Biochem. J. (1991) 276, 47–52 (Printed in Great Britain)

Amino acid sequences of *Euglena viridis* ferredoxin and cytochromes *c*

R. P. AMBLER,* M. D. KAMEN,† R. G. BARTSCH‡ and T. E. MEYER‡

*Department of Molecular Biology, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, Scotland, U.K.,
†1390 Plaza Pacifica, Montecito, CA 93108, U.S.A. and ‡Department of Biochemistry, University of Arizona, Tucson, AZ 85721, U.S.A.

The Order Euglenida comprises many species and perhaps 40 genera, but almost all biochemical and genetic studies have been limited to a single species, *Euglena gracilis*, because of its ease of growth in the laboratory. Sequence studies of chloroplast and mitochondrial proteins from *E. gracilis* show that they have diverged widely from other eukaryotic lines. In the present paper we report the sequences of three proteins from another euglenoid, *Euglena viridis*, using material isolated from a natural bloom. The mitochondrial cytochrome *c* shows more than 90% sequence identity with that from *E. gracilis*, and contains the same characteristic features. The chloroplast cytochrome c_6 has diverged to a greater extent and shows only 77% identity. The chloroplast ferredoxin from *E. viridis* is similar in sequence to those of cyanobacteria and algal chloroplasts, with sequence identities of up to 75%. Details of the purification, analysis and sequence determination experiments on the peptides have been deposited as Supplementary Publication SUP 50163 (32 pages) at the British Library Document Supply Centre, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in Biochem. J. (1991) **273**, 5.

INTRODUCTION

Euglena gracilis is a green flagellate protist that has been important for many years as an organism that confuses clear distinctions between the eukaryotic kingdoms. It is photosynthetic and possesses chloroplasts, but morphologically it is a typical flagellate, and very similar colourless forms like *Astasia* are well characterized. The Order Euglenida comprises many species, divided into perhaps 40 genera (Leedale, 1967). Different forms can readily be isolated from many water environments, but very few will grow in axenic culture, and virtually all biochemical and physiological studies have been confined to *E. gracilis*, which is readily grown in the laboratory.

Pettigrew *et al.* (1975) isolated the mitochondrial cytochrome *c* and the chloroplast cytochrome c_6 (Pettigrew, 1974) from *E. gracilis*, and determined their amino acid sequences. They showed that these proteins were homologous with the functionally equivalent cytochromes from other eukaryotes, but had diverged considerably from the other eukaryotic lines. Cytochromes with similar sequences to those of mitochondrial cytochromes *c* have been found in many Gram-negative bacteria, particularly in the Rhodospirillaceae, the purple non-sulphur phototrophic bacteria (Ambler *et al.*, 1987). Dayhoff (1983) has interpreted these similarities between particular prokaryotic and eukaryotic sequences as suggesting that mitochondria have arisen from bacteria independently in several different lines. In contrast, Meyer *et al.* (1986) argue that these similarities, which only amount to around 60% identity at the most, are expected from the balance between divergence and convergent back-mutation for proteins that cannot diverge further because of structural and functional constraints.

Features in the *E. gracilis* cytochrome sequences are so different from the corresponding structures in homologous cytochromes from other organisms that we have wanted to look at the proteins from another euglenoid to confirm that they are present in other representatives of the line. This has not been easy to do because of the difficulty in culturing euglenoids other than *E. gracilis* in the laboratory. However, we have been able to take advantage of a natural bloom to study the proteins from a marine euglenoid.

In 1980 the sewage processing plant in Tijuana, Mexico, apparently failed, and raw sewage was discharged into the Pacific Ocean. Soon afterwards a green bloom was noticed on the sand at low tide on the beach at La Jolla, about 25 miles to the north. The bloom persisted for several years and we cannot be certain that it was not present before the sewage spill. Whether or not there is a cause-and-effect relationship between these events, the bloom was as dense as those of *Microcystis* that are often found in Nature. The bloom was most pronounced when the sun was obscured, and faded in strong illumination. Microscopic examination of the surface layer showed the presence of an apparently pure culture of a euglenoid protist, and large amounts of cells were obtained by scraping off the surface sand and washing it with sea-water. The organism has been identified as *Euglena viridis* by Professor F. T. Haxo of the Department of Marine Biology, University of California at San Diego (personal communication), and called strain LJ-1. It has not yet been found possible to grow this organism in the laboratory.

