A genomic clone encoding a cryptophyte phycoerythrin α-subunit

Evidence for three α-subunits and an N-terminal membrane transit sequence

Jon Jenkins[1], Roger G. Hiller[1], Jim Speirs[2] and Jasminka Godovac-Zimmermann[3]

[1]*School of Biological Sciences, Macquarie University, N.S.W. 2109, Australia,* [2]*CSIRO Division of Horticulture, North Ryde, N.S.W. 2113, Australia* and [3]*John Curtin School of Medical Research, Australian National University, Canberra, A.C.T. 2600, Australia*

Received 11 July 1990; revised version received 10 September 1990

A genomic library of *Chroomonas* (Cryptophyceae) DNA has been constructed in λEMBL4. Using a synthetic oligomer as a hybridization probe, a clone containing a phycoerythrin α-subunit has been obtained and sequenced. The principal α_1- and α_2-subunits of the holoprotein have been partially sequenced by sequential Edman degradation and differ from the DNA derived sequence, providing evidence for at least 3 α-subunit genes. The nucleotide sequence of the α-subunit gene is GC rich and encodes an N-terminal extension which is putatively thylakoid-lumen directing.

Phycoerythrin; α-Subunit; Amino acid sequence; Genomic clone; Signal sequence; *Chroomonas* CS24; Cryptophyceae

1. INTRODUCTION

There are two groups of eucaryotic algae, the Rhodophytes and Cryptophytes, which have phycobilin accessory light-harvesting pigments. In the rhodophytes the phycobilins are organised into phycobilisomes attached to the exterior of the thylakoids, an arrangement also found in the cyanobacteria. In the cryptophytes, phycocyanin and phycoerythrin are never present together and the single phycobilin is located in the intrathylakoid space [1]. A common feature of the cryptophyte phycobilins is the existence of multiple forms, revealed by isoelectric focusing, although the molecular basis for these forms is not known [2-5]. The structure of cryptophyte phycoerythrin (PE) is $\alpha_1\alpha_2(\beta)_2$ where the α_1 and α_2 are of different mobilities on SDS PAGE and have different sequences [5-7]. From a study of *Chroomonas* CS24 phycoerythrin it was suggested that the origin of the multiple forms of PE lies in the α-subunits and a model was proposed in which the $(\alpha\ \beta)_2$ structure could comprise any combination of 2 out of 4 different α-units with an invariant β-unit. The same α-units could be present in two copies in this model, which could generate up to 10 electrophoretically distinct forms of the holoprotein [4].

To further investigate the structure of cryptophyte PE, we have made a genomic library of *Chroomonas* DNA and have isolated and sequenced a clone encoding an α-subunit. The peptide sequence deduced from the clone, together with conventional Edman sequencing of α_1- and α_2-subunits, provides clear evidence for the existence of three α-subunit genes as well as for a leader sequence with similarities to known thylakoid lumen-targeting sequences.

Correspondence address: R.G. Hiller, School of Biological Sciences, Macquarie University, N.S.W. 2109, Australia

2. MATERIALS AND METHODS

2.1. *DNA isolation and purification*

Chroomonas sp CS-24 was originally from CSIRO (Dept. of Fisheries, Hobart, Australia 7001) and was grown in a modified Fe medium as previously described [7]. One litre of exponentially growing *Chroomonas* cells was harvested by centrifugation (700 × *g*, 5 min, 4°C), washed twice in 50 ml of 0.85% saline and resuspended in 10 ml of extraction buffer (400 mM NaCl, 40 mM EDTA, 100 mM Tris-HCl, pH 8.0). The resuspended cells were made to 1% with SDS, proteinase K (Boehringer) was added to 100 $\mu g \cdot ml^{-1}$ and the suspension incubated at 50°C for 2 h with occasional inversion. The solution was extracted by gentle shaking with an equal volume of phenol/chloroform (1:1, v/v), centrifugation in a bench centrifuge and collection of the aqueous phase. The extraction was repeated 3 further times. DNA was precipitated from the aqueous phase by the addition of two volumes of cold ethanol, gentle mixing of the phases and spooling of the DNA onto a glass rod. The DNA was washed in 70% ethanol, was air-dried and was allowed to dissolve in TE (1 mM Tris-HCl, 0.1 mM EDTA, pH 8), for 12 h at 4°C. The DNA was then treated with RNase A (100 $\mu g \cdot ml^{-1}$), for a further 1 h at 37°C. After a further 3 extractions with equal volumes of phenol/chloroform the DNA was precipitated by the addition of two volumes of ethanol and spooling, washed in 70% ethanol and air-dried. The DNA pellet was resuspended in TE and repricipitated twice from 3.75 M ammonium acetate, pH 4.8, by the addition of two volumes of ethanol. The DNA was finally purified by isopycnic centrifugation in CsCl (density 1.7 $g \cdot ml^{-1}$) (Beckman Ti 80 rotor, 50 000 rpm, 72 h), 20°C, the band collected, analysed by spectrophotometry and dialysed against TE buffer.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

