

Biochimica et Biophysica Acta 1276 (1996) 117-123



Two distinct forms of the peridinin-chlorophyll *a*-protein from *Amphidinium carterae*¹

Frank P. Sharples, Pamela M. Wrench, Keli Ou, Roger G. Hiller *

School of Biological Sciences, Macquarie University, NSW 2109, Australia Received 20 February 1996; revised 22 April 1996; accepted 22 April 1996

Abstract

Peridinin-chlorophyll *a*-proteins (PCPs) have been purified by combination of ammonium sulphate precipitation and cation exchange chromatography. The amino acid sequences of several of the most abundant forms have been deduced by direct protein sequencing and from DNA and indicate a highly conserved multi-gene family. At least two of the PCP genes are tandemly arranged. A novel form of the protein was also obtained in low yield with fewer peridinins (six vs eight) per chlorophyll *a* and with a different molecular mass (34 kDa vs 32 kDa) of its apoprotein. It had only 31% sequence identity with any of the more abundant PCP forms but retained a two-domain structure.

Keywords: Peridinin; Peridinin-chlorophyll a-protein; Amino acid sequence; Light-harvesting; Transit sequence; Dinophyceae; (Amphidinium carterae)

1. Introduction

Dinoflagellates possess two distinct light-harvesting systems, an intrinsic Chl *a*-*c*-caroteno-complex with a 19 kDa polypeptide [1,2] which is related to the other Chl *a*-*b*-binding proteins [1,3] and the unique water-soluble peridinin chl *a*-protein (PCP) [4–15]. The latter has been extensively studied and consists of either an apoprotein of 32 kDa binding 8 peridinins and 2 Chl *a* molecules (monomeric PCP) or is a dimer of 15 kDa apoproteins each of which binds 4 peridinins and 1 Chl *a* (dimeric PCP), although variations on these data for the number of peridinins/Chl have been reported [4,12,15]. A full-length cDNA of PCP from *Symbiodinium* has been sequenced and shows the 32 kDa form of PCP to consist of 2 similar domains each containing a double histidine residue as a putative Chl *a*-binding site, providing evidence that the 32

kDa apoprotein may have resulted from fusion of two similar genes [14].

A feature of all PCPs is the occurrence of electrophoretically distinct isoforms ranging in pI from 4.0 to 8.0. Although neither the physiological nor the molecular basis of these isoforms is established, it has been suggested that on the basis of the cDNA clones and Southern blotting they may represent distinct gene products [13,14]. On the other hand it has been reported that the different PCP iso-forms from *Alexandrium* separated by ion exchange are no longer present if protease inhibitors are included throughout the preparation [15]. All forms of PCP so far isolated are strongly antigenically related, recognising PCPs across species and generic boundaries [6,7,10].

During the purification of PCP of one predominant isoform for protein sequencing and crystallographic studies [9], we also purified a minor PCP form which, as detailed below, is distinguished from the most abundant forms by spectra, pigment composition, molecular mass, antigenicity and most compellingly by its primary sequence. We have also sequenced a number of genomic and RT-PCR clones which encode different variants of the most abundant PCP species, at least two of which are arranged in tandem on the genome. A preliminary account of part of this work appears in the Proceedings of the 10th International Photosynthesis Congress [16].

Abbreviations: Chl, chlorophyll; PCP, peridinin-chlorophyll *a*-protein; MFPCP, mainform of peridinin-chlorophyll *a*-protein; HSPCP, high salt form of peridinin-chlorophyll *a*-protein; LHC, light-harvesting complex; PCR, polymerase chain reaction.

^{*} Corresponding author. Fax: +61 2 8508245; e-mail: rhiller@rna.bio.mq.edu.au

¹ The sequences reported here will appear in the EMBL data base under numbers Z50792, Z50793, and X94549 and in the Swiss Protein data base under numbers P80483 and P80484.

2. Materials and methods

Amphidinium carterae CS21 was cultured as previously described [1,9]. Cells were harvested by flocculation with 0.5 mM AlK(SO₄)₂.12H₂O and centrifugation at $3000 \times g$ using a Beckman J-10 rotor. After washing in 50 mM tricine 20 mM KCl (pH 7.5), they were disrupted by a single passage through a French pressure cell at 80 MPa. Thylakoids and cell debris were removed by centrifugation at $20\,000 \times g$ for 30 min using a Beckman J-20 rotor. The resulting supernatant was subjected to ammonium sulphate precipitation and most of the PCP was recovered in the 70%-90% fraction. Precipitated PCP was dialysed exhaustively against 5 mM Na acetate (pH 5.0) then applied to a CM-Trisacryl column and developed with a linear salt gradient. The PCP fraction eluting between 0.05 and 0.15 M NaCl and having an $A_{478 \text{ nm}}/A_{280 \text{ nm}}$ ratio > 4.5 was designated mainform PCP (MFPCP). A minor PCP fraction eluting at 0.25 M salt was designated high salt PCP (HSPCP).

The ratio of peridinin to Chl *a* and HSPCP-peridinin to MFPCP-peridinin was determined in 80% acetone extracts using millimolar extinction coefficients of 76.8 at 663 nm for Chl *a* [17] and 85 at 478 nm for peridinin [18]. The ratio of peridinin in the two forms of PCP was also determined after separation of peridinin and Chl *a* by HPLC using a C₁₈ Bondpak column (Waters) and a gradient of 100% solvent A (1 MeOH; 1 CH₃CN:2H₂0) to 100% solvent B (1 MeOH:1 CH3CN). Areas under the peridinin peaks were compared after normalising the Chl *a* peaks in HSPCP and MFPCP.

