ELSEVIER

Biochimica et Biophysica Acta 1322 (1997) 1–7

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Short sequence-paper

The orange carotenoid protein of *Synechocystis* PCC 6803 [1]

Yi Pyng Wu, David W. Krogmann *

Department of Biochemistry, Purdue University, West Lafayette, IN 47907-1153, USA

Received 2 April 1997; accepted 7 July 1997

Abstract

A water soluble protein with the carotenoid 3′-hydroxyequinenone bound to it has been purified from the cyanobacterium *Synechocystis* PCC 6803. Based on partial amino acid sequencing of the protein, oligonucleotides were synthesized and used as primers for PCR to obtain a substantial fragment of the gene. This DNA was sequenced and the sequence data and the size of the protein indicate that the protein is encoded by gene *slr 1963* in the Kazusa DNA sequence data bank containing the *Synechocystis* 6803 genome. This protein is very similar to 3′-hydroxyechinenone proteins found in several other cyanobacteria but it shows very little resemblance in its amino acid or gene sequence to other carotenoid binding proteins. The protein binds 1–2 molecules of 3′-hydroxyechinenone and is slowly cleaved by proteases in the cell extract to give a molecule of approximately half the original mass which retains the carotenoid and which shows a striking change in color. © 1997 Elsevier Science B.V.

Keywords: Carotenoid protein; Cyanobacteria; Carotenoid; Water-soluble carotenoid protein; Photoprotection

Carotenoid proteins that dissolve in water in the absence of detergents are rarely encountered. A few such proteins have been characterized from crustacians [1]. Powls and Britton reported the first water-soluble violaxanthin binding protein from a photosynthetic source – the alga *Scenedesmus obliquus* [2]. In 1981, Holt and Krogmann described an orange carotenoid protein from cyanobacteria [3]. This protein contained 3′-hydroxyequinenone and could be found in *Arthrospira* (formerly *Spirulina*) *maxima*, *Microcystis aeruginosa* and *Aphanizomenon flos-aquae*.

Most of the carotenoids of cyanobacteria are embedded in the photosynthetic membrane which is both, the site of their synthesis and the site of their major functions of light harvesting and photoprotection. Some carotenoid is exported to the periphery of the cell, especially during high intensity light exposure, where it may provide additional photoprotection. Engle et al. [4] have clear evidence for two carotenoid proteins that reside on the exterior of the cell. Other carotenoid proteins may be involved in the transport of carotenoid from the photosynthetic membrane to the exterior.

Here we identify the gene and report on the primary structure of the orange cartenoid protein from *Synechocystis* PCC 6803.

Purification of the orange carotenoid protein. The strain of *Synechocystis* PCC 6803 used in this study was a gift from Dr. Lee McIntosh of the DOE Plant Research Laboratory at Michigan State University

Abbreviations: OCP, orange carotenoid protein; RCP, red carotenoid protein; PCR, polymerase chain reaction

* Corresponding author. Fax: +1 765 494 7897; E-mail: krogmann@biochem.purdue.edu

[1] Publication No. 15377.

PII S0005-2728(97)00067-4

and it was grown in 4 l Povitsky bottles containing BG-11 medium [5] at 25°C illuminated with 75–100 mE n^{-2} s^{-1} with continuous aeration. *Escherichia coli* strains were grown in Luria Bettini broth or 2× tryptone yeast extract using standard microbial methods.