We have isolated and characterized the mitochondrial cytochrome *c* and the chloroplast cytochrome c_6 and ferredoxin from this organism, and determined their amino acid sequences. We compare these sequences with those from algae and from other protists. This is the first report of the sequence of a ferredoxin from a euglenoid.

EXPERIMENTAL

Preparation of organism

Euglena viridis cells were collected at La Jolla Shores Beach, in front of the Scripps Institute of Oceanography, La Jolla, CA, U.S.A. Cells were scraped from the top surface of the sand at low tide on 7 October 1981, when the sky was overcast, and where a green stain on the sand was very apparent. The mixture of sand and cells was gently swirled in sea-water, and the cell suspension was decanted after the sand settled out. Cells were harvested in a Sharples continuous-flow centrifuge. Approx. 1400 g wet wt. of

* To whom correspondence should be addressed.

Table 1. Amino acid compositions of *E. viridis* ferredoxin and cytochromes *c*

The 'Analysis' value is the average of the accepted amounts from three to five independent determinations, except for tryptophan (derived after hydrolysis with 3 M-mercaptoethanesulphonic acid) and cysteine (analysed as cysteic acid after performic acid oxidation), which were only single determinations. The 'Sequence' values are derived from the results given in Figs. 1–3.

	Amino acid composition (mol of residue/ml)					
	Chloroplast ferredoxin		Chloroplast cytochrome c_6		Mitochondrial cytochrome *c*	
	Analysis	Sequence	Analysis	Sequence	Analysis	Sequence
Glycine	7.4	7	9.6	11	9.7	11
Alanine	7.2	7	7.1	7	11.6	11
Valine	3.9	4	6.4	7	4.4	5
Leucine	5.7	6	3.3	3	4.4	5
Isoleucine	5.6	6	4.4	5	3.9	4
Serine	6.3	7	6.8	9	6.8	7
Threonine	9.1	10	2.3	2	5.1	5
Aspartic acid	13.2	12	9.1	3	10.4	4
Asparagine		2		6		7
Glutamic acid	11.5	6	12.3	9	7.8	5
Glutamine		5		3		2
Phenylalanine	2.6	3	1.3	1	3.0	3
Tyrosine	3.7	4	5.1	5	4.8	6
Tryptophan	< 0.08	0	1.3	2	0.8	1
Cysteine	6.1	7	1.1	2	0.7	1
Methionine	1.1	1	1.8	2	1.9	2
Proline	4.1	4	3.5	4	3.1	4
Lysine	4.1	4	3.3	3	10.2	12
ε-Trimethyl-lysine					0.9	1
Histidine	0.9	1	1.5	2	1.0	1
Arginine	0.9	1	1.1	1	4.5	4
Total		96		87		102

cells was recovered from about 37 m^2 (400 ft^2) of beach. Bacterial and algal contamination appeared to be less than a few percent, as based on microscopic examination.

Purification of proteins

The cells were suspended in 3 litres of 0.1 M-Tris/HCl buffer, pH 7.5, and sonicated for five 2 min intervals in 500 ml portions in a stainless-steel beaker on ice with a Heat Systems Ultrasonics model W-220F cell disrupter with a 13 mm (0.5 in) tip. Almost all membranous material was removed by centrifugation at 145000 *g* for 4 h. The slightly turbid orange supernatant was passed through a column (15 cm × 7 cm diam.) of Whatman DE-52 DEAE-cellulose equilibrated with 0.1 M-Tris/HCl buffer, pH 7.5, and both ferredoxin and what proved to be the cytochrome c_6 were adsorbed. The unadsorbed solution was desalted by gel filtration through Sephadex G-25 into 10 mM-Tris/HCl buffer, pH 7.5, and passed through another similar DEAE-cellulose column, which adsorbed some further cytochrome, but left some coloured material still unadsorbed. The DEAE-cellulose columns were eluted with 20 mM-Tris/HCl buffer, pH 7.5, containing 0.5 M-NaCl, the coloured material being collected in as small a volume as possible and then subjected to gel filtration through a Sephadex G-75 (fine grade) column (100 cm × 8 cm diam.) in 20 mM-Tris/HCl buffer, pH 7.5, containing 0.1 M-NaCl. The main coloured band was of low M_r, and as judged spectrally contained both cytochrome and ferredoxin. This mixture of proteins was chromatographed on a column (15 cm × 7 cm diam.) of Whatman DE-52 DEAE-cellulose equilibrated with 20 mM-Tris/HCl buffer, pH 7.5, and eluted with a linear gradient of 0.08–0.24 M-NaCl in the same buffer. The cytochrome c_6 (about 30 μmol) was eluted at about 0.16 M-NaCl, and was judged to be pure after $(NH_4)_2SO_4$ precipitation. The 60–80%-saturation fraction contained about 25 μmol of cytochrome with a purity index (A_{280}/A_{416}) of 0.20.