Polypeptides were separated by the Laemmli method [19] using a 12% or 15% resolving gel and standard proteins to estimate relative molecular masses. Polyclonal antibodies were prepared by injecting 600 μ g purified MFPCP emulsified with Freunds complete adjuvant into a female White New Zealand rabbit followed by a boost of



Fig. 1. Elution of PCP forms from a CM-Sephacryl cation exchange column using an NaCl gradient. \blacktriangle , $A_{478 \text{ nm}}$; \blacksquare , $A_{478 \text{ nm}} / A_{670 \text{ nm}}$; \blacklozenge , NaCl concentration.



Fig. 2. Absorbance and circular dichroism spectra of MFPCP and HSPCP (a) native proteins (b) pigments in 80% acetone (c) circular dichroism. The spectra have been normalised to the Chl a Qy absorbance at 670 nm or 663 nm, respectively, or to the negative CD at 670 nm.

300 μ g MFPCP in incomplete adjuvant after 4 weeks. Western blotting was performed as previously described [1,20,21]

Total DNA was isolated using SDS and proteinase-K and purified by caesium chloride gradient centrifugation [22]. DNA fragments for cloning and sequencing were amplified from genomic DNA by PCR using primers designed from protein sequence data. For the HSPCP prim er the forw ard was 1235 5'-GC(ACTG)GA(AG)AA(AG)TT(TC)GA(TC)TGGGG corresponding to the amino acid sequence AEKFDWG and the reverse primer was 1795 5'-GC(AGCT)GC(GA)AA(AGCT)A(GA)(AGCT)GCCAT-(AT)TG corresponding to the sequence QMALFAA. To obtain the 3' end of the gene for MFPCP, cDNA fragments obtained by reverse transcription from polyA + RNA were amplified. First strand cDNA synthesis utilised a NotI primer adaptor 2360 5'-AATTCGCGGCCGC(T)₁₅ and reagents and protocol as described for the 5'-Amplified RACE Kit (Clontech). Second strand synthesis and PCR amplification was with the forward primer 1444 5'-AAG-



Fig. 3. SDS polyacrylamide electrophoresis and Western blotting of purified PCP forms. Lane 1, HSPCP; lane 2, protein standards; lane 3, HSPCP + MFPCP; lane 4, MFPCP; lane 5, protein standards; lane 6, MFPCP; lane 7, protein standards; lane 8, HSPCP + MFPCP; lane 9, HSPCP + MFPCP; lane 10, 11, 12 and 13, Western blot of lanes 6, 7, 8 and 9 with antibody to MFPCP; lane 14, HSPCP after prolonged storage at -20° C; lane 15, protein standards. Lanes 1–5 and 14–15 were obtained using 12% acrylamide gels and lanes 6–9 using 15% acrylamide gels.

GTACCTGCGTA(TC)ATGAA corresponding to the sequence KVPAYMK and 2360 as the reverse primer. A DNA fragment encoding the 3' end of one MFPCP gene, the intergenic space, leader sequence and 5'end of the next gene was amplified from genomic DNA. For this reaction Taq Extender PCR additive (Stratagene) was included. The forward primer was 1457 5'-TCCTGAAG-GAGAT(TCA)GA(TC)TGG corresponding to the sequence FLKEIDW and the reverse primer was 819 5'-ATI(GC)(AT)ICC(TCA)ATIGC(TC)TT(AG)TG(GA)-

TG corresponding to the sequence HHKAIGSI. A fulllength cDNA clone was obtained using the primer 2364 5'-GTTGTCATCATGGTTCGCAG corresponding to the 5'

```
A1 DEIGDAAKKLGDASYAFAKEVDWNN----GIFLQAPGKLQPLEALKAIDK 46
               : ::..:
 B1
       DGIADASKKFSDATYPIAEKFDWGGSSAVAKYIADASASNPRQAALAVEK 50
 47 MIVMGAAADPKLLKAAAEAHHKAIG-SISGPNGVTSRADWDNVNAALGRV 95
 :: | . ||||::||..|| ||:: .:|.|. |.|:.||.||:|:
51 lletgltmdpklvraavaahskaldtavsnpklvasKEDFAAVNEALARM 100
        IASVPENMVMDVYDSVSKITDPKVPAYMKSLVNGADAEKAYEGFLAFKDV 145
  96
                                                                            IASADKqkfaalrtafpesrelqsslfagn--ngyeaekaydsfkaltsa 148
101
        VKKSQVTSAAGPATV-----PSGDKIGVAAQQLSEASYPFLKEIDWL 187
146

    i...
    i...

149
188
        .SDVYMKPLPGVSAQQS---LKAIDKMIVMGAQADGNLLKAAAEAHHKAI 233
                          1:.....
                                                     .:|||
                                                                : ::
                                                                                KSPEIskylatasskdpkmmapgidktlevaltmnqnlinnavyahvrai 248
199
        .GSIDATG.VTSVADYAAVNAALGRVIASVPKGTVMDVYNAMAGVTDTSI 281
234
                                     249
       kgalntpgfvaerddfarvnlalakMIGSADPAKFKALLTAFPGNAD--L 296
282 PLNMFSKVNPLDANAAAKAFYTFKD-VVQAaqr 313
       .:.:|..||:|.||..|.:...||
QMALFA-ANPEQAKAAYETFVALTSAVV 322
297
```

Fig. 4. Amino acid sequence of MFPCP and HSPCP determined by a combination of protein sequencing and genomic PCR and aligned using the 'Bestfit' programme. A (upper sequence): MFPCP; B (lower sequence): HSPCP. Upper case letters show the protein-derived sequences and lower case letters the DNA-derived sequence. I. = identity; : and . = similarity; - = a space; an arrow shows the cleavage point of HSPCP on prolonged storage and Chl-binding histidines are over/underlined.

end encoding the leader sequence together with primer 2360 above. A genomic clone corresponding to the full coding region together with an additional 400 bases at the 3' end was obtained using the primer 2364 alone. For genomic and RT-PCR a regime of 3 min at 94°C, 1 min at 52°C and 2 min at 72°C followed by 34 cycles of 1 min at 94°C, 1 min at 52°C and 2 min at 72°C was employed. PCR products were separated on agarose gels, the products purified using a Gene Clean Kit (BIO-101) and cloned into the pGEM-T vector (Promega).