Synechocystis cells were harvested by centrifugation at 7000 × *g* for 10 min at 4°C and washed with 10 mM Tris buffer, pH 8. The packed cells were either frozen or directly resuspended in batches of 75 g per 250 ml of 20 mM Tris buffer, pH 8, 0.2% lysozyme, 5 mM EDTA, pH 8 and 100 μM phenylmethylsulfonyl fluoride. The cells were resuspended to a chlorophyll concentration of 2.5–3 mg ml^{-1} and stirred for 6 h at 4°C. DNase I (20 mg) was added and the suspension was subjected to two cycles of disruption in a French press (Aminco) at 1500 psi. All subsequent steps were carried out at 4°C. The broken cell suspension was diluted with an equal volume of 20 mM Tris buffer, pH 7.8 and centrifuged at 17 000 × *g* for 30 min to remove coarse debris. Solid ammonium sulfate was added to 20% of saturation and the preparation was centrifuged at 8000 × *g* for 20 min to remove green membrane fragments. Ammonium sulfate was added to 35% saturation and the centrifugation repeated to remove more than half of the phycobiliproteins. The ammonium sulfate concentration was raised to 75% of saturation. The centrifugation was repeated and the precipitate was resuspended in and dialysed against 10 mM Tris buffer, pH 7.8. The dialysed solution was clarified by centrifugation at 100 000 × *g* for 1 h and loaded on a 4.6 × 25 cm column of DEAE Trisacryl Plus M which had been equilibrated with the dialysis buffer. The column was washed with 40 mM NaCl in 10 mM Tris buffer, pH 7.8 to remove cytochrome *c*6 and then eluted with a linear gradient of 40–90 mM NaCl in 10 mM Tris buffer. The eluted protein fractions containing orange carotenoid protein (OCP) were pooled and concentrated by ultrafiltration on an Amicon YM10 membrane to a volume of 25 ml. The concentrate was brought to 50% ammonium sulfate saturation and was loaded onto a 1.2 × 20 cm Sephacryl 200 HR column equilibrated with 65% saturated ammonium sulfate, 250 mM NaCl and 50 mM K^+ Na^+ phosphate buffer, pH 7.0. Elution was done with a linear gradient of diminishing ammonium sulfate concentration (65–30%) with constant buffer strength and NaCl concentration. The orange colored fractions which eluted between 40 and 35% ammonium sulfate were pooled and concentrated to 1 ml for gel filtration chromatography. A 1.2 × 90 cm column of Sephacryl-200 was equilibrated with 100 mM NaCl and 5% glycerol in 100 mM imidazole buffer, pH 6.8. Fractions containing OCP were combined and dialysed against 10 mM Na^+ K^+ phosphate buffer, pH 7.0 containing 10% glycerol and 2 mM dithiothreitol and then subjected to two cycles of preparative isoelectric focusing using the BioRad Rotofor apparatus with Ampholyte, pH range 4–6 at 11 W at 8–10°C. The OCP enriched fractions were dialysed against 20 mM Tris buffer, pH 8 and concentrated to a minimal volume before loading on 8% cross-linked native PAGE. Electrophoresis was at a constant current of 20 mA. The colored gel segments were excised, finely fragmented by passing through a 1 ml syringe with a fine needle and then incubated in 3–5 ml of 20 mM Tris buffer, pH 7.8 containing 0.05% octanoyl-*N*-methyl-glucamide on a 150 rpm shaker overnight. The passively eluted OCP was diluted to 10 mM buffer strength and further purified on a 1.5 × 5 cm Whatman DE52 cellulose column using a 60–80 mM NaCl linear gradient in 10 mM imidazole buffer, pH 6.8.

The orange and red carotenoid proteins (RCP) of *M. aeruginosa* and *A. maxima* were purified to 90% of homogeneity according to the procedures of Holt and Krogmann [3] and further purified with isoelectric focusing as described above.

Characterization of the protein and its chromophore. The molecular mass of the apoOCP was determined by standard discontinous SDS-PAGE [6]; the size of the apoprotein band was determined with 14–94 kDa molecular mass standards (Bio-Rad). The molecular mass of holoOCP was determined with a size exclusion column and by matrix absorption laser desorption ion (MALDI) mass spectroscopy with a Perspective Biosystems mass spectrometer.

The amino acid sequences of the N-terminal region and of some internal protein fragments were obtained on a gas phase Applied Biosystems 473 Protein Sequencer. Two milligrams of OCP were purified by SDS-PAGE and digested with clostripain and trypsin in mixture at a 20:1 ratio as recommended by Boehringer. The digested protein fragments were separated on SDS-PAGE and blotted onto polyvinyl-

idene difluoride membrane [7] for amino acid sequence determination.