About 36 μmol of ferredoxin was eluted from the DEAE-cellulose column at between 0.24 M- and 0.32 M-NaCl, and was then chromatographed on a hydroxyapatite column (5 cm × 8 cm diam.) equilibrated with 5 mM-potassium phosphate buffer, pH 7.0, containing 0.1 M-NaCl and developed with a linear gradient of 0.01–0.1 M-potassium phosphate buffer, pH 7.0. About 22 μmol of ferredoxin was eluted at about 0.02 M-phosphate, and was followed by large quantities of nucleic acid. The protein was subjected to $(NH_4)_2SO_4$ fractionation, and the 70–90%-saturation fraction was judged to be pure ferredoxin, yielding 18 μmol with a purity index (A_{280}/A_{460}) of 1.85.

The material that had not adsorbed on the initial DEAE-cellulose columns was equilibrated with 1 mM-potassium phosphate buffer, pH 5.0, by gel filtration through Sephadex G-25, and then adsorbed on a column (4 cm × 4 cm diam.) of Whatman CM-52 CM-cellulose equilibrated with the same buffer, and eluted with 0.5 M-NaCl in the same buffer. This fraction was further separated on a Sephadex G-75 (fine grade) column (100 cm × 8 cm diam.) equilibrated with 20 mM-potassium phosphate buffer, pH 7.0, containing 0.1 M-NaCl. About 3.5 μmol of the mitochondrial cytochrome *c* (Pettigrew *et al.*, 1975) separated from a larger-M_r flavoprotein. This cytochrome was then chromatographed on a column (4 cm × 4 cm diam.) of CM-52 CM-cellulose equilibrated with 15 mM-potassium phosphate

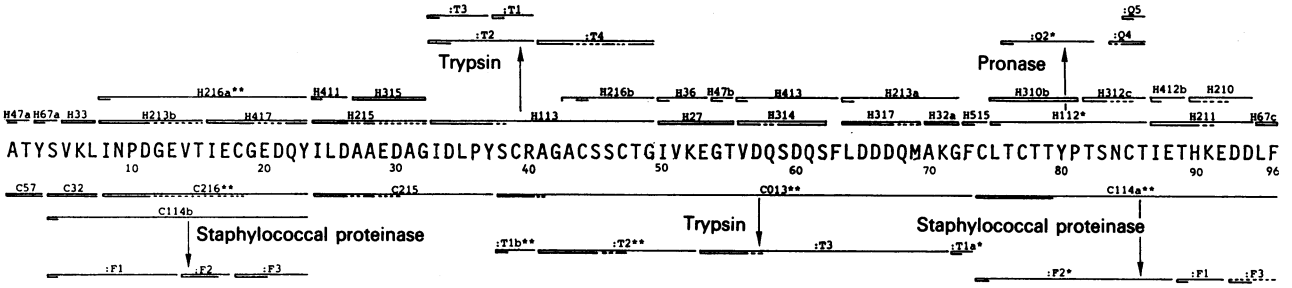

Fig. 1. Amino acid sequence of *D. viridis* LJ-1 ferredoxin

Peptides derived by digestion with thermolysin (H) are shown above the sequence, and by digestion with chymotrypsin (C) below the sequence. Peptides from sub-digests are labelled with a second letter indicating the secondary method of cleavage (T, trypsin; Q, Pronase; F, staphylococcal proteinase). Continuous lines indicate quantitative amino acid analyses, substandard if marked*, and particularly bad if marked **. Dashed lines indicate peptides that were recognized as being present, but were not isolated in a pure state. The peptide lines are doubled where the sequence has been determined by the dansyl/phenyl isothiocyanate method, with the lower line broken at residues where the identification was substandard. *C*-Terminal residues identified as free amino acids after removal of the remainder by phenyl isothiocyanate degradation are indicated by a vertical line joining the double lines at the end of the peptide.

buffer, pH 7.0, and eluted with a linear gradient of 15–60 mM-potassium phosphate buffer, pH 7.0. The cytochrome was eluted at about 46 mM-phosphate. The yield was about 3.0 μmol with a purity index (A_{280}/A_{414}) of 0.26. This material was found to be not completely pure (see below), but nevertheless was used for sequence determination. This cytochrome had the same red-shifted absorption peaks, as found for *E. gracilis* mitochondrial cytochrome *c* (Pettigrew *et al.*, 1975), and the haem is presumably bound by a single thioether bond.