Standard dideoxy termination sequencing techniques [23] were used to sequence the DNA, employing a Sequenase II kit according to the manufacturer's instructions. Clones were sequenced in both directions using a combination of standard subcloning and custom oligonucleotide primers. DNA sequences were aligned using the Bestfit and Pileup programmes [24] accessed through the Australian Genomic Information Service

3. Results

The 70–90% ammonium sulphate fraction was applied to a cation-exchange column and developed with a salt gradient as shown in Fig. 1. Under the starting condition of pH 5.0 and 5 mM acetate, $\sim 20\%$ of the PCP was eluted. The salt gradient was commenced when the $A_{478 \text{ nm}}$ of the



Fig. 5. The tandem arrangement of MFPCP genes as determined by genomic PCR. Boxes indicate the coding regions of the genes, arrows the sites of the primers, heavy vertical lines the putative Chl-binding sites (His residues) and hatched the leader sequences. * position of stop codon. The nucleotide sequence for this figure is X94549.

eluate had fallen to 1/10 of the starting value. A major peak of PCP, 70% as judged by the $A_{478 \text{ nm}}$, was eluted in fractions 19–24 at ~ 0.1 M NaCl. The $A_{478 \text{ nm}}/A_{670 \text{ nm}}$ ratio had a mean value of 4.15 through this region but then a minor peak of 478 nm absorbance (2% of total $A_{478 \text{ nm}}$) was observed at fractions 29–35 (0.15 M NaCl) with an $A_{478 \text{ nm}}/A_{670 \text{ nm}}$ ratio of 3.37. After this latter PCP form had eluted, the $A_{478 \text{ nm}}/A_{670 \text{ nm}}$ ratio began to increase again. The spectral properties of the two PCP forms are shown in Fig. 2. The change in $A_{478 \text{ nm}}/A_{670 \text{ nm}}$ ratio

Α

												~ ~ -							60
GT	TGT	GAT	CAT	GGT	TCG	CAG	TGG	TAA'	GAA	GGC	TGT	GGT	GTT	GGC	GGC	AGT	CGC	TTT'T'	CTGT
			м	v	R	3	G	ĸ	ĸ	A	v	v	ь	A	A	v	A	F.	120
GC	AAC	FAG	TGT	GGT	GCA	GAA	GAG	CCA	TGG	CTT	CGT	ccc	TTC	ACC	CTT	GCG	CCA	GCG	TGCT
A	т	S	v	v	0	ĸ	S	H	G	F	v	P	S	P	L	R	Q	R	A
					~														180
GC	TGC	AGC.	AGG	AGC	AGC	AGC.	AGC	GTC.	AGC	TGC	CAC	CAT	GTT	CGC	TCC	AGC	GGC	CTT	CGCA
A	A	A	G	A	A	A	A	S	A	A	T	м	F	A	P	A	A	F	A
~~						maa		~ * *	000	.	003	.	1 mc	0.002	m00	2 000	~~~		240
GA	TGA	JAT T	TGG G	1GA n	TGC A	TGC A	GRAU K	R CAR	GCI T.	001 0	n U	A	s s	V	A DEL	11A 7	کان ک	K K	E
Ă.	5	-		2	~	~	~	••	-		2	••	5	•	••	•			300
\mathbf{I}_{T}	GGA	СТG	GAA	CAA	CGG	TAT	CTT	CTT	GCA	GGC	ACC	AGG	GAA	GCT	TCA	ACC	ACT	CGA	GGCA
v	D	W	N	N	G	I	F	L	Q	A	P	G	к	L	Q	P	L	Е	A