The mass of the chromophore in the OCP and the RCP from *M. aeruginosa* and *A. maxima* were determined with a Finningan 4000 mass spectrometer with Data General Nova/4 using a solid probe at 250°C and 70 eV. The carotenoid was removed from 10 μg protein by adding TFA to 0.1% and mixing with ice chilled acetone (HPLC grade, Fisher) to reach 90% followed by centrifugation for 15 min at $20\,000 \times g$ to separate the carotenoid from the white precipitate of denatured protein. The carotenoid in the acetone phase was extracted with an equal volume of hexane phase. Subsequently, the yellow hexane phase was transferred to a fresh tube, washed twice with water, and concentrated with a gentle stream of argon gas. The carotenoid was then purified by TLC on a silica gel plate, 20×20 cm (Easterman Chromagram 13179) with ethyl ether and petroleum ether mixture at ratio of 1:2 as the developing solvent.

The isolation of OCP from *Synechocystis* PCC6803 is difficult. A unique problem occurs with the *Synechocystis* PCC6803 preparation when using DEAE cellulose in the first ion exchange column. A large amount of orange pigment becomes irreversibly bound to the column and the yield of OCP is diminished. Rapid elution reduces this loss but it is still significant. The use of DEAE Trisacryl Plus-M eliminated this problem. After size exclusion chromatography, OCP was found to aggregate above concentrations greater than 1 μg per ml in less than 100 mM buffer with 10% glycerol. Non-ionic detergents such as NP-40, octanoyl-*N*-methyl-glucamide, dodecyl-β-D-maltoside or Mega 8 prevent this aggregation. The final DE-52 ion exchange chromatography gave fractions with an A_{465}/A_{275} higher than 1.6 which were pooled and were found to be homogeneous by mass spectroscopy (MS) and by amino acid sequence analyses. The p*I* is 4.7. The final yield of pure OCP (A_{465}/A_{275} above 1.89) from a 50 l cell culture was 150–250 μg protein.

In the early stages of protein purification, two distinctly colored carotenoid proteins were present in the preparation. These two proteins were not separated by ion exchange or hydrophobic interaction chromatography or by isoelectric focusing. They had