In addition to these proteins, about 3 μmol of a high-spin protohaem protein was weakly adsorbed on the first DEAE-cellulose columns, and appeared to have an M_r of about 70000 from analysis on Sephadex G-75. It was not characterized further.

Amino acid sequence determination

The sequences were determined by cleaving the proteins with several different enzymic or chemical methods, fractionating the peptides formed by gel filtration and high-voltage paper electrophoresis, analysing quantitatively for amino acid composition, and sequencing by the dansyl/phenyl isothiocyanate method. The methods used have been described previously (Ambler & Wynn, 1973; Ambler *et al.*, 1979, 1984). The *N*-terminal sequence of the cytochrome c_6 was determined by using automatic sequencers (Applied Biosystems models 477 and 477A). Amide groups have largely been assigned from peptide electrophoretic mobilities and exopeptidase analysis.

RESULTS

Properties of the protein preparations

The chloroplast ferredoxin and cytochrome c_6 were obtained in large quantities. The spectral purity of the ferredoxin was satisfactory, and during sequence determination no indications were given of the presence of any other protein, no peptides being found that did not fit the sequence shown in Fig. 1. The spectral purity of the cytochrome c_6 was similar to that reported by Pettigrew (1974) for the *E. gracilis* protein, but the amino acid composition of the protein preparation (Table 1) showed some divergence from that deduced for the protein. Thus the serine and histidine values were low, and the phenylalanine value was higher than would be expected were the preparation to have consisted completely of material of the deduced sequence. Nevertheless, no peptides were detected that did not fit the sequence shown in Fig. 2, nor did automated sequencing give any suggestion of significant contamination.

The yield of the basic mitochondrial cytochrome *c* was much lower than those of the chloroplast proteins, with the amount of material produced being judged barely sufficient for sequence determination by the methods that we had available. The spectrophotometric purity index was adequate ($A_{280}/A_{414} = 0.26$), which compares with 0.22 for horse cytochrome *c* (Margoliash & Frohwirt, 1959), and so we resolved to use the material at this stage to compare with the protein from *E. gracilis*, rather than purifying it further with the accompanying loss in yield. During the sequence analysis some peptides were characterized (and listed in Supplementary Publication SUP 50163) that could not be reconciled with the proposed sequence (Fig. 3), but they were recovered in much lower yield than most of the peptides shown in the Figure. Amino acid analysis of the preparation (Table 1) was in quite good agreement with the values subsequently deduced from the sequence.

Amino acid sequences

The evidence for the proposed amino acid sequences of the *E. viridis* ferredoxin, cytochrome *c*-558 and cytochrome c_6 are summarized in Figs. 1, 2 and 3. Details of the purification, analysis and sequence determination experiments on the peptides are given in Supplementary Publication SUP 50163.

DISCUSSION

We have determined the sequences of three electron-transport proteins from *E. viridis*. Two cytochromes from *E. gracilis* have already been sequenced (Pettigrew, 1974; Pettigrew *et al.*, 1975), but this is the first report of a ferredoxin sequence from a euglenoid.

Sequence of the ferredoxin

In our hands, the determination of amino acid sequences of non-haem iron proteins through enzymic digestions is less satisfactory than for cytochromes *c*, as we do not have a really effective method of making the protein digestible because the iron is so firmly held to the native protein. On this occasion performic acid oxidation was used, as we were satisfied that the protein did not contain tryptophan. This amino acid was not detected after hydrolysis of the protein with mercaptoethanesulphonic acid, and no strongly fluorescent peptides were seen in digests of the oxidized protein.

The sequence was deduced from the properties of peptides from a chymotryptic and a thermolysin digest (Fig. 1). To obtain high yield the larger chymotryptic-digest peptides were only subjected to a single step of electrophoretic purification, with the result that their analyses showed some contamination with adjoining peptides. The thermolysin-digest peptides and those