																			360
CT	GAA	GC	CAT	CGA	CAA	GAT	GAT	TGT	CAT	GGG	GGC	GGC	TGC	TGA	ccc	GAA	GCT	GCT	GAAG
L	ĸ	A	I	D	ĸ	м	I	v	м	G	A	A	A	Б	Р	ĸ	ь	'n	K 420
ac	AGC	rac	m C A	200	тсъ	CCA		ഹറ	ጥልጥ	таа	CAG	TGT	GAG	CGG	ccc	AAA	TGG	тGT	GACC
A	A	A	E	A	н	H	ĸ	A	I	G	s	v	s	G	P	N	G	v	т
																			480
TC.	ACG	AGC	TGA	TTG	GGA	CAA	CGT	GAA	TGC	TGC	GCT	TGG	CCG	TGT	CAT	TGC	ATC	AGT	TCCG
s	R	λ	D	W	D	N	v	N	A	A	L	G	R	v	I	A	S	v	P
	~	~ ~ ~		~ ~	~~ 1	— ——	0.002			200	0.000	~	0.0	C 1 C	101			<u>с</u> ст	540 NCCT
GA	GAA	CAT	GGT	GAT	GGA	TGT	UTA V	TGA	ante e	AG1	GIC C	CAA	GAT T	CAC m	AGA	.ICC	KAA W	v	P
E.	IN	м	•	ы	D	v	+	5	5	v	9	£	-	•	2	-		•	600
GC	GTA	CAT	GAA	.GTC	TCT	GGT	GAG	CGG	TGC	TGA	TGC	TGA	.GAA	GGC	СТА	CGA	GGG	CTT	CCTT
A	Y	м	к	s	L	v	s	G	A	D	A	E	ĸ	A	Y	E	G	F	L
																			660
GĊ	TTA	CAA	GGA	TGT	GGT	GAA	GAA	GAG	CCA	AGT	GAC	CTC	CGC	AGC	TGG	GCC	TGC	AAC	AGTG
A	F	ĸ	D	v	v	ĸ	к	s	Q	v	т	s	A	A	G	Р	A	т	720
cc	CTC	100	നവം	C 2 2	CAT		CGT	aac	man	CCA	GCA	GOT	mrc.	CGA	AGC	GTC	ста	ccc	ATTC
P	S S	G	D.	ĸ	I	G	v	A	A	0	0	L	s	E	A	s	Y	P	F
-	-	-	-							~	-								780
СТ	GAA	GGA	GAT	CGA	CTG	GCT	TTC	TGA	CGT	GTA	TAT	GAA	.GCC	ACI	CCC	TGG	TGT	GAG	CGCA
L	ĸ	E	I	D	W	L	S	D	v	Y	м	ĸ	P	L	P	G	v	s	A
~ •	~~~		~~~					~ • •	~		1000		000	maa	000		000	~~~	840 നഗരന
CA	.GCA	GTC	CTT T	GAA	1999 1	CAT	TGA	CAA 7	GA'I M	GA1	17	GAI	ມຍູຍ. ກ	700C	,	NGGC	מטט. ח	.000 A	N
¥	¥	3	5	ĸ	^	-	5	n	14	-	•			••	¥	•••	-	-	900
GC	ATT	GAA	.GGC	AGC	TGC	GGA	.GGC	CCA	CCA	CAA	AGGC	GAT	TGG	AAG	CAG	CGA	TGC	CAC	AGGT
A	L	к	A	A	A	Е	A	н	н	к	A	I	G	s	I	D	A	т	G
													_						960
GT	GAC	GTC	AGC	TGC	TGA	CTA	TGC	GGC	'AG'I	GAA	ATGC	AGC	CAC1	TGO	GCC	CG1	'GA'I	'AGC	TTCT
v	т	s	A	A	D	Ŷ	A	A	v	N	A	A	Ц	G	R	v	1	A	1020
GТ		مدت	GAG	CAC	n na n	'C A T	GGA	TGT	مىت	CAZ	ATGO	TAAT	GGC	TGO	CGC	'AAC	CGA	TAC	CTCG
v	P	K	S	T	v	M	D	v	Y	N	A	M	A	G	A	т	D	т	S
	_																		1080
AT	TCC	ACT	GAA	CAI	GTI	CTC	AAA	.GGT	GAA	.ccc	CACI	TGA	TGC	AAA	CGC	CAGC	CGC	GAA	.GGCC
I	Р	L	N	м	F	s	ĸ	v	N	₽	L	D	A	N	A	A	A	ĸ	A
																	_		1140
ΤT	°CTA	CAC	CTI	CAP	GGA	TGT	CGI	GCA	AGC	CGC	CTCA	ACC	ATC	AGC	TCF	AGCI	CCG	AAC	:rtTT
F	Y	T	F	ĸ	D	v	v	Q	A	A	Q	R	*						1200
CЛ	C L C	ייבב	чтCл	TGO	יררי	ጥርን	AGG		GAC	TGZ	ACCT	''TGT	GTT	TGC	CAC	CAGO	TCT	GTA	GTGA
91	. CAG	r1 (1	101	. 100	. CCI	. 1 GM	000			Gr								÷ • • •	1260
СТ	TGC	CGC	GAT	GAT	TGC	CAG	GTG	СТС	TTT	GAA	\TT]	GCF	ACAI	CTC	GTGI	rgTī	TTT	GAG	ACGT
														_					1320
GC	GAT	TGG	GGA	ATGF	LAAI	rece	GTG	ATO	CTO	CTI	\AGA	ACA	AAA	AA/	IAA	\AAA	AAA	AAA	AAAA
	* * * * * * * * * * CCCCCCCCCTT**																		