2.2. *Preparation of a library*

Forty μg of DNA from *Chroomonas* sp. were partially restricted with *Sau*3A in the presence of 2.5 mM spermidine, and size fractionated on a 0.4% agarose gel. The DNA from the 10–20 kb region was extracted and purified by standard procedures.

The size fractionated *Chroomonas* DNA was ligated into λEMBL4 [8], packaged, and a library produced by plating on *E. coli* strain NM539.

Partial amino acid sequences of the α-subunits of cryptophyte phycocyanin (PC) and phycoerythrin (PE) have been published [6] and there is a conserved region of relatively low redundancy beginning at residue 37 of Fig. 3.

		37					
PC-645	α_1	D	D	E	M	C	V
		GAU/C	GAU/C	GAA/G	AUG	UGU/C	GU*
PC-645	α_2	D	D	Q	-	C	V
		GAU/C	GAU/C	CAA/G	-	UGU-C	GU*
PE-545	α_1	D	D	E	M	M	
		GAU/C	GAU/C	GAA/G	AUG	AUG	
PE-545	α_2	D	D	E	M	L	V
		GAU/C	GAU/C	GAA/G	AUG	UUA/G CU*	GU*

The sequence for PE-545 α_1 was chosen as the basis for a mixed synthetic oligomer with the additional assumption that valine is conserved at position 43. A 17-nucleotide mixed oligomer ($GA^T/_CGA^T/_CGA^A/_GATGATGGT$, 8 degeneracies) was synthesised on a Pharmacia no. 560 Gene Synthesiser.

Labelling of the synthetic oligomer, screening of the phage library, restriction digestions, fragment cloning and subcloning and Southern blot hybridizations were carried out by standard methods [9,10]. Nucleotide sequencing was carried out by the oligopriming and chain termination method [11] using a Sequenase II kit (USB). Sequence was determined using universal primer and oligonucleotide primers synthesised on the basis of determined sequence. Sequence analysis was undertaken with the aid of the Cornell Package [12].

2.3. *Phycoerythrin isolation and subunit separation*

This was by means of ammonium sulphate precipitation and size exclusion chromatography [7].

2.4. *HPLC separation of α_1- and α_2-subunits*

The separation of α-subunits has been carried out by HPLC on LiChrosorb RP-2 (4.6 × 250 mm; Merck, Darmstadt, FRG) column in 0.1 M sodium acetate/5% formic acid. The chromatographic separation was carried out in an acetonitrile gradient of 37–60%, in 47 min using a Varian LC 5000 instrument. The eluate was measured at 280 nm.

2.5. *Amino acid sequence determination*

The N-terminal amino acid sequence of 60–200 pmol of the α-subunits was determined on an Applied Biosystems pulsed liquid phase 477 A sequencer equipped with an on-line 120 A PTH analyser (ABI, Foster City, CA, USA) using microbore chromatography and a gradient solvent system.