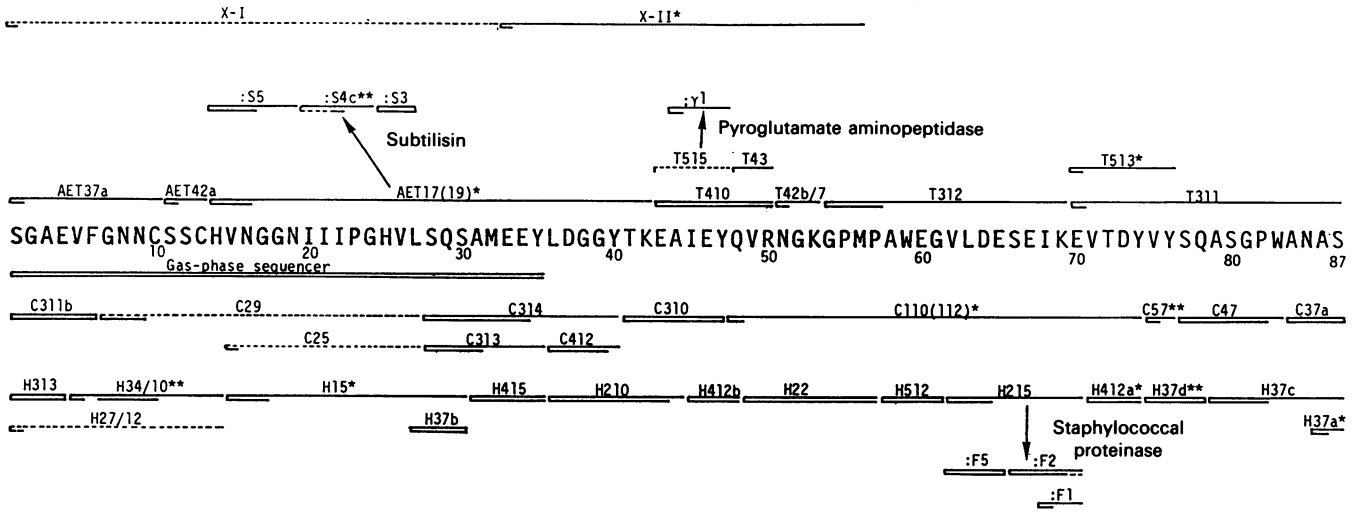

Fig. 2. Amino acid sequence of *E. viridis* LJ-1 cytochrome c_6

The abbreviations and conventions used are as explained in the legend to Fig. 1, with the additions of: X, CNBr; S, subtilisin; γ, pyroglutamate aminopeptidase. Peptides marked AE were formed by tryptic cleavage at *S*-β-aminoethylcysteine. The *N*-terminal sequence obtained by automatic sequencer degradation is shown.