noted in Fig. 1 is reinforced when the pigments are removed from the apoprotein (Fig. 2B). The ratio of absorbance at 478 nm for HSPCP: MFPCP was 0.777 corresponding to peridinin: Chl a of 6.2 if the MFPCP has eight. In a separate experiment in which the pigments were separated by HPLC, the ratio of absorbance under the peridinin peaks was 0.75 (HSPCP/MFPCP) after normalisation with respect to Chl a, consistent with a 6:8 ratio of peridinin in the two forms. The main features of the circular dichroism spectrum (Fig. 2C) of HSPCP compared

в																			60
GT'	TGT	GAT	CAT	GGT	TCG	CAG	TGG	AAA	GAA	GGC	TGT	GGT	GTT	GGC	GAC	AGT	CGC	TTT	CTGT
			м	v	R	s	G	ĸ	ĸ	A	v	v	L	A	т	v	A	F	C 120
GC	CAC	TAG	TGT	GGT	GCA	GAA	GAC	TTG	TGG	CTT	CGI	ccc	TTC	ACC	CTT	GCG	CCA	GCG	TGCT
A	т	s	v	v	Q	ĸ	T	С	G	F	v	P	s	P	L	R	Q	R	A
acr	TGC	200	caa	200	100	ACC	The second se	നവന	man	ጥልሮ	ירמי	- C T T T T T T T T T T T T T T T T T T	cac	maa	200	100	നസ		180 AGAT
A	A	A	G	A	A	A	S	v	A	T	M	F	A	P	AGC	A.	F	A	D
																		1	240
GA(GAT T	TGG	TGA	CGC	TGC	CAA	GAC 77	GCT	TGG	AGA	TGC.	ATC	TTA V	TGC	ATT	CGC	CAA	GGA	GGTG
	-		2	^	^	r	•	"	ų.	5	î	5	1	'n	£	n	ĸ	-	300
GA	CTG	GAA	CAA	TGG	TAT	CTT	CTT	GCA	.GGC	ACC	AGG	GAA	GCT	GCA	ACC	ACT	CGC	GGC	ACTG
D	W	N	N	G	I	F	L	Q	A	P	G	ĸ	L	Q	₽	L	A	A	L 360
AA	GGC	CAT	CGA	CAA	GAT	GAT	TGT	CAT	GGG	GGC	AGC	GGC	TGA	ccc	GCA	GCT	GTT	GAA	GGCA
ĸ	A	I	D	ĸ	М	I	v	м	G	A	A	A	D	P	Q	L	L	ĸ	A
ccr	ncc	ጥር እ	NGC	т с л	~~ a	~ A A	aac	СУД	יידערי	~ ^ ^	тст	GAG	cee	cac	222	TCC	ጥርተሳ	220	420 CTCA
A	A	E	A	H	H	K	.000C	I	T	T	v	S	G	A	N	G	v	T	S
																			480
CG	CGC	TGA	CTG	GGA	CAA	TGT	GAA	.CGC	TGC	TCT	TGG	CCG	TGT	CAT	CTC	AGC.	AGT	TCC	GGAG
R	~	D	'n	D	14	v	14	~	î		9	n	v	+	5	^	•	F	540
AA	GAT	GGT	GAT	GGA	TGT	CTA	CAA	CTC	AGT	GGC	CAA	GAT	CAC	AGA	TCC	CAA	GGT.	ACC	GGCA
K	М	v	М	D	v	Y	N	s	v	A	ĸ	I	т	D	P	K	v	₽	A
TAC	САТ	GAA	стс	тст	GGT	GAA	CGG	TGC	TGA	TGC	TGA	.GAA	.GGC	ста	CGC	AGG	TTT	сст	CGCG
Y	м	Ŋ	s	L	v	N	G	A	D	A	Е	ĸ	A	Y	A	G	F	L	A
The second se	~ n n	003	നാന	0.00		~ ~ ~ ~	~ ~ ~	~~~	۱Cm	~ <i>~</i> ~	1 mc	mee	100	maa	000	maa	2 2 0	000	660
TTV F	CAA K	D.GGA	V	v	GAA K	GAA K	GAA N	.CCA 0	AGT V	GAC T	ATC S	A A	AGC	TGG G	P	A I	AAC T	V	P
-		-						-						-					720
TC'	TGG	TGA	CAC	TAA	TGG	CGT	GGC	GGC	CCA	GAA	GCT	TTC	AGA	TGC	GTC	CTA	cco	GTT	CCTG
8	G	D	т	1	G	v	A	A	Q	ĸ	Ц	5	μ	A	5	r	P	F.	780
AA	GCA	AAT	CGA	CTG	GCI	TTC	TGA	TGT	GTA	TCI	GAA	GCC	CCT	ccc	TGG	CGT	GAG	CGC	ACAG
ĸ	Q	I	D	W	L	S	D	v	Y	L	ĸ	P	L	P	G	v	s	A	Q
CA	GTC	CTT	GAA	AGC	CAT	TGA	TAG	GAT	GAT	CGT	'GAT	GGG	TGC	CCA	GGC	GGA	CGG	CAA	840 CGCA.
Q	s	L	ĸ	A	I	D	R	M	I	v	M	G	A	Q	A	D	G	N	A
	~		~~~			~~~		~~`						~					900
1719 L	GAA	.GGC A	GGC	TGC A	GGA	LGGC A	CCA H	LCCA H	ICAA K	.GGC A	GA'I T	rigg G	AAG S	CA1 T	CGA D	ATGC	CAA K	AGG G	v v
-											-	-	-	-	-			-	960
AC.	ATC	CGC	TGC	CGP	CTA	TGC	GGA	AGI	GAA	.TGC	AGC	TAA:	TGG	GCG	GGI	GGT	GGC	TTC	TGTG
т	s	A	A	D	Y	A	E	v	Ņ	A	A	I	G	R	v	v	A	s	v 1020
cc	CAA	GAG	CAC	GGI	CAT	GGA	TGT	CTA	CAA	.GGC	TAA	GGC	TAG	TGI	TAC	CGA	TAC	CGG	GATT
P	ĸ	S	т	v	М	D	v	Y	ĸ	A	м	A	s	v	т	D	т	G	I
~~				om	000		0.05		~~~	12.00	- -	TCC		~~~			مە	~~~	1080 COTTC
CC. P	AC'I L	TAA N	CA'i M	GTT F	.crc s	:ААА: К	UGG1 V	GAP N	P	AC1 L	D D	A A	.AAA N	.000 A	AGC A	.000. A	K	A	F
•	-			-	-				-	_	-		-						1140
TA	CAC	CTT:	CAA	.GGA	TG1	CGI	GCI	TGC	CGC	GCA	ACO	CTA	AGC	TCC	GAI	GTT	CGT	TGA	GCGI
Y	т	F	ĸ	D	v	v	Г	A	A	Q	R	*							1200
TT	GCC	ACA	.GAC	TCC	TGT	GAC	TGG	CTO	CAA	GGI	GTO	CTG	CTG	CAA	GG1	GCC	CTT	TTT	TAAT
	~~~						~				mon		COL	maa	000		n m m n	00	1260
ΤG	CGC	CATC	TGT	GTC	C'I'I	TTP	GAC	GTG	ICCA	CCF	ATG1	CTA	661	TŲC	CGI	GAC	TTA	.GC 1	1320
ΤG	ATG	CAG	TGA	TGC	TCC	GGG	GAI	GGI	TGC	TTT	GCC	GCAA	GAC	TGO	TGC	TCG	CTT	CAC	GCGGC
		000	mer ÷								mor	101mc	000	1 <b>.</b>	000	nn/nn	പന്ന		1380
AG	AGG	JCGG	rr'I'G	ACI	4CAC	.CAP	AAG	GC.	₽ئاي.	ATA	1.90	ເອກະ	ריסיי	ATG	1997	.101	919		1440
тT	TAT	TGT	GTG	CAC	TCC	TTT	TTZ	AAC	GAA	ATT	GGI	GTO	GAI	TGC	GTI	'GAT	GAG	AGC	GAAC
	mma	11.00	1 m m m			~~ ~ ~	• • • • •			• <b>አ</b> .መ/	י ה הי <i>ר</i>	י א מי	22.24	יאאר	ነጥ አ ብ	-220		GCI	1500 מבסדיו
ΤC	-1-1-C	ACC	ATC.	.900	.GA'I	GAC	.MA(	AA(	,CAC	.A.I.(		, AAU	MAC	.nnl	IAC	.nnc	-GCR		
CT	GAG	CTG	CGA	ACC	CATC	ATC	ACA	AAC											

Fig. 6. Nucleotide and amino acid sequences of 2 different MFPCPs obtained by (A) RT-PCR and (B) genomic PCR. Start of mature proteins is shown by an arrow. The underlined motif SPLR occurs in the intrinsic light-harvesting protein sequence and is a possible protease recognition site [3].

to that of MFPCP are that the positive CD at 430 nm is less pronounced, the isobestic point for the Cotton effect of the peridinin shifts from 479 nm to 492 nm and the negative component in the region of 500 nm to 560 nm is more asymmetric. The fluorescence emission at 77 K is shifted from 669 nm in MFPCP to 674 nm in HSPCP (results not shown).

