and 13.4 kDa proteins. *Plant Cell Physiol.* 35(3): 341-351.

Mayes, S.R., K.M. Cook, S.J. Self, Z. Zhang and J. Barber (1991). Deletion of the gene encoding the photosystem II 33 kDa protein from *Synechocystis* sp. PCC 6803 does not inactivate water-splitting but increases vulnerability to photoinhibition. *Biochim. Biophys. Acta* 1060: 1-12.

Mayfield, S.P., P. Bennoun and J.-D. Rochaix (1987). Expression of the nuclear encoded OEE1 protein is required for oxygen evolution and stability of photosystem II particles in *Chlamydomonas reinhardtii*. *EMBO J.* 6(2): 313-318.

Mayfield, S.P., M. Schirmer-Rahire, G. Frank, H. Zuber and J.D. Rochaix (1989). Analysis of the genes of the OEE1 and OEE3 proteins of photosystem II complex from *Chlamydomonas reinhardtii*. *Plant Mol. Biol.* 12: 683-693.

McFadden, G. and P. Gilson (1995). Something borrowed, something green: lateral transfer of chloroplasts by secondary endosymbiosis. *TREE* 10(1): 12-17.

McFadden, G.I., P.R. Gilson and S.E. Douglas (1994a). The photosynthetic endosymbiont in cryptomonad cells produces both chloroplast and cytoplasmic-type ribosomes. *J. Cell. Sci.* 107: 649-657.

McFadden, G.I., P.R. Gilson and D.R.A. Hill (1994c). Goniomonas: rRNA sequences indicate that this phagotrophic flagellate is a close relative of the host component of cryptomonads. *Eur. J. Phycol.* 29: 29-32.

McFadden, G.I., P.R. Gilson, C.J.B. Hoffman, G.J. Adock and U.-G. Maier (1994b). Evidence that an amoeba aquired a chloroplast by retaining part of an engulfed eukaryoti alga. *Proc. Natl. Acad. Sci. USA* 91: 3690-3694.

McGrath, J.M., W.B. Terzaghi, P. Sridhar, A.R. Cashmore and E. Pichersky (1992). Sequence of the fourth and fifth photosystem II type I chlorophyll a/b-binding protein genes of *Arabidopsis thaliana* and evidence for the presence of a full complement of the extended CAB gene family. *Plant Mol. Biol.* 19: 725-733.

Medlin, L.K., G.L.A. Barker, M. Baumann, P.K. Hayes and M. Lange (1994). Molecular biology and systematics. *The Haptophyte Algae*. Oxford, Clarendon Press. 393-411.

Melkonian, M. (1990). Phylum Chlorophyta: Class Prasinophyceae. *Handbook of Protoctista*. Boston, Jones and Bartlett. 600-607.

Meyer, T.E., M.A. Cusanovich and M.D. Kamen (1986). Evidence against use of bacterial amino acid sequence data for construction of all-inclusive phylogenetic trees. *Proc. Natl. Acad. Sci. USA* 83: 217-220.

Michel, H. and J. Deisenhofer (1986). X-ray diffraction studies on a crystalline bacterial photosynthetic reaction center: a progress report and conclusions on the structure of photosystem II reaction centers. *Photosynthesis III: photosynthetic membranes and light-harvesting systems.* New York, Springer-verlag. 371-381.

Moran, R. (1982). Formulae for determination of chlorophyllous pigments extracted with N,-N-dimethylformamide. *Plant Physiol.* 69: 1376-1381.

Morden, C.W., C.F. Delwiche, M. Kuhsel and J.D. Palmer (1992). Gene phylogenies and the endosymbiotic origin of plastids. *Biosystems* 28: 75-90.

Morden, C.W. and S.S. Golden (1989). psbA genes indicate common ancestry of prochlorophytes and chloroplasts. *Nature* 337: 382-385.

Morishige, D.T. and P.J. Thornber (1992). Identification and analysis of a barley cDNA clone encoding the 31-kilodalton LHC IIa (CP29) apoprotein of the light-harvesting antenna complex of photosystem II. *Plant Physiol.* 98: 238-245.

Mörschel, E. and E. Rhiel (1987). Phycobilisomes and thylakoids: the light-harvesting system of cyanobacteria and red algae. *Membranous structures.* New York, Academic Press. 209-254.

Muchhal, U.S. and S.D. Schwartzbach (1992). Characterization of a *Euglena* gene encoding a polyprotein precursor to the light-harvesting chlorophyll *a/b*-binding protein of photosystem II. *Plant Mol. Biol.* 18: 287-299.