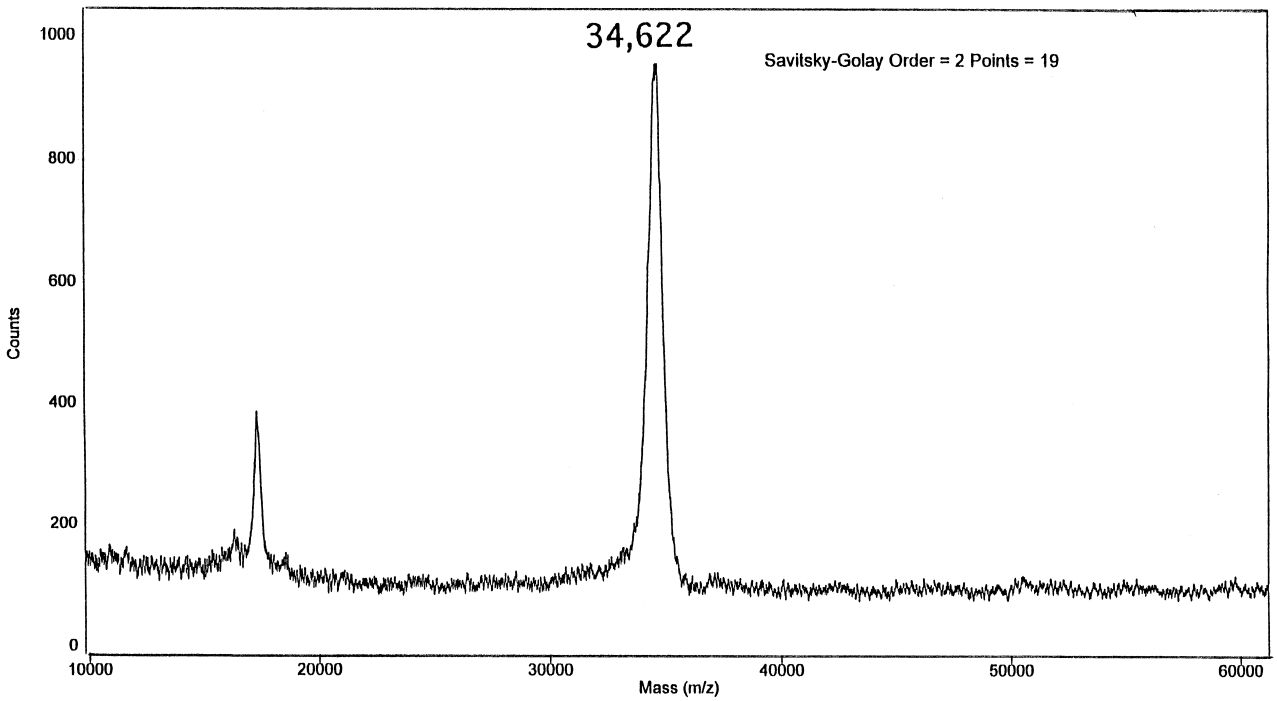

Fig. 1. Determination of the mass of OCP by matrix absorption laser desorption ion mass spectrometry. The accelerating voltage was 28 125 and the pressure was 3.35e-07. The result is the average of 34 scans.

similar p*I*s of 4.7 and 5.0. Gel filtration chromatography separated the large OCP from the RCP. The yield of RCP was usually one-tenth that of OCP. The RPC had been observed in earlier work [3] and its absorption spectrum was identical to that reported earlier with a single, broad peak which had its maximum at 505 nm. Using chromatography on a Sephacryl S-200 column with molecular mass standards, the OCP has an apparent mass of 40–41 kDa and the RCP is 19–20 kDa. When the molecular masses were measured by SDS-PAGE, the apoOCP was 35 kDa and the apoRCP was 16.5 kDa. The SDS removes the carotenoid from the protein and this accounts for some of the difference in mass between the two measurements. The OCP was used in mass spectrometric determination of molecular mass. Fig. 1 shows the result of a MALDI-MS determination. A mass of 34 622 was obtained. Virtually identical results were obtained by the electrospray technique. A MALDI-MS determination of the RCP molecular weight gave a value of 16 739.

The purified OCP was subjected to N-terminal amino acid sequencing and a sample of the protein was digested with clostripain and trypsin to yield peptide fragments for sequencing. SDS-PAGE was used to separate the fragments produced by protease digestion. The peptides were transferred to a PVDF membrane and subjected to N-terminal amino acid sequencing. The N-terminal sequences of a 22 kDa peptide and the intact OCP were the same. Peptides of 17 and 14 kDa gave identical sequences from the interior of the protein. Degenerate oligonucleotide primers were designed from these amino acid sequences.

Since all attempts at measuring the mass of the OCP holoprotein by mass spectrometry failed, an estimation of the stoichiometry of carotenoid to protein was undertaken with the expectation of at least a 10% error in the colorimetric measurements and possible loss of chromophore during extraction and purification of the carotenoids. The carotenoids from all OCPs and RCPs had the same mobility on thin layer chromatography, virtually identical visible absorption spectra, and were identified by mass spectra and NMR in samples from *M. aeruginosa* as 3′-hydroxyequinenone [3]. Samples of purified OCP from *Synechocystis* PCC6803 were used in estimates of the molarity of the protein from the UV absorbance peak and by the Bradford assay [12] and estimates of the molarity of the extracted carotenoid from the absorbance of the peak at 455 nm using the ‘‘average’’ extinction coefficient of 2500 [13]. The results indicated a ratio of 1.7 molecules of 3′-hydroxyequinone per molecule of protein. Estimates repeated with several protein samples agreed within 10%.