The PCP forms obtained from the cation exchange column were concentrated and their apoproteins separated by SDS-PAGE (Fig. 3, lanes 1, 3 and 4). PCP comprising the major peak probably includes several isoforms as judged by chromatofocusing and separation on a size-exclusion chromatography column at low ionic strength (results not shown), but all MFPCP forms have an apoprotein of 32 kDa, whereas that of HSPCP is 34 kDa. On prolonged storage the HSPCP breaks down to two main peptides of 17 and 16 kDa (Fig. 3, lane 14). Polyclonal antibodies to MFPCP react very weakly with HSPCP (Fig. 3, compare lanes 10 and 13 with lane 12), suggesting that the two PCP forms are very different in primary structure, a view reinforced by N-terminal sequencing. MFPCP was sequenced by automated Edman degradation combined with cyanogen bromide and proteolytic cleavage and the sequence shown in Fig. 4A was obtained. Some heterogeneity was observed, notably Ala/Leu at residue 221, and more significantly for the existence of isoforms, Asp/Pro at residue 76 (not shown in figure). Genomic and RT-PCR derived sequences (see below) and comparison with the cDNA derived sequence from Symbiodinium suggest that the three C-terminal residues are Ala-Gln-Arg, although this has not been unequivocally established by direct protein sequencing. On Edman sequencing of HSPCP it was immediately apparent that it was very different from all MFPCPS and this was confirmed by sequencing a number of internal peptides generated by enzymatic degradation and separated by reverse-phase HPLC. A combination of genomic PCR and direct sequencing was used to generate the sequence of HSPCP shown in Fig. 4B. We cannot be sure that the sequence is complete at the Cterminus but the calculated mass is 34 289 Da corresponding to the value determined from migration in SDS polyacrylamide gels.

The cleavage observed on prolonged storage was shown by sequencing to take place at the Arg residue at position 170 (arrow in Fig. 4) since the upper band of Fig. 3 (lane 14) contained the N-terminus DGIAD... and the lower band began with the sequence YVPDG..... MFPCP and HSPCP are 31% identical and 55% similar in comparison to > 85% identity between monomeric MFPCPs from *Amphidinium* and *Symbiodinium*. HSPCP does, however, retain the two-domain structure of monomeric MFPCPs with a putative Chl-binding histidine in each domain.

Since the protein sequencing provided compelling evidence that MFPCPs are heterogenous, at least in part as a result of different gene expression, we investigated by genomic PCR whether such genes might be adjacent.

а	MVRSGKKAVV	LAAVAFCATS	VVOKSHGEVP	SPLRORAAAA	GAAAASAATM	FAPAAFADEI
b	********	**T******	****TC****	*******	*-*** <u>*</u> V***	******
с	***GAR**IA	VG-**VAVAC	GL**HLN***	G*RH**	PV***A*SM*	<u>M*****</u> ***
d						* * *
e						* * *
	61					120
а	GDAAKKLGDA	SYAFAKEVDW	NNGIFLQAPG	KLQPLEALKA	IDKMIVMGAA	ADPKLLKAAA
b	*****_****	*******	*******	*****A****	*******	***Q******
С	******	**S******	*******	*F******	****E****	******D**
d	*******	********	*******	*******	*******	********
e	*******	********X	*******	*F***X***X	*-*HX*	
	121					180
a	EAHHKAIGSV	SGPNGVTSRA	DWDNVNAALG	RVIASVPENM	VMDVYDSVSK	ITDPKVPAYM
b	******TT*	**A******	********	***SA***K*	*****N**A*	
С	********	********	***A****I*	**V****KAK	**A*****TA	****G****
d	********I	*******	********	*******	********	********
						0.4.0
	181					240
a	KSLVSGADAE	KAYEGFLAFK	DVVKKSQVTS	AAGPATVPSG	DKIGVAAQQL	SEASYPPLKE
b	N***N****	****	N		-TK-	*D******
ç	****N*P***	***Q***E**	***E*N**AT	SATIV	E. NA	*D*******
a					****	
r	• • • • • • • • • • •	• • • • • • • • • • • •				
	2.4.1					300
	Z41 TDUI CDUVMV	DI DOVENOOR	- VATOVMTUM	CADADONALK	******************	CSTDATCUTS
à	1046304146	FLFGV3AQQ3	******	*********	********	*********
~	******	********	*********	********	********	********
a	*********	*********	*********	*******	*******	*******
÷	********	- * XTANPEDV	*********	**AM*SA***	*GL******	****K**T
~						
	3.01					360
a	AADYAAVNAA	LGRVIASVPK	STVMDVYNAM	AGA-TDTSIP	LNMFSKVNPL	DANAAAKAFY
b	****E****	I***V*****	******K**	*SV-***G**	********	*******
c	****E****	I**LV*****	T******	**V-V*S*V*	N*L******	**V****G**
d	V*******	********	G*******	**V-*****	********	********
f	L***E****	I*HMV**AGM	*KT * * * * * * L	* * FNLGKDVG	PY*M****AA	**S**YS*LL
	361					
a	TFKDVVQLAQ	R				
b	******L***	*				
с	******E*S*	*				
d	*******	*				
f						

Fig. 7. Alignment of all MFPCP complete and partial amino acid sequences. a,b and d, *Amphidinium* (this paper); c, *Symbiodinium* [14]; e, *Alexandrium* [15]; f, *Heterocapsa* bold-face Pisano and Hiller (unpublished) and [13]. Leader sequences are underlined. All mature protein monomeric sequences begin at the D (Asp) at position 57. The dimeric PCP of *Heterocapsa* has been aligned by the Pileup programme with the C-terminal domain of other PCPs. The *Heterocapsa* sequence is incomplete at the C-terminus as Ref. [13] has a reading frame error.