Mullet, J.E. (1983). The amino acid sequence of the polypeptide segment which regulates membrane adhesion (grana stacking) in chloroplasts. *J. Biol. Chem.* 258(16): 9941-9948.

Mullet, J.E., J.J. Burke and C.J. Arntzen (1980). Chlorophyll proteins of photosystem I. *Plant Physiol.* 65: 814-822.

Murata, N. and K. Satoh (1986). *Light-emission by plants and bacteria.* San Diego, Academic Press. 137-159.

Murray, E.E., J. Lotzer and M. Eberle (1989). Codon usage in plant genes. *Nucl. Acids Res.* 17(2): 477-498.

Mustardy, L., F.X. Cunningham and E. Gantt (1990). Localization and quantitation of chloroplast enzymes and light-harvesting components using immunocytochemical methods. *Plant Physiol.* 94: 334-340.

Nanba, O. and K. Satoh (1987). Isolation of a PS II reaction center consisting of D-1 and D-2 polypeptides and cytochrome b-559. *Proc. Natl. Acad. Sci. USA* 84: 109-112.

Nilsson, F., B. Andersson and C. Jansson (1990). Photosystem II characteristics of a constructed Synechocystis 6803 mutant lacking synthesis of the D1 polypeptide. *Plant Mol. Biol.* 14: 1051-1054.

Norris, B.J. and D.J. Miller (1994). Nucleotide sequence of a cDNA clone encoding the precursor of the peridinin-chlorophyll a-binding protein from the dinoflagellate *Symbiodinium* sp. *Plant Mol. Biol.* 24: 673-677.

190

Ortiz, W., E. Lam, M. Ghirardi and R. Malkin (1984). Antenna function of a chlorophyll *a/b* protein complex of photosystem I. *Biochim. Biophys. Acta* 766: 505-509.

Osafune, T., J.A. Schiff and E. Hase (1991). Stage-dependent localization of LHCP II apoprotein in the golgi of synchronized cells of *Euglena gracilis* by immunogold electron microscopy. *Exper. Cell Res.* 193: 320-330.

Owens, T.G. (1986). Photosystem II heterogeneity in the marine diatom *Phaeodactylum tricornutum. Photochem. Photobiol.* 43(5): 535-544.

Owens, T.G. (1986b). Light-harvesting in the diatom *Phaeodactylum tricornutum* II. Distribution of excitation energy between the photosystems. *Plant Physiol.* 80: 739-746.

Owens, T.G. and E.R. Wold (1986). Light-harvesting function in the diatom *Phaeodactylum tricornutum*: I. isolation and characterization of pigment-protein complexes. *Plant Physiol.* 80: 732-738.

Pakrasi, H.B., H.C. Riethman and L.A. Sherman (1985). Organization of pigment proteins in the photosystem II complex of the cyanobacterium *Anacystis nidulans* R2. *Proc. Natl. Acad. Sci. USA* 82: 6903-6907.

Pancic, P.G. and H. Strotmann (1993). Structure of the nuclear encoded γ subunit of CF_0CF_1 of the diatom *Odontella sinensis* including its presequence. *FEBS lett.* 320(1): 61-66.

Passaquet, C., J.C. Thomas, L. Caron, N. Hauswirth, F. Puel and C. Berkaloff (1991). Light-harvesting complexes of brown algae. Biochemical characterization and immunological relationships. *FEBS lett.* 280(1): 21-26.

Perasso, R., A. Baroin, L.H. Qu, J.P. Bachellerie and A. Adoutte (1989). Origin af algae. *Nature* 339: 142-144.

Peter, G.F. and J.P. Thornber (1991). Biochemical composition and organization of higher plant photosystem II light-harvesting pigment-proteins. *J. Biol. Chem.* 266(25): 16745-16754.

Peyriere, M., L. Caron and H. Jupin (1984). Pigment complexes and energy transfers in brown algae. *Photosynthetica* 18(2): 184-191.

Pichersky, E., T.G. Brock, D. Nguyen, N.E. Hoffman, B. Piechulla, S.D. Tanksley and B.R. Green (1989). A new member of the CAB gene family: structure, expression and chromosomal location of Cab-8, the tomato gene encoding the Type III chlorophyll a/b-binding polypeptide of photosystem I. *Plant Mol. Biol.* 12(257-270):

Pichersky, E., N.E. Hoffman, R. Bernatzky, B. Piechulla, S.D. Tanksley and A.R. Cashmore (1987). Molecular characterization and genetic mapping of DNA sequences encoding the Type I chlorophyll a/b-binding polypeptide of photosystem I in *Lycopersicon esculentum* (tomato). *Plant Mol. Biol.* 9: 205-216.