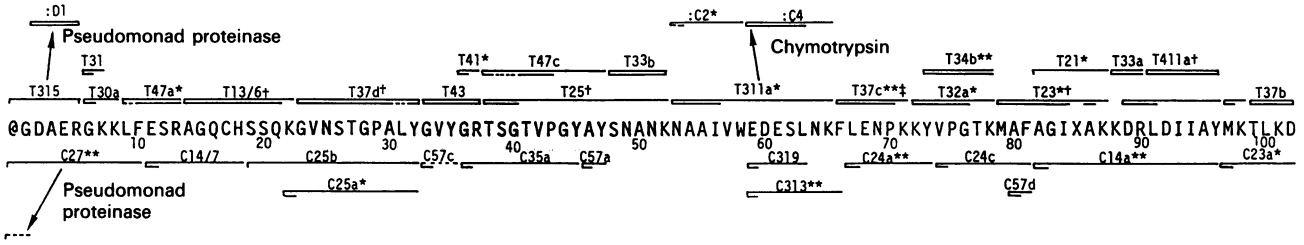

Fig. 3. Amino acid sequence of *E. viridis* LJ-1 mitochondrial cytochrome *c*

The abbreviations and conventions used are explained in the caption to Fig. 1, with the addition of: D, pseudomonad proteinase; @, α-*N*-acetyl (at *N*-terminus). Peptides marked † were examined by carboxypeptidase A digestion, and ‡ by aminopeptidase M digestion.

```
                1         2         3         4         5         6         7         8         9         9
                0         0         0         0         0         0         0         0         0         6
        *** * *       *   *  *** *** *  ** ******** *|*    *  *** ****** *   *  ***  ** *|*   *  * * **
(1)     ATYSVKLIN-PDG-EVTIECGEDQYILDAAEDAGIDLPYSCRAGACSSCTGIVKEGTVDQSDQSFLDDDQMAKGFCLTCTTYPTSNCTIETHKEDDLF Euglena viridis

(2)     ATYKVTLKT-PSG-DQTIECPDDTYILDAAEEAGLDLPYSCRAGACSSCAGKVEAGTVDQSDQSFLDDSQMDGGFVLTCVAYPTSDCTIATHKEEDLF Scenedesmus

(3)     ATYSVTLVNEEKNINAVIKCPDDQFILDAAEEQGIELPYSCRAGACSTCAGKVLSGTIDQSEQSFLDDDQMGAGFLLTCVAYPTSDCKVQTHAEDDLY Bumilleriopsis

(4)     ATYKVTLVR-PDGSETTIDVPEDEYILDVAEEQGLDLPFSCRAGACSTCAGKLLEGEVDQSDQSFLDDDQIEKGFVLTCVAYPRSDCKILTNQEEELY Synechococcus sp.

(5)     ATYKVTLISEAEGINETIDCDDDTYILDAAEEAGLDLPYSCRAGACSTCAGKITSGSIDQSDQSFLDDDQIEAGYVLTCVAYPTSDCTIQTHQEEGLY Spirulina maxima
```

Fig. 4. Alignment of selected chloroplast-type (2Fe-2S) ferredoxins

The sequences shown are: (1) *Euglena viridis* (the present investigation); (2) *Scenedesmus quadricauda* (green alga); (3) *Bumilleriopsis filiformis* (golden-yellow alga); (3) *Synechococcus* sp. (cyanobacterium); (4) *Spirulina maxima* (cyanobacterium). Sequences (2)–(5) are taken from Matsubara & Hase (1983). Positions marked * are identical in all the five sequences shown, while those marked ¦ (48 and 84) are identical in sequences (2)–(5), although different in sequence (1). Position 48 is threonine, as in *Euglena*, in *Phytolacca* iso-I and *Gleichenia japonica*, and position 84 is also asparagine in *Cyanidium caldarium* and *Nostoc* strain MAC iso-II.

from secondary digestion of chymotryptic-digest fragments were obtained satisfactorily pure, and overlaps were established throughout the sequence except at peptide bond 23–24. The sequence contains a high proportion of amides and acidic residues, but assignments could be made without ambiguity from electrophoretic mobilities. There are two positions shown in Fig. 4 where the four comparison sequences have identical residues, but at which the *Euglena* sequence is different. However, Thr-48 (in the numbering of Fig. 1) is also found in the ferredoxins from the dicotyledon *Phytolacca* and a staghorn fern *Gleichenia* (see Matsubara & Hase, 1983), and Asn-84 is also found in *Cyanidium caldarium* and a strain of the cyanobacterium *Nostoc*.

Sequences of the cytochromes *c*

For the two cytochromes *c*, sequences of the proteins from *E. gracilis* were available (Pettigrew, 1974; Pettigrew *et al.*, 1975), and so peptides could be positioned by comparison as they were isolated and characterized. Nevertheless, our eventual sequences

Table 2. Identity matrix for 2Fe-2S ferredoxins

The first number is the number of identities in the 96–98 residues compared; the second (in parentheses) is the number of insertion or deletion events needed in reaching the alignment shown in Fig. 4. The sequences of proteins (2)–(4) are taken from Matsubara & Hase (1983).

	(1)	(2)	(3)	(4)	(5)
(1) *Euglena viridis*	100	71 (0)	63 (2)	63 (1)	63 (2)
(2) *Scenedesmus quadricauda*		100	66 (2)	68 (1)	72 (2)
(3) *Bumilleriopsis filiformis*			100	62 (1)	70 (0)
(4) *Synechococcus* sp.				100	71 (1)
(5) *Spirulina maxima*					100

showed that the cytochrome c_6 differed in 20/87 residues (23%) and that the mitochondrial cytochrome differed in 9/102 residues (9%).

The main difficulty in determining the sequence of the cytochrome c_6 was in the region on the *C*-terminal side of the haem-attachment site (residues 15–27), as it was found difficult to isolate suitable pure peptides in good yield with any of the proteolytic enzymes available. The region was also made more difficult to study by the presence of a labile -Asn-Gly- sequence and three consecutive isoleucine residues. Automatic sequencer degradations of the whole protein gave satisfactory results as far as residue 35, and were completely compatible with the peptide evidence. Satisfactory overlaps between tryptic-digest, chymotryptic-digest and thermolysin-digest peptides were obtained for the remainder of the sequence (Fig. 2).