Using a forward primer between the putative Chl-binding sites and a reverse primer made to the Chl-binding sites, a number of PCR products were obtained (results not shown) including one of 1272 bp. When this was sequenced it encoded the C-terminus of one MFPCP, an intergenic region and the leader sequence and N-terminal region as far as the first chlorophyll binding site of another MFPCP (Fig. 5). Full-length MFPCP sequences (Fig. 6A,B) were obtained by further genomic PCR or RT-PCR experiments utilising forward primers made to the region encoding the N-terminus of the leader sequence (see Section 2). In no case did we find evidence from any of the genomic PCR experiments for introns within the PCP coding regions. The information derived from these studies and from the literature is combined in Fig. 7 and demonstrates clearly that all MFPCPs (even those with a 15 kDa apoprotein) are closely related and are more closely related overall than HSPCP is to MFPCP within Amphidinium.

# 4. Discussion

As shown here by DNA and protein sequencing, the origin of isoforms of PCP lies at least in part in the expression of different genes but in some species [15] may be accentuated by proteolytic activity. Protease inhibitors were not employed in our experiments as no effect of PMSF was seen in preliminary purifications and our purified MFPCP gave crystals diffracting to 2.3 Å [9]. Nterminal sequencing of minor MFPCP forms gave termini as in Fig. 4, indicating that any proteolytic activity must have been C-terminal and resulting in differences in mass below the resolution of our polyacrylamide gels. Interestingly, the slow proteolytic activity associated with HSPCP which results in cleavage at an arginine residue between the two putative Chl-binding domains was not inhibited by PMSF or a cocktail of well-known protease inhibitors (Sharples and Hiller, unpublished).

Previous studies have shown that different PCP isoforms are remarkably similar in spectroscopic properties and amino acid composition. This similarity is reemphasised by the direct amino acid sequences and by the PCR-derived sequences reported here for MFPCP. The identity between Symbiodinium and Amphidinium PCPs is greater than 85% and there is a similar level of identity with the partial sequences of Heterocapsa [13] and Alexandrium [15]. It may be noted that the cDNA sequences previously reported [13] from Heterocapsa as indicating major differences in amino acid sequence at the C-terminus are in error as the start codon was incorrectly assigned. The error was compounded by a reading frame error in the region of the stop codon since moving the reading frame by one base produces a C-terminal sequence with strong identity with that of Amphidinium and Symbiodinium. The minor HSPCP form differs from MFPCP in its absorbance, CD and fluorescence emission but some of these differences are also found in a comparison of Amphidinium MFPCP with Heterocapsa PCP. HSPCP is set apart by its very different amino acid sequence and by its pigment composition being relatively deficient in peridinin. These factors, its relatively low concentration and longer wavelength fluorescence emission suggest it may play a special role, perhaps as a link between MFPCP and an intrinsic Chl-protein.

Although other workers [4,12,15] have reported variation in the ratio of peridinin to Chl *a*, most of the variations have been for peridinin/Chl *a* ratios greater than that of usually accepted 8:2. Our suggestion that HSPCP has only six peridinins per 2 Chl *a* again reinforces the unusual nature of this PCP form. The crystal structure of MFPCP has recently been solved and shows unequivocally that *Amphidinium* MFPCP has a peridinin:Chl *a* composition of 8:2 [25]. This suggests that values of 12 peridinins per monomeric PCP [12,15] in *Alexandrium* may require re-evaluation, especially in light of the similarity of apoprotein size, primary sequence, amino acid composition, CD spectrum and most importantly identity of  $A_{478 \text{ nm}}/A_{670 \text{ nm}}$  of the holoprotein with that (4.1/1) of MFPCP of *Amphidinium*.

Interpretation of the role of HSPCP is hampered by our

lack of knowledge of the location of even MFPCP. Immuno-gold labelling of Amphidinium thylakoids is inconclusive (Vesk and Hiller, unpublished observations). Although the thylakoid lumen is slightly larger than that of higher plants and green algae, it is not as distended and electron dense as it is in Cryptophytes, which contain a lumenal phycobilin [26-29]. The putative leader sequence of Symbiodinium PCP has been interpreted [14] as being lumen-directing, having an N-terminal region which resembles that of proteins exported from the endoplasmic reticulum together with a hydrophobic C-terminal region. We found similar leader sequences for 2 Amphidinium MFPCPs, although the identity between the two leader sequences is much less than between the mature proteins. Both leader sequences possess the A-X-A motif which precedes the