Pichersky, E., D. Soltis and P. Soltis (1990). Defective chlorophyll a/b-binding protein genes in the genome of a homosporous fern. *Proc. Natl. Acad. Sci. USA* 87: 195-199.

Pichersky, E., R. Subramaniam, M.J. White, J. Reid, R. Aebersold and B.R. Green (1991). Chlorophyll *a/b* binding (CAB) polypeptides of CP29, the internal chlorophyll *a/b* complex of PSII: characterization of the tomato gene encoding the 26 KDa (type I) polypeptide, and evidence for a second CP29 polypeptide. *Mol. Gen. Genet.* 227: 277-284.

Plumley, F.G., T.A. Martinson, D.L. Herrin, M. Ikeuchi and G.W. Schmidt (1993). Structural relationships of the photosystem I and photosystem II chlorophyll *a/b* and *a/c* light-harvesting apoproteins of plants and algae. *Photochem. and Photobiol.* 57(1): 143-151.

Prézelin, B.B. and F.T. Haxo (1976). Purification and characterization of peridinin-chlorophyll *a*-proteins from the marine dinoflagellates *Glenodinium sp.* and *Gonyaulax polyedra. Planta (Berl.)* 128: 133-141.

Pueschel, C.M. (1990). Cell structure. *Biology of the red algae.* New York, Cambridge University Press. 7-41.

Pyszniak, A.M. and S.P. Gibbs (1992). Immunocytochemical localization of photosystem I and the fucoxanthin-chlorophyll a/c light-harvesting complex in the diatom *Phaeodactylum tricornutum. Protoplasma* 166: 208-217.

Randall, S., A.L. Friedman, D.L. Gustafson, M.S. Rudnick and H. Lyman (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.

Rau, W. (1988). Functions of carotenoids other than in photosynthesis. *Plant Pigments.* New York, Academic Press. 231-255.

Raven, P.H. (1970). A multiple origin for plastids and mitochondria. *Science* 169: 641-646.

Redlinger, T. and E. Gantt (1983). Photosynthetic membranes of *Porphyridium cruentum.* An analysis of chlorophyll-protein complexes and heme-binding proteins. *Plant Physiol.* 73: 36-40.

Reith, M. and R.A. Cattolico (1986). Inverted repeat of *Olisthodiscus luteus* chloroplast DNA contains genes for both subunits of ribulose-1,5-bisphosphate carboxylase and the 32,000-dalton QB protein: Phylogenetic implications. *Proc. Natl. Acad. Sci. USA* 83: 8599-8603.

Reith, M. and J. Munholland (1993). A high-resolution gene map of the chloroplast genome of the red alga *Porhyra purpurea. Plant Cell* 5: 465-475.

Rhiel, E., W. Lange and E. Mörschel (1993). The unusual light-harvesting complex of *Mantoniella squamata*: supramolecular composition and assembly. *Biochim. Biophys. Acta* 1143: 163-172.

Rhiel, E. and E. Mörschel (1993). The atypical chlorophyll *a/b/c* light-harvesting complex of *Mantoniella squamata*: molecular cloning and sequence analysis. *Mol. Gen. Genet.* 240: 403-413.

Rhiel, E., E. Morschel and W. Wehrmeyer (1987). Characterization of and structural analysis of a chlorophyll a/c light-harvesting complex and of photosystem I particles isolated from thylakoid membranes of *Cryptomonas maculata* (Cryptophyceae). *Eur. J. Cell Biol.* 43: 82-92.

Riley, J.P. and T.R.S. Wilson (1967). The pigments of some marine phytoplankton species. *J. Mar. Biol. Ass. U.K.* 47: 351-362.

Robinson, M., R. Lilley, S. Little, J.S. Emtage, G. Yarranton, P. Stephens, A. Millican, M. Eaton and G. Humphreys (1984). Codon usage can affect efficiency of translation of genes in *Escherichia coli. Nucl. Acids Res.* 12(17): 6663-6671.

192

Round, F.E. and R.M. Crawford (1990). Phylum Bacillariophyta. *Handbook of Protoctista*. Boston, Jones and Bartlett. 574-597.

Rutherford, A.W. (1989). Photosystem II, the water-splitting enzyme. *Trends Biochem. Sci.* 14: 227-232.