3. RESULTS

A genomic library of 1.56×10^6 clones was obtained with an efficiency of 2.4×10^6 pfu · μg^{-1} *Chroomonas* DNA. The library was screened by hybridization using the synthetic oligomer, and a clone, OL1A1, was identified as containing sequence homology with the oligomer. DNA from OL1A1 was isolated, restricted with *Sal*1 restriction enzyme and the products analysed by gel electrophoresis. Digestion produced 5 fragments: 6.8, 2.5, ca 2.0 kb and two smaller fragments. Hybridization of the fragments with the synthetic oligomer, identified the 6.8 kb fragment as containing the homologous sequence. The 6.8 kb fragment was cloned into the *Sal* site of pUC 18 [13], giving the plasmid pAPEJJ2. A map of pAPEJJ2, derived by restriction analysis, is shown in Fig. 1.

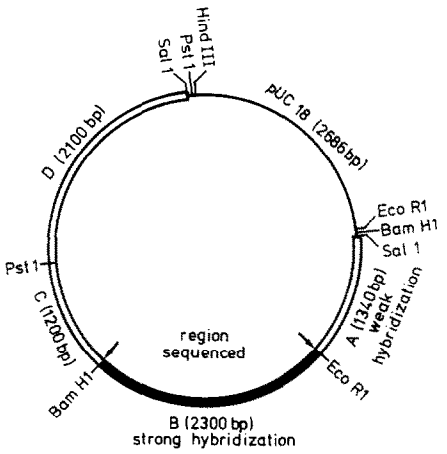

Fig. 1. Restriction map of pAPEJJ2. A 6.8 kb *Sal*1 fragment from clone OL 1A1 was subcloned into the *Sal*1 site of pUC 18 to give the plasmid pAPE JJ2.

Two areas of of pAPEJJ2 were found to hybridize with the synthetic oligomer (Fig. 1). The area designated A and delineated by *Sal* and *Eco* sites, hybridizes weakly, while area B, delineated by *Eco* and *Bam* sites, hybridizes strongly. Area B was subcloned into M13 mp8 and mp9 sequencing vectors [14] and the nucleotide sequence was determined. From the DNA sequence (Fig. 2), the amino acid sequence of a PE α-subunit precursor protein, with a size of 14 349 Da, can be derived.

Partial sequences were obtained by Edman degradation for the expressed α_1 and α_2 PE proteins of *Chroomonas*. The three *Chroomonas* sequences are compared in Fig. 3 with the α_1 and α_2 PE sequences of *Cryptomonas* [6].