cleavage site of typical higher plant lumenal proteins [30] and additionally there are negatively charged amino acids in the first four residues following the cleavage site, another feature of lumenally located thylakoid proteins [31]. Symbiodinium PCP leader has the sequence NFVGP which also occurs near the C-terminus of several Cab proteins and there is a similar motif GFVPSP present in the Amphidinium leader at the same point. More striking is the occurrence in the Amphidinium leader of the SPLR motif which is the C-terminus of the Chl a-ccaroteno protein [3] and postulated to be the recognition site for the arginine-specific protease responsible for conversion of the LHC polyprotein to its mature form. A similar coincidence has been noted in processing sites for Euglena preLHC. The C-terminus of the Euglena LHC leader sequence has the GXKX motif which is conserved at the C-terminus of the decapeptide which separates each mature LHC peptide in the precursor polyprotein; a common peptidase was suggested to be responsible for processing in both instances [32]. We speculate that in Amphidinium as well, an arginine-specific protease located in the stroma cleaves both preLHC to its mature polypeptides and processes prePCP to a form with a hydrophobic N-terminal extension which is lumen-directing. The same protease may also be responsible for the cleavage of purified HSPCP observed on prolonged storage.

# Acknowledgements

We thank Prof. Keith Williams and Dr. Andrew Gooley of Macquarie University Centre for Analytical Biotechnology for assistance with protein sequencing. The work was supported by the Australian Research Council (A19131108).

## References

 Hiller, R.G., Wrench, P.M., Gooley, A.A., Shoebridge, G. and Breton, J. (1993) Photochem. Photobiol. 57, 125-131.

- [2] Inglesias-Prieto, R., Govind, N.S. and Trench, R.K. (1993) Phil. Trans. R. Soc. Lond. B 340, 381–392.
- [3] Hiller, R.G., Wrench, P.M and Sharples, F.P. (1995) FEBS Lett. 363, 175-178.
- [4] Haxo, F.T., Kycia, J.H., Somers, G.F., Bennet, A and Siegelman, H.W. (1976) Plant Physiol. 57, 297–303.
- [5] Koka, P. and Song, P.-S. (1976) Biochim. Biophys. Acta 495, 220-231.
- [6] Chang, S.S. and Trench, R.K. (1984) Proc. Roy. Soc. Lond. B 222, 259–271.
- [7] Prezelin, B.B. (1987) in Botanical Monographs, Vol.21, The Biology of Dinoflagellates (ed. Taylor, F.J.R.), pp. 174–223, Blackwell, Oxford.
- [8] Govind, N.S., Roman, S.J., Inglesias-Prieto, R., Trench, R.K., Triplett, E.L. and Prezelin, B.B. (1990) Proc. Roy. Soc. Lond. B 240, 187-195.
- [9] Steck, K., Wacker, T., Welte. W., Sharples, F.P. and Hiller, R.G. (1990) FEBS Lett. 268, 48–50.
- [10] Inglesias-Prieto, R., Govind, N.S. and Trench, R.K. (1991) Proc. Roy. Soc. B 246, 275–283.
- [11] Jovine, R.V.M., Triplett, E.L. Nelson, N.B. and Prezelin, B.B. (1992) Plant Cell Physiol. 33, 733-741.
- [12] Ogata, T. and Kodama, M. (1993) in Toxic Phytoplankton Blooms in the Sea (Smayda, T.J., Shimizu, Y., eds.), pp. 901–905, Elsevier, New York.
- [13] Triplett, E.L., Jovine, R.V.M., Govind, N.S., Roman, S.J., Chang, S.S. and Prezelin, B.B. (1993) Mol. Mar. Biol. and Biotech. 2, 246-254.
- [14] Norris, B.J. and Miller, D.J. (1994) Plant Mol. Biol. 24, 673-677.
- [15] Ogata, T., Kodama, M., Nomura. S., Kobashi, M., Nozawa, T., Katoh, T. and Mimuro, M. (1994) FEBS Lett. 356, 367–371.
- [16] Hiller, R.G., Wrench, P.M. and Sharples, F.P. (1995) in Photosyn-

thesis: from Light to Biosphere (Mathis, P., ed.), Vol. 1, pp. 29-34, Kluwer, Dordrecht.

- [17] Porra, R.J., Thompson, W.A. and Kriedeman, P.E. (1989) Biochim. Biophys. Acta 975, 384–395.
- [18] Johansen, J.E., Svec, W.A., Liaaen-Jensen, S. and Haxo, F.T. (1974) Phytochemistry 13, 2261–2271.
- [19] Laemmli, U.K. (1970) Nature 227, 680-685.
- [20] Khyse-Andersen, J. (1984) J. Biochem. Biophys. Methods 10, 203– 209.
- [21] Hawkes, R., Niday, E. and Gordon, J. (1982) Anal. Biochem. 119, 142-147.
- [22] Scaramuzzi, C.D., Stokes, H.W. and Hiller, R.G. (1992) Plant Mol. Biol. 18, 467–476.
- [23] Sanger, F., Nichlen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
- [24] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
- [25] Hofmann, E., Wrench, P.M., Sharples.F.P., Hiller, R.G., Welte, W. and Diederichs, K. (1996) Science, in press.
- [26] Gantt, E., Edwards, M.R. and Provosoli, L. (1971) J. Cell Biol. 48, 412–419.
- [27] Spear-Bernstein, L. and Miller, K.R. (1989) J. Phycol. 25, 412-419.
- [28] Ludwig, M. and Gibbs, S.P. (1992) J. Cell Biol. 108, 875-884.
- [29] Vesk, M. Dwarte, D., Fowler, S. and Hiller. R.G. (1992) Protoplasma 70, 166–170.
- [30] von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1993) Eur. J. Biochem. 180, 535-545.
- [31] Clausemeyer, S. Klosgen, R.B. and Herrmann, R.G. (1993) J. Biol. Chem. 268, 13869–13876.
- [32] Kishore, R., Muchal, U.S. and Schwartzbach, S.D. (1993) Proc. Natl. Acad. Sci. USA 90, 11845–11849.