Saffo, M.B. (1987). New light on seaweeds. *Bioscience* 37(9): 654-664.

Sagan, L. (1967). On the origin of mitosing cells. *J. Theoret. Biol.* 14: 225-274.

Saitou, N. and M. Nei (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4(4): 406-425.

Sambrook, J., E.F. Fritsch and T. Maniatis (1989). Molecular cloning, a laboratory manual. Cold Spring Harbour Laboratory Press.

Sauer, K. (1986). Photosynthetic light reactions--physical aspects. *Photosynthesis III: photosynthetic membranes and light-harvesting systems*. New York, Springer-Verlag. 85-97.

Schirmer, T., W. Bode, R. Huber, W. Sidler and H. Zuber (1985). *J. Mol. Biol.* 184: 257-277.

Schlegel, M. (1991). Protist evolution and phylogeny as discerned from small subunit ribosomal RNA sequence comparisons. *Europ. J. Protistol.* 27: 207-219.

Schmitt, A., A. Herold, C. Welte, A. Wild and C. Wilhelm (1993). The light-harvesting system of the unicellular alga *Mantoniella squamata* (Prasinophyceae): evidence for the lack of a photosystem I-specific antenna complex. *Photochem. Photobiol.* 57(1): 132-138.

Schuler, G.D., S.F. Altschul and D.J. Lipman (1990). A workbench for multiple alignment construction and analysis. *Proteins: Structure, function, and Genetics* in press:

Schuster, D.M., G.W. Buchman and A. Rashtchian (1992). A simple and efficient method for amplification of cDNA ends using 5' RACE. *Focus* 14: 41-47.

Schuster, G., D.C. Owens, Y. Cohen and I. Ohad (1984). Thylakoid polypeptide composition and light-independent phosphorylation of the chlorophyll a,b-protein in Prochloron, a prokaryote exhibiting oxygenic photosynthesis. *Biochim. Biophys. Acta* 767: 596-605.

Schwartz, E. and E. Pichersky (1990). Sequence of two tomato nuclear genes encoding chlorophyll a/b-binding proteins of CP24, a PSII antenna component. *Plant Mol. Biol.* 15: 157-160.

Schwartz, E., D. Shen, R. Aebersold, J.M. McGrath, E. Pichersky and B.R. Green (1991). Nucleotide sequence and chromosomal location of Cab11 and Cab12, the genes for the fourth polypeptide of the photosystem I light-harvesting antenna (LHCI). *FEBS lett.* 280(2): 229-234.

Sharp, P.M. and W.-H. Li (1986). Codon usage in regulatory genes in Eschericia coli does not reflect selection for 'rare' codons. *Nucl. Acids Res.* 14(19): 7737-7749.

Sidler, W., W. Wehrmeyer and H. Zuber (1988). Structural studies on chlorophyll a/c light-harvesting complex from the cryptomonad *Cryptomonas maculata*: Partial amino acid sequences. *Experientia* 44: A60.

Sleigh, M. (1989). Protozoa and other protists. New York, Cambridge University Press.

193

Sober, E. (1988). Reconstructing the past: parsimony, evolution and inference. Cambridge, MA, The MIT Press.

Somerville, C.C., S. Jouannic and S.L. Goër (1992). Sequence, proposed secondary structure, and phylogenetic analysis of the chloroplast 5S rRNA gene of the brown alga *Pylaiella littoralis* (L.) Kjellm. :

Song, P.S., P. Koka, B.B. Prézelin and F.T. Haxo (1976). Molecular topology of the photosynthetic light-harvesting pigment complex, peridinin-chlorophyll *a*-protein, from marine dinoflagellates. *Biochem.* 15(20): 4422-4427.

Spear-Bernstein, L. and K.R. Miller (1989). Unique location of the phycobiliprotein light-harvesting pigment in the Cryptophyceae. *J. Phycol.* 25: 412-419.

Staehelin, A. (1986). Chloroplasts structure and supramolecular organization of photosynthetic membranes. *Photosynthesis III : Photosynthetic membranes and light harvesting systems.* New York, Springer-Verlag. 1-84.

Steinkötter, J., D. Bhattacharya, I. Semmelroth, C. Bibeau and M. Melkonian (1994). Prasinophytes from independent lineages within the Chlorophyta: evidence from ribosomal RNA sequence comparisons. *J. Phycol.* 30: 340-345.