4. DISCUSSION

The sequences shown in Fig. 3 provide clear evidence for the existence of 3 α-subunit genes. The N-terminal sequences of the 3 *Chroomonas* and the 2 *Cryptomonas* α-subunits show a high degree of conservation, the α_3-subunit being almost identical to that of the α_1-subunit of *Cryptomonas* in the NH_2 terminal region. None of the differences between the α_1- and α_2-subunits would account for their separation by ion exchange chromatography but this may result from variations in charged amino acids towards the carboxyl end.

```
                                                   10         20         30         40         50         60
                                          GTGGCCATGATTCCTACCTGACCTGCACACCCGCTACCAATTCCGAGGGGGACGCAGTAGCGCT

   70         80         90        100        110        120        130        140        150        160        170
GAAGGAGTTTGCCGTCGCAAAAACGGCACATTCACATAGCGTTTCGTCTTCCTTTCCCTCGTTTCAGCACCTCTGCGCGTCACCAGCGCTGCCGGGGGTGAGCTCA

     180        190        200        210        220        230        240        250        260        270
GCGCCACCCGCACAAAGCTCGAGCGCTGAGACCGACAGCCCACAACCATGATGCCCAAAACCCACAAACTTAAGCTGCCGCCGTGTGCGAGTGCCCACGTCTGCAC

        290        300        310        320        330        340        350        360        370        380
TCCGCGACCAGAATCGTGTCACATTTCGTTGTGGCCATTTTGCGTCCCCTTTGACAAGTTTTGTTCGAGGTGATGTGTTTCGAGTTGTCATTGTCAAGGTGTCCGC

 390        400        410        420        430        440        450        460        470        480        490
AACGAACCCCCATTTCTGCAAAATGTTTTAAAAACAATATGTTTTAAACGTAGGTGATAATAGAGACCTGAAACAAGTCATGGATCAGTCCGAGGCAAACTCCCCA

    500        510        520        530        540              550             560                570            580
AAAAACCCAGATCGACCCTCTGCGAAAACACCACCGACTTGTTGCC ATG TTC GCC AAG ACC CTC GCC TCC CTC GCT GTG ATC GGC TCT GCC
                                                M   F   A   K   T   L   A   S   L   A   V   I   G   S   A

   590            600             610            620             630            640            650             660
CC GCC TAC GTG CCC ATG ATG TCC ATG GAC ATG GGC CGC CGC GAG GTC GTG CAG GCT GGC GCT GCT GCC GCC GCG GTC ACC
A   A   Y   V   P   M   M   S   M   D   M   G   R   R   E   V   V   Q   A   G   A   A   A   A   A   V   T

  670            680             690             700            710             720             730            740
CC TTC CTC TCG GGT GCC CCG GCT GGC GCT GCC ATG GAC AAG TCG GCC AAG GCT CCC CTG ATC ACC ATC TTC GAC CAC CGT
P   F   L   S   G   A   P   A   G   A   A   M   D   K   S   A   K   A   P   L   I   T   I   F   D   H   R
                                        ↑
 750            760             770            780             790            800             810            820
GC TGC TCC CGC GCG CCC AAG GAG TCC ACC GGC GGC AAG GCC GGC GGC CAG GAC GAC GAG ATG ATG GTC AAG GTC GCC TCC
G   C   S   R   A   P   K   E   S   T   G   G   K   A   G   G   Q   D   D   E   M   M   V   K   V   A   S

30              840             850            860             870            880             890            900
CC AAG GTG ACC GTC TCG GAG TCT GAC GCC GCC AAG AAG CTC CAG GAG TTC ATC ACC TTC GAG AAG GGC ATC GAC GGA CCT
T   K   V   T   V   S   E   S   D   A   A   K   K   L   Q   E   F   I   T   F   E   K   G   I   D   G   P

                920            930        940        950        960        970        980        990       1000
TC ACC TCC AAG AAC TAA ATGTTCAGTCAAAAGTGTAGATGGTGACGTGGGCATGACGTGGGCTGTTGTGCTTTCGCGGGAGTTCGGGAGTGCACTCGTGT
F   T   S   K   N   *

      1020       1030       1040       1050       1060       1070       1080       1090       1100       1110
GAATAAAATCGTACCTCGCCCACACTCTGAATTTCGTGGCAACCAGGACAGGACGTGCGGTTGCCTTGGCGCGATTGTTTGGCACTGTTGGCAGGAAGGGGGTCCA

120       1130       1140       1150       1160       1170       1180       1190       1200       1210       1220
TATGGATCAGTCCAGCTCTTGGGGAGCAAGAGTTGAAAGAGAGCGCGGTTCCAACGCAAAAAAGGCGCAGACACGCAAAAAAGGGCAGACGCGAGAGTGGCTTCCA

  1230       1240       1250       1260       1270       1280       1290       1300       1310       1320       1330
AGGATGGGTGGCCGCACGATTTCAGGGCGTGCGATTCTCCAATGCTGGTGAACATCGCGCCTCAACACCCGCCGCTGCGAAAGTTGATTTTGCTCATTCAGGCTTG

     1340       1350       1360       1370       1380       1390       1400       1410       1420       1430
GCCCACACTTCCACAAATGCCGAACAAGGGCTTCGATTCAAGACCAAGTCCCGACCAAGCAATGCAGAAGGCCCAGCCTACAGGTGTGTCTGGTGTCGTGCTTCGT

        1450       1460       1470
AGCGATACGTACAGGGCGCCCAAGAAGCACAAAG...... unsequenced                           ca. 500 bases

                                    10         20         30         40         50         60         70         80
                   ...........GGCTGCCCTCGAGCGCACTGTGAGCTGACCGAGCACTGCGCATGCGAACGACCTCTGTCTGTGGGGGTCACGTCCACCTC

        90        100        110        120        130        140        150        160        170        180
GTCGTGCGGCGGAAGACGGCTCACATTCGCTGCGTGACATGGTGTCGAGAAGCGTTCGGGGATAAGCAGATTGCGACCCACGACGCAGACGGAACCGTACGAGGCA

90        200        210        220        230        240        250        260        270        280        290
AGGGCTCCGGTCACCGTGCAGATTCAGACTTAGGGAGGACAGAGCCATCGATTCGGCGATTTACAGCAACAGCACGAGCAGGAGCTGCTAGTCGAGACAGCACGTC

  300
TGAAC
```