Small amounts of RCP were observed in purifications of *Synechocystis* PCC6803. Larger samples of RCP were available from *M. aeruginosa* and *A. maxima*. Amino acid sequencing was done on the RCPs from these two species and on the OCP from *A. maxima*. The results are compared to the N-terminal sequence of the OCP from *Synechocystis* PCC6803 (Table 1). The RCP is a fragment of the OCP. The first 15 residues from the RCP of *A. maxima* are identical to residues 16–29 of the OCP from that source. The next 20 residues of the *A. maxima* RCP are 75% identical to the corresponding residues in *Synechocystis* PCC6803 OCP. On proteolysis to RCP, the N-terminus of the OCP has lost 15 residues and the molecular weight of *Synechocystis* PCC6803 RCP indicates that in the C-terminal half of the OCP has been removed.

Gene fragment isolation. *Synechocystis* PCC6803 chromosomal DNA was purified from 2L of late log phase cells according to the method of Williams [8]. The degenerate oligonucleotide primers designed according to the N-terminal amino acid sequence determination results are 5′-CCYTAYAAYCCNAAYAC-

Table 1
N-Terminal sequences of OCPs and RCPs

	1	10	20	30	40	50
Synechocystis 6803 OCP	PFTIDSARGI	FPNTLAADVV	PATIARFSQL	NAEDQLALIW	FAYLEMGKTL	TIAAPGAAS
A. maxima OCP	PFTIDSARSI	FPETLAADVV	PATIARFKQL	·····	·····	·····
A. maxima RCP	·····	····· AADVV	PATIARFKQL	SAEDQLALIF	FAYLGMKKTI	IAAP·····
M. aeruginoza RCP	·····	····· SADVV	PATIARFNQL	NTEDQLALIY	·····	·····

3′ and 5′-GCRGTRTCYTGRGGRGG-3′ (Y = T,C; N = A,T,G,C; R = A,C). The partial gene sequence of the OCP was obtained with polymerase chain reaction (PCR) against chromosomal DNA. The PCR was performed with 0.1 μg of chromosomal DNA, 50 pmole of each primer, 0.2 mM deoxyribonucleotide triphosphate and 0.5 units of Vent DNA polymerase in a reaction volume of 50 μl. PCR cycling conditions were programmed at 50 s at 93°C, 45 s at 48°C, 1 min at 72°C for 30 cycles followed by 10 min at 72°C (Perkin-Elmer Cetus, GeneAmp 9600). The amplified gene fragment of OCP was cloned into the TA system (Invitrogen) and sequenced with Sequenase version 2 according to the manufacturer's instructions.

The following reagents and supplies were obtained from the indicated sources: lysozyme, U S Biochemicals; DNase I, Sigma; octonoyl-*N*-ethyl-glucamide (Mega-8), Bobingen; PVDF membrane, Applied Biosystems; DEAE-Trisacryl Plus M, Sigma; Sephacryl-200, Pharmacia; clostripain and trypsin, Boehringer-Mannheim; Vent polymerase, New England Biolabs; Sequenase version 2, Amersham; Ampholyte, BioRad. The Gene Runner program is a product of Hastings, Box 567, Hastings New York, 10706. The programs BLAST, FASTA, and MSA are from the Wisconsin Software Package, Version 8, September 1994, The Genetics Computer Group, 575 Science Drive, Madison, Wisconsin 53711.

The polymerase chain reaction was used to generate a 519 base pair segment of DNA which encodes the amino acid sequence bounded by the N-terminal and the internal peptide sequences. This fragment was sequenced and was found to encode the N-terminal amino acid sequence of the OCP. The *Synechocystis* PCC6803 genome in the Kazusa Institute database (http://www.kazusa.or.jp) was searched and yielded the gene *slr 1963* – residues 1767270–1768233 which encodes a 317 amino acid sequence. This sequence begins with a methionine which is missing from the N-terminal sequence of the protein. Our sequencing of the DNA from the region encoding the first 171 residues shows five single base changes that would alter the identity of five hydrophobic residues – $G_{73} \rightarrow A$, $T_{75} \rightarrow G$, $L_{121} \rightarrow F$, $V_{165} \rightarrow A$ and $I_{169} \rightarrow G$. There is no evidence of a leader sequence preceding the OCP.