Stewart, A., U. Ljungberg, H.-E. Akerlund and B. Andersson (1985). Studies on the polypeptide composition of the cyanobacterial oxygen-evolving complex. *Biochim. Biophys. Acta* 808: 353-362.

Stewart, C.B. (1993). The powers and pitfalls of parsimony. *Nature* 361: 603-607.

Sukenik, A., K.D. Wyman, J. Bennett and P.G. Falkowski (1987). A novel mechanism for regulating the excitation of photosystem II in a green alga. *Nature* 327: 704-707.

Swift, H. and B. Palenik (1993). Prochlorophyte evolution and the origin of chloroplasts: morphological and molecular evidence. *Origins of plastids. Symbiogenesis, prochlorophytes, and the origins of chloroplasts.* New York, Chapman and Hall. 123-139.

Swofford, D.L. (1991). PAUP: Phylogenetic analysis using parsimony. Champaign Il, Illinois Natural History Survey.

Swofford, D.L. and G.J. Olsen (1990). Phylogeny reconstruction. *Molecular Systematics.* Sunderland, Sinauer Assoc., Inc. 411-501.

Tae, G.-S. and W.A. Cramer (1994). Topography of the heme prosthetic group of the cytochrome b-559 in the photosystem II reaction center. *Biochemistry* 33: 10060-10068.

Taguchi, F., Y. Yamamoto, N. Inagaki and K. Satoh (1993). Recognition signal for the C-terminal processing protease of D1 precursor protein in the photosystem II reaction center. *FEBS lett.* 326(1-3): 227-231.

Tanksley, S.D. and E. Pichersky (1988). Organization and evolution of sequences in the plant nuclear genome. *Plant Evolutionary Biology.* New York, Chapman and Hall. 55-83.

Taylor, F.J.R. (1987). An overview of the status of evolutionary cell symbiosis theories. *Ann. NY Acad. Sci.* 503: 1-16.

Taylor, F.J.R. (1990). Phylum dinoflagellata. *Handbook of Protoctista.* Boston, Jones and Bartlett. 419-437.

Taylor, F.J.R. (1992). The taxonomy of harmful marine phytoplanton. *Giornale Botanico Italiano* 126(2): 209-219.

Taylor, F.J.R. (1993). Current problems with harmful phytoplankton blooms in British Columbia waters. *Toxic Phytoplankton Blooms in the Sea*. Elsevier Science Publishers. 699-703.

Taylor, F.J.R. and R. Haigh (1993). The ecology of fish-killing blooms of the chloromonad flagellate *Heterosigma* in the Straight of Georgia and adjacent waters. *Toxic Phytoplankton Blooms in the Sea*. Elsevier Science Publishers. 705-710.

Thornber, J.P., D.T. Morishige, S. Anandan and G.F. Peter (1991). Chlorophyll-carotenoid proteins of higher plant thylakoids. *Chlorophylls*. Boca Ranton, CRC Press. 549-586.

Trebst, A. (1987). The three-dimensional structure of the herbicide binding niche on the reaction center polypeptides of photosystem II. *Z. Naturforsch.* 42c: 742-750.

Turner, S., T. Burger-Wiersma, S.J. Giovannoni, L.R. Mur and N.R. Pace (1989). The relationship of a prochlorophyte *Prochlorothrix hollandica* to green chloroplasts. *Nature* 337: 380-382.

Valentin, K., R.A. Cattolico and K. Zetsche (1993). Phylogenetic origin of the plastids. *Origins of plastids*. New York, Chapman and Hall. 193-212.

van Amerongen, H., B.M. van Bolhuis, S. Betts, R. Mei, R. van Grondelle, C.F. Yocum and J.P. Dekker (1994). Spectroscopic characterization of CP26, a chlorophyll *a/b* binding protein of the higher plant photosystem II complex. *Biochim. Biophys. Acta* 1188: 227-234.

van der Staay, G.W.M. (1992). Functional localization and properties of the chlorophyll *b*-binding antennae in the prochlorophyte *Prochlorothrix hollandica*. Universiteit van Amsterdam.

van der Staay, G.W.M. and L.A. Staehelin (1994). Biochemical characterization of protein composition and protein phosporylation patterns in stacked and unstacked thylakoid membranes of the prochlorophyte *Prochlorothrix hollandica*. *J. Biol. Chem.* 269(40): 24834-24844.

Vermaas, W.F.J., S. Styring, W.P. Schröder and B. Andersson (1993). Photosynthetic water oxidation: the protein framework. *Photosyn. Res.* 38: 249-263.

von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucl. Acids Res.* 14(11): 4683-4690.

von Heijne, G. (1990). The signal peptide. *J. Membrane Biol.* 115: 195-201.