Fig. 2. Partial nucleotide sequence of pAPEJJ2. The sequence contains a putative PE α-subunit gene with the coding sequence starting at nucleotide 542. The gene encodes a precursor PE α-subunit, the amino acid sequence of which is shown. The processed protein begins with the alanine at 697 (arrow). Possible TATA and CAAT boxes and a polyA addition signal are underlined.

```
                     -50                           -40                           -30
α3 CS540           M F A K T L A S L A V I G S A A A Y V P M M S M
                   -20                  -10
α3 CS540 D M G R R E V V Q A G A A A A A V T P F L S G A P A G A
         1                 10                  20
α3 CS540 A M D K S A K A P L I T I F D H R G C S R A P K E S T G
α1 CS540 * * * * * * * * * Q * * * * * * * * * * * * * * S E * *
α2 CS540 * * * * * * * * * V * * * * * * * * * * * * * * * - * S
α1 PE545 * * * * * * * * * Q * * * * * * * * * * * * * * * * * *
α2 PE545 * * * * * * * * * V L * * * * * * * * * * * * * * N * S

           30                  40                  50
α3 CS540 G K A G G - - Q D D E M M V K V A S T K V T V S E S D A
α1 CS540 * T * T K - - - * * Q * * * * * S Q V
α2 CS540 - * S * K S G * * * T * * * * A Q * V
α1 PE545 A * * * * - - * * * * * *
α2 PE545 - * S * * - K - * * * * L * * A Q

             60                  70
α3 CS540 A K K L Q E F I T F E K G I D G P F T S K N
```

Fig. 3. Comparison of the amino acid sequences of the α-subunits of Cryptophyte phycoerythrins. CS 540 refers to the *Chroomonas* α-subunits (this work), and PE 545 to those of *Cryptomonas* [6]. Conserved amino acids are shown as *.

Since we have not so far detected a peptide containing the sequence of α_3, we cannot conclude that multiple genes are the cause of the large number of PE isomers, and it is always possible that the α_3 gene is a pseudogene.

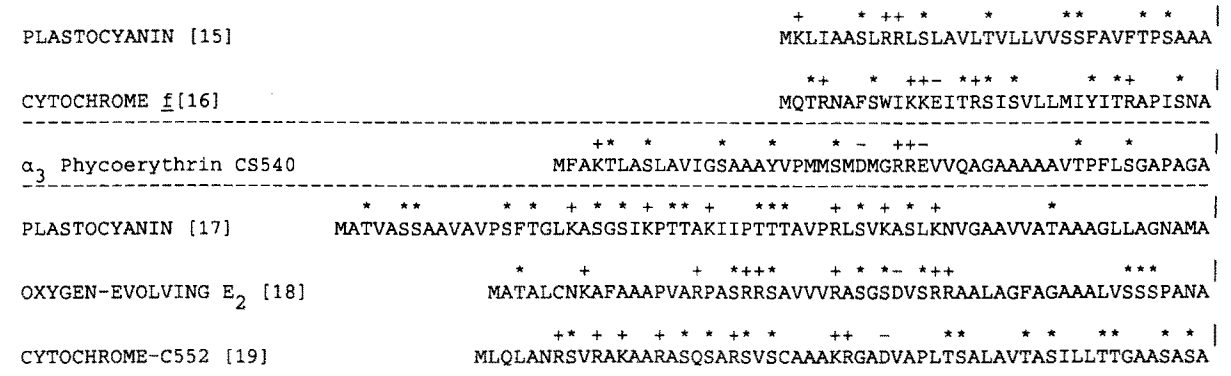

Fig. 4. Comparison of the thylakoid-lumen targetting sequences. Chloroplast-encoded or procaryotic proteins are shown above α_3 phycoerythrin CS 540, and nuclear encoded proteins below α_3 phycoethrythrin CS 540. Positive charged (+), negative charged (−) and hydrophylic (*) amino acids are indicated.