The mitochondrial cytochrome *c* was produced in smaller amounts than the chloroplast protein under natural growth conditions, and so the amount available for sequencing was smaller. The preparation that we used was also not completely pure, and so we were confused by high-yield peptides from a contaminant protein. None of them was recovered with a yield greater than 6%, whereas more than half the tryptic-digest peptides shown in the sequence in Fig. 3 were obtained with yields greater than 20%. Information about these contaminant peptides is given in Supplementary Publication SUP 50163. If the assumption is made that the protein is closely similar to that from *E. gracilis* (Pettigrew *et al.*, 1975), direct evidence was obtained for the sequence shown in Fig. 3, with the *E. viridis* protein differing in nine positions.

Pseudomonad proteinase split *N*-terminal peptides from the protein before Asp-2, and formed a compound with the same electrophoretic mobility as synthetic acetylglycine, yielded glycine on hydrolysis and did not give a colour reaction with ninhydrin on paper. Residue 85 is basic, and has properties expected for ϵ-trimethyl-lysine, but has not been definitively shown to be this amino acid.

The *E. viridis* protein shares two unusual features with that from *E. gracilis* (Pettigrew *et al.*, 1975). Compared with other mitochondrial cytochromes *c* there is a single residue deletion that is best located between residues 24 and 25 (Fig. 3) and that probably corresponds to Lys-25 in the animal sequence. There is also the substitution of Ala-14 for one of the haem-binding cysteine residues, a change responsible for the unusual visible-absorption spectrum of the protein.

Comparison with homologous proteins: the ferredoxin

An alignment of chloroplast and cyanobacterial 2Fe-2S ferredoxins is given in Fig. 4. The proteins from higher plants, algae and cyanobacteria are similar in sequence, with an average of about 70% identity. This value compares with 59% identity across the kingdoms for the closest known cytochromes *c* (Ambler *et al.*, 1987), or the values around 50% for 'highly conserved proteins' such as glyceraldehyde-3-phosphate dehydrogenase or the *groEL* 'chaperonin'. The chloroplast-type ferredoxin from *E. viridis* is not particularly close in sequence to the proteins from any other class of photosynthetic organism (Fig. 4 and Table 2), and it can be aligned with the chloroplast ferredoxins from higher and lower plants and green algae without the need for any internal insertions or deletions.

Comparisons with chloroplast cytochrome c_6

This protein appears to be more variable than either the mitochondrial cytochrome *c* or the chloroplast ferredoxin. The two *Euglena* sequences show only 76% identity, compared with 91% for the corresponding mitochondrial cytochrome *c*, and this probably reflects a smaller number of residues that need to be conserved to maintain the structure and function. In direct sequence identity, the *Euglena* proteins are not significantly close to any of the other species for which the corresponding sequence is known. However, the *Euglena* proteins do share a two-residue insertion in the middle of the molecule with the proteins from two green algae *Chlamydomonas* (Merchant & Bogorad, 1987) and *Bryopsis* (Okamoto *et al.*, 1987), and all these proteins contain tryptophan at position 59, rather than the phenylalanine present in other algal proteins.

Comparisons with mitochondrial cytochrome *c*

The *Euglena* mitochondrial cytochromes *c* are not particularly close in sequence to any other mitochondrial protein or to any of the homologous bacterial cytochromes c_2. With prokaryotes the maximum similarity yet observed is 50% with the *Rhodopseudomonas globiformis* protein (Ambler *et al.*, 1987). The similarity to the proteins from the ciliate protozoa of the genus *Crithidia*, which also have haem-binding sites that contain only a single cysteine residue (Pettigrew *et al.*, 1975; Hill & Pettigrew, 1975), is in the range 55–58% identity, higher than the 40–44% identity with the mitochondrial cytochromes *c* from the green algae *Enteromorpha intestalis* (Meatyard & Boulter, 1974) and *Chlamydomonas reinhardtii* (Amati *et al.*, 1988), or the 52–53% with a plant/animal/fungus deduced ancestral sequence (Baba *et al.*, 1981).

The haem-binding site in the *E. gracilis* mitochondrial cytochrome c_1 also has only a single cysteine residue (Matsubara *et al.*, 1989; Mukai *et al.*, 1989). A possible explanation of this unexpected phenomenon is that mutation in post-translational processing resulted in selection for change in a cysteine residue in the two mitochondrial cytochromes *c*, but not in the chloroplast cytochrome c_6. It is known that a haem lyase is required to attach the haem to cytochrome *c* in yeast and in *Neurospora* (Dumont *et al.*, 1987; Drygas *et al.*, 1989), but the mechanism is unknown. Perhaps the haem lyase is required to reduce the cysteine disulphide group before the haem can bind. A mutational loss of the lyase could have precluded haem binding unless one of the cysteine residues was also altered.

