The Purification and Amino Acid Sequence of Cytochrome *c*-552 from *Euglena gracilis*

By G. W. PETTIGREW

Department of Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, U.K.

(Received 4 October 1973)

Cytochrome *c*-552 from *Euglena gracilis* was purified and the amino acid sequence determined. The protein is a single peptide chain of 87 residues with the haem prosthetic group bound through two thioether linkages to two cysteine residues near the amino-terminal region. The amino acid sequence shows some similarities to mitochondrial cytochrome *c* and to two prokaryote *c*-type cytochromes. The sequence, taken with the known characteristics of cytochrome *c*-552, indicates that it is an *f*-type cytochrome. The possible functional and evolutionary significance of these features in common is discussed. Detailed evidence for the amino acid sequence of *Euglena* cytochrome *f* has been deposited as Supplementary Publication SUP 50027 at the British Library, Lending Division (formerly the National Lending Library for Science and Technology), Boston Spa, Yorks. LS23 7QB, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1973) 131, 5.

Two classes of easily extractable *c*-type cytochromes have been found in eukaryotic organisms in contrast with the considerable diversity of such proteins isolated from bacteria (Kamen & Horio, 1970). The mitochondrial cytochromes *c* have been extensively studied and are closely related in primary structure (Margoliash, 1970; Boulter & Ramshaw, 1972; Sugeno *et al.*, 1971). The acidic high-redox-potential cytochromes of many algae are also readily extracted (Yakushiji, 1971) and cytochrome *c*-552 from *Euglena* is a member of this group. They are thought to be related in function to the membrane-bound cytochrome *f* of higher plants (Davenport & Hill, 1951; Nelson & Racker, 1972).

Cytochrome *f* is thought to act as the reductant of photosystem I, receiving electrons, either from photosystem II or by cyclic electron transport and involving cytochrome *b₆* (Boardman, 1968). Structural investigations on cytochrome *f* might be justified on the following grounds.

The postulated homology between the small, soluble algal proteins and the membrane-bound proteins of minimum mol. wt. 30000 from higher plants can be studied. The classification of the algae is a disputed area of taxonomy with radically different interpretations of the available data (e.g. Lee, 1972; Raven, 1970). The evolutionary history of the group is relevant to the evolution of eukaryotic organelles and the evolution of photosynthesis. It may eventually be possible to compare various gene phylogenies with species phylogeny in order to test the symbiotic theory for eukaryotic organelles.

Structural relationships among functionally diverse classes of *c*-type cytochromes have been detected

(Ambler, 1971) on the basis of primary structure. The tertiary structures of two bacterial cytochromes, cytochrome *c₂* from *Rhodospirillum rubrum* (Salemme *et al.*, 1973) and cytochrome *c*-550 from *Micrococcus denitrificans* (Timkovich & Dickerson, 1973) show considerable homology to the tertiary structure of horse heart ferricytochrome *c* (Dickerson *et al.*, 1971). The primary structure of *Euglena* cytochrome *c*-552 will be discussed in relation to these results.

The amino acid sequence of one algal cytochrome *f*, that of *Monochrysis lutherii*, has been reported (Laycock, 1972). *Euglena* cytochrome *f* or *c*-552 has been purified and characterized by Perini *et al.* (1964a), Cusanovich *et al.* (1971) and Mitsui & Tsushima (1971). Its role in photosynthesis has been investigated by Perini *et al.* (1964b) and Katoh & San Pietro (1967).

Experimental and Results

Materials

Euglena gracilis strain z (1224/5z) was obtained from The Culture Centre of Algae and Protozoa, 36 Storey's Way, Cambridge CB3 0DT, U.K.

Ion-exchange celluloses, Sephadex grades and enzymes were obtained as described by Ambler & Wynn (1973).

Aminopeptidase M was obtained from P.L. Biochemicals, Milwaukee, Wis. 53205, U.S.A.

Ethyleneimine was obtained from Koch-Light Ltd., Colnbrook, Bucks., U.K.

Methods

Growth of organism. The photosynthetic protozoan was grown in batches of 10 litres on the acidic

heterotrophic medium of Hutner *et al.* (1966). The cultures were stirred and illuminated for 7–10 days on a New Brunswick 'Microferm' (New Brunswick Co., New Brunswick, N.J., U.S.A.). The cells were collected by centrifugation (Alfa-Laval LAB 102B-25 centrifuge) at a yield of 20–25 g wet wt. of cells/l and stored frozen.

Extraction and purification of cytochrome c-552. Cells (1 kg) were suspended in 1 vol. of cold 0.1 M-NaH₂PO₄-Na₂HPO₄, pH 7.0, and broken by passage through a Manton-Gaulin pressure cell operated at 5×10^7 Pa (500 kg/cm²). Cell breakage was monitored by microscopic examination. The broken-cell suspension was centrifuged at 25000g for 30 min and the debris was re-extracted with 1 vol. of phosphate buffer at 0–4°C for 1 h. After centrifugation of the suspension, the pH of the pooled supernatants was adjusted to 5.0 with 2 M-acetic acid. Centrifugation at 25000g for 30 min sedimented the green material and left a clear red supernatant, which was subjected to gel filtration on a column (9 cm × 40 cm) of

Sephadex G-25 (medium grade) in 0.01 M-NaH₂PO₄-Na₂HPO₄, pH 7.0.

The void-volume eluate was passed through a pad of DEAE-cellulose (Whatman DE-23; approx. 4 cm × 6 cm on a coarse glass sinter operated under gentle suction). The red-