Cryptophyte phycobilins are located in the thylakoid lumen and it is therefore not surprising that the derived α_3 sequence clearly suggests a putative membrane transit N-terminal extension. All thylakoid lumen proteins, whether procaryotic, chloroplast or nuclear encoded, possess such extensions (see Fig. 4). The extensions, however, show little homology and it is not possible to decide, on this basis, whether α_3 is nuclear encoded, which would result in it having to cross 5 membranes, or chloroplast encoded, and having to cross a single membrane.

By analogy with the rhodophytes it might be anticipated that the PE genes would be chloroplast encoded in cryptophytes. However, the sequence of the cloned DNA fragment outside the α_3-subunit, does not appear to encode any known chloroplast protein. In general, chloroplast and procaryote promoter sequences are similar to each other but are not totally conserved. In Fig. 2 potential TATA and CAAT boxes and a polyA attachment signal are indicated but are not conclusive evidence that the gene is nuclear as opposed to chloroplastic. A feature of the cloned DNA fragment containing the α_3-gene is the G+C content which rises to 66% in the α_3-coding region, exceptionally high for a chloroplastic gene.

Again, by analogy with the cyanobacteria, it might be expected that a phycobilin α-subunit gene would be flanked by a β-subunit gene, but this is clearly not so in the case of the cloned gene.

Acknowledgement: This work was supported in part by the Australian Research Grants Commission (Grant AO 8516033 to R.G.H.).

REFERENCES

[1] Gantt, E. (1979) Biochemistry and Physiology of Protozoa, (2nd edn) (Levandowsky, M. and Hunter, S.H. eds), vol. II, pp. 121–137.
[2] Mörschel, E. and Wehrmeyer, W. (1975) Arch. Microbiol. 105, 153–158.
[3] Mörschel, E. and Wehrmeyer, W. (1977) Arch. Microbiol. 113, 83–89.
[4] Hiller, R.G. and Martin, C.D. (1987) Biochim. Biophys. Acta 923, 98–102.
[5] Guard-Friar, D. and MacColl, R. (1986) Photochem. Photobiol. 43, 81–85.
[6] Sidler, W., Kumpf, B., Frank, G., Suter, F., Morisset, W., Wehrmeyer, W. and Zuber, H. (1985) Hoppe-Seylers Z. Physiol. Chem. 366, 233–244.
[7] Martin, C.D. and Hiller, R.G. (1987). Biochim. Biophys. Acta 923, 88–97.
[8] Frischauf, A., Lehrach, H., Poustka, A. and Murray, N. (1983) J. Mol. Biol. 170, 827–842.
[9] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
[10] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
[11] Sanger, F., Nicklen, S., Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA. 74, 5463–5467.
[12] Fristensky, B., Lis, J. and Wu, R. (1982) Nucleic Acids Res. 10, 6451–6463.
[13] Yanisch-Peron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103–119.
[14] Messing, J. and Vieira, J. (1982) Gene 19, 269–276.
[15] Van der Plas, J., Bovy, A., Kruyt, F., de Vrieze, G., Dassen, E., Klein, B. and Weisbeeck, P. (1989) Mol. Microbiol. 3, 275–284.
[16] Willey, D.L., Auffret, A.D. and Gray, J.C. (1984) Cell 36, 555–562.
[17] Rother, C., Jansen, T., Tyagi, A., Tittgen, J. and Herrmann, R.G. (1986) Curr. Genet. 11, 171–176.
[18] Merchant, L. and Bogorad, L. (1987) J. Biol. Chem. 262, 9062–9067.
[19] Mayfield, S., Rahire, M., Frank, G., Zuber, H. and Rochaix, J.D. (1987) Proc. Natl. Acad. Sci. USA 84, 749–753.