Relationship of *Euglena* to other unicellular eukaryotes

The chloroplasts of *Euglena* and the green algae resemble those from higher plants, and differ from other algae, in having both chlorophyll *a* and chlorophyll *b*. There are suggestions that the chloroplast proteins cytochrome c_6 and ferredoxin (Table 2) are slightly closer in sequence to those of green algae than to other phototrophs, but, as discussed above, the mitochondrial cytochrome *c* does not seem to be close in sequence to those from green algae. One conclusion is that amino acid sequence comparisons, even with proteins that are as strongly 'conserved'

as the ones studied in the present paper, may not be an effective probe for elucidating the past history of a group as divergent as the euglenoids. Our results are in concordance with the conclusions of Meyer *et al.* (1986). Another possibility is that in an alga the mitochondrion and the chloroplast can have evolved in separate lines. Thus *Euglena* could be a chimera containing flagellated-protozoon mitochondria and green-algal chloroplasts.

We thank Margaret Daniel for skilled technical assistance. We also thank Dr. Linda Fothergill-Gilmore, Miss Linda Kerr and Mr. Andrew Cronshaw of the WELMET Protein Characterization Facility, Department of Biochemistry, University of Edinburgh, for the automatic sequence determination, and the Wellcome Trust and the Salvesen Foundation for supporting this facility. A preliminary sequencer run had kindly been done for us by Professor John Fothergill, University of Aberdeen, Aberdeen, U.K. The work was also supported in part by Grant GM21277 from the National Institutes of Health to Dr. M. A. Cusanovich, University of Arizona.

REFERENCES

Amati, B. B., Goldschmidt-Clermont, M., Wallace, C. J. A. & Rochaix, J.-D. (1988) J. Mol. Evol. **28**, 151–160

Ambler, R. P. & Wynn, M. (1973) Biochem. J. **131**, 485–498

Ambler, R. P., Daniel, M., Meyer, T. E., Bartsch, R. G. & Kamen, M. D. (1979) Biochem. J. **177**, 819–823

Ambler, R. P., Daniel, M., Melis, K. & Stout, C. D. (1984) Biochem. J. **222**, 217–227

Ambler, R. P., Meyer, T. E., Cusanovich, M. A. & Kamen, M. D. (1987) Biochem. J. **246**, 115–120

Baba, M. L., Darga, L. L., Goodman, M. & Czelusniak, J. (1981) J. Mol. Evol. **17**, 197–213

Dayhoff, M. O. (1983) Precambrian Res. **20**, 299–318

Drygas, M. E., Lambowitz, A. M. & Nargang, F. E. (1989) J. Biol. Chem. **264**, 17897–17906

Dumont, M. E., Ernst, J. F., Hampsey, D. M. & Sherman, F. (1987) EMBO J. **6**, 2135–2141

Hill, G. C. & Pettigrew, G. W. (1975) Eur. J. Biochem. **57**, 265–270

Leedale, G. F. (1967) Euglenoid Flagellates, Prentice–Hall, New York

Margoliash, E. & Frohwirt, N. (1959) Biochem. J. **71**, 570–572

Matsubara, H. & Hase, T. (1983) in Proteins and Nucleic Acids in Plant Systematics (Jensen, U. & Fairbrothers, D. E., eds.), pp. 170–181, Springer-Verlag, Berlin

Mukai, K., Yoshida, M., Toyosaki, H., Yao, Y., Wakabayashi, S. & Matsubara, H. (1989) Eur. J. Biochem. **178**, 649–656

Meatyard, B. T. & Boulter, D. (1974) Phytochemistry **13**, 2777–2782

Merchant, S. & Bogorad, L. (1987) J. Biol. Chem. **262**, 9062–9067

Meyer, T. E., Cusanovich, M. A. & Kamen, M. D. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 217–220

Mukai, K., Wakabayashi, S. & Matsubara, H. (1989) J. Biochem. (Tokyo) **106**, 479–482

Okamoto, Y., Minami, Y., Matsubara, H. & Sugimura, Y. (1987) J. Biochem. (Tokyo) **102**, 1251–1260

Pettigrew, G. W. (1974) Biochem. J. **139**, 449–459

Pettigrew, G. W., Leaver, J. K., Meyer, T. E. & Ryle, A. P. (1975) Biochem. J. **147**, 291–302

Received 21 September 1990/7 November 1990; accepted 13 November 1990