Structure of OCP and RCP. Fig. 2 shows the amino acid sequence of the OCP as translated from gene *slr 1963* in the Kazusa Institute database. This

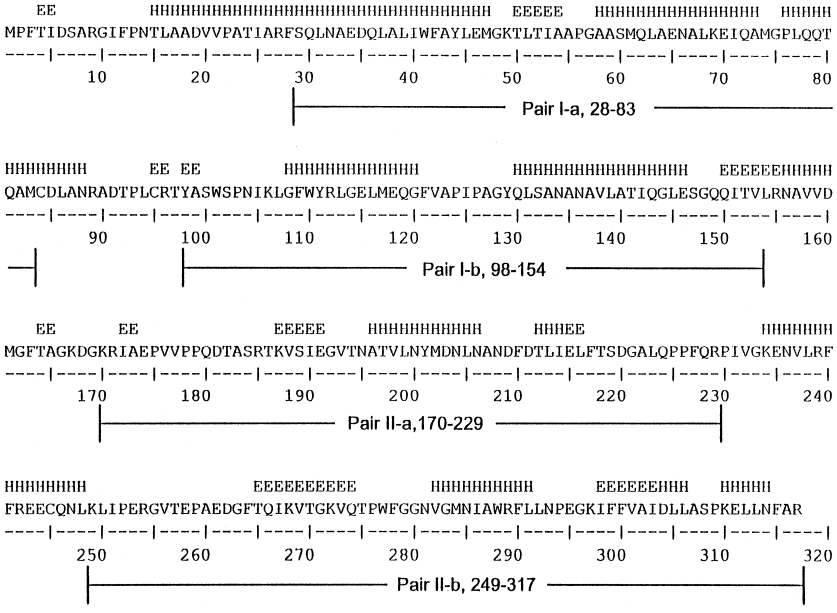

Fig. 2. The amino acid sequence of the OCP as deduced from the sequence of the gene *slr 1963* of the *Synechocystis* PCC 6803 genome in the Kazusa Institute database. The helical regions are designated by H and the β-sheet regions by E above the sequence. The pairs of regions with similar sequences are indicated by brackets below the sequence.

sequence has a mass averaged for natural isotope composition of 34 660. This is quite close to the value of 34 622 we obtained by mass spectrometry which, at its best, gives values accurate to 0.1% of the total mass. Given that our protein does not have the N-terminal methionine and may have some amino acid substitutions (four of the first 170 residues) due to strain differences, the similarity is quite close. The observed mass suggests that the carotenoid chromophores are lost in the mass spectrometry. A single 3′-hydroxyequinenone molecule has a molecular weight of 566. A variety of ways to introduce the holoprotein into mass spectrometers gave the same value which is closer to the calculated apoprotein mass.

The Gene Runner software program calculates an isoelectric point of 4.76 for the polypeptide sequence deduced from the gene while the measured isoelectric point is 4.7. This program identifies an ATP/GTP binding site at residues 43–50 and possible phosphorylation sites at Ser 7, Tyr 211, Tyr 270 and Ser 309. A possible glycosylation site is identified at Asn 195. However, the electrospray mass spectrometer measurement gave a single peak at the mass of the apoprotein. In our experience, this method has been sensitive to small amounts of oxidized methionine containing protein, so we would judge that adducts like phosphate are not present in our sample. The Gene Runner program was used to generate a hydropathy plot which revealed two regions of lipophilicity – residue 8–62 and 118–165 – in the N-terminal half of the protein and a smaller lipophilic region – 280–311 – near the C-terminus.