Wada, M., M. Shono, O. Urayama, S. Satoh, H. Yukichi, Y. Ikawa and T. Fujii (1994). Molecular cloning of P-type ATPases on intracellular membranes of the marine alga *Heterosigma akashiwo*. *Plant Mol. Biol.* 26: 699-708.

Wallbraun, M., S. Kim, B.R. Green, B. Piechulla and E. Pichersky (1994). Nucleotide sequence of a tomato *psbS* gene. *Plant Physiol.* 106: 1703-1704.

Walne, P.L. and P.A. Kivic (1990). Phylum Euglenida. *Handbook of Protoctista*. Boston, Jones and Bartlett. 270-287.

Whatley, J. and F.R. Whatley (1981). Chloroplast evolution. *New Phytol.* 87: 233-247.

White, M.J. and B.R. Green (1987). Antibodies to the photosystem I chlorophyll a + b antenna cross-react with polypeptides of CP29 and LHCII. *Eur. J. Biochem.* 163: 545-551.

White, M.J. and B.R. Green (1987b). Polypeptides belonging to each of the three major chlorophyll *a* + *b* protein complexes are present in a chlorohyll-*b*-less barley mutant. *Eur. J. Biochem.* 165: 531-535.

White, M.J. and B.R. Green (1988). Intermittent-light chloroplasts are not developmentally equivalent to chlorina f2 chloroplasts in barley. *Photosyn. Res.* 15: 195-203.

Wiedemann, I., C. Wilhelm and A. Wild (1983). Isolation of chlorophyll-protein complexes and quantification of electron transport components in *Synura petersenii* and *Tribonema aequale*. *Photosyn. Res.* 4: 317-329.

Wilhelm, C., C. Büchel and B. Rousseau (1988). The molecular organization of chlorophyll-protein complexes in the Xanthophycean alga *Pleurochloris meiringensis*. *Biochim. Biophys. Acta* 934: 220-226.

Winhauer, T., S. Jäger, K. Valentine and K. Zetsche (1991). Structural similarities between *psb*A genes from red and brown algae. *Curr. Genet.* 20: 177-180.

Witt, D. and E. Stackebrandt (1988). Disproving the hypothesis of a common ancestry for the *Ochromonas danica* chrysoplast and *Heliobacterium chlorum*. *Arch. Microbiol.* 150: 244-248.

Wolfe, G.R., F.X. Cunningham, D.G. Durnford, B.R.G. Green and E. Gantt (1994). Evidence for a common origin of chloroplasts with light-harvesting complexes of different pigmentation. *Nature* 367: 566-568.

Wolfe, G.R., F.X. Cunningham and E. Gantt (1992). In the red alga *Porphiridium cruentum* photosystem I is associated with a putative LHCI complex. *Research in photosynthesis. Proc. IX int. Cong. Photosynthesis, Nagoya, Japan*. Dordrecht, Kluwer. 315-318.

Wolfe, G.R., F.X. Cunningham, B. Grabowski and E. Gantt (1994b). Isolation and characterization of photosystem I and II from the red alga *Porphyridium cruentum*. *Biochim. Biophys. Acta* 1188: 357-366.

Yu, L., J. Zhao, U. Mühlenhoff, D.A. Bryant and J.H. Golbeck (1993). PsaE is required for in vivo cyclic electron flow around photosystem I in the cyanobacterium *Synechococcus* sp. PCC 7002. *Plant. Physiol.* 103: 171-180.

Yurina, N.P., G.V. Karakashev, N.V. Karapetyan and M.S. Odintsova (1991). Composition and biosynthesis of thylakoid membrane polypeptides in the red alga *Cyanidium caldarium*: Comparison with the thylakoid polypeptide composition of higher plants and cyanobacteria. *Photosyn. Res.* 30: 15-23.

Zhou, Y.H. and M.A. Ragan (1993). cDNA cloning and characterization of the nuclear gene encoding chloroplast glyceraldehyde-3-phosphate dehydrogenase from the marine red alga *Gracilaria verrucosa*. *Curr. Genet.* 23: 483-489.

Zuber, H. (1986). Primary structure and function of the light-harvesting polypeptides from cyanobacteria, red algae, and purple photosynthetic bacteria. *Photosynthesis III: Photosynthetic membranes and light harvesting systems*. New York, Springer-Verlag. 238-251.