The programs for predicting protein structure, TOPOLOGY [9] and PREDICT PROTEIN [10] indicate that the OCP is a globular protein dominated by helix and random coil structure. According to the self-repeating sequence analysis program REPRO [11], OCP shows two pairs of repeating sequences indicated in Fig. 2. On subjecting the OCP gene sequence to the similarity comparison programs BLAST and FASTA, the *slr 1963* sequence showed less than 30% similarity to anything in GenBank.

Discussion. Several cyanobacteria contain a water soluble carotenoid protein with a spectrum characteristic of bound hydroxyequinenone. Purification of this protein is made difficult by its similarity in isoelectric point to the phycobiliproteins which constitute a very large percentage of the total protein in the cell. With *Synechocystis* PCC6803, the purification is further complicated by the tendency of the OCP in crude extracts to become immobilized on DEAE. The immobilized carotenoid could not be released by treatment with 5 M NaCl, 5% Triton X100, 10% SDS, 1 N HCl, 1 N NaOH, organic solvents or by protease digestion. The increased binding of OCP with prolonged exposure of crude extract to DEAE and the absence of binding with purified OCP suggests that some agent in the dialysed crude extract is responsible for binding. This binding might relate to a function of carotenoids which finds them embedded in the cell wall in a precise orientation [14] where they presumably protect the cell from excess light. Carotenoids are synthesized in the photosynthetic membrane and they may associate with a protein to migrate across the cytosol to reach the cell periphery where they are inserted into the cell wall.

The removal of parts of the OCP polypeptide cause a shift in the absorption spectrum which signals a substantial change in the environment of the carotenoid. The RCP contains two of the three major hydrophobic domains of the OCP. The two large hydrophobic domains in the N-terminal half of the protein are the likely binding site of carotenoid(s) and the third hydrophobic domain near the C-terminus might provide additional shielding of the carotenoids(s) from the aqueous phase as suggested by the large red shift in the spectrum of the RCP which has lost the C-terminal half of the protein. The OCP of *A. maxima* has been crystallized [15] and its three-dimensional X-ray structure will be sought.

The authors are grateful to Ms. C. Bonham and Dr. Karl Wood of the Purdue Mass Spectrometer Center and to Ms. M. Bauer of the Purdue Laboratory for Molecular Structure for their help. Financial support for this research was provided by grant no. 89-3729 from the Department of Agriculture and grant 917600 from the National Science Foundation.

References

[1] P.F. Zagalsky, E.E. Eliopoulus, J.B.C. Findlay, Comp. Biochem. Physiol. 97B (1990) 1–18.
[2] R. Powls, G. Britton, Biochim. Biophys. Acta 453 (1976) 270–276.

[3] T.K. Holt, D.W. Krogmann, Biochim. Biophys. Acta 637 (1981) 408–414.
[4] J.M. Engle, W. Burkhart, D.W. Sherman, G.S. Bullerjahn, Arch. Microbiol. 155 (1991) 453–458.
[5] M.B. Allen, J. Phycol. 4 (1968) 1–3.
[6] U.K. Laemmli, Nature 227 (1970) 680–685.
[7] P. Matshidara, J. Biol. Chem. 262 (1987) 10035–10038.
[8] J.G.K. Williams, Methods Enzymol. 167 (1988) 766–778.
[9] B. Rost, Methods Enzymol. 266 (1996) 525–539.
[10] B. Rost, C. Sander, J. Mol. Biol. 232 (1993) 584–599.
[11] J. Heringa, P. Argos, Proteins Struct. Funct. Genet. 17 (1993) 391–410.
[12] M.M. Bradford, Anal. Biochem. 72 (1976) 248–252.
[13] S. Liaaen-Jensen, A. Jensen, Methods Enzymol. 23 (1971) 586–602.
[14] U.J. Jurgens, W. Mantele, Biochim. Biophys. Acta 1067 (1991) 208–212.
[15] C.A. Kerfeld, Y.P. Wu, C. Chan, D.W. Krogmann, T.O. Yeats, Acta Crystallogr. (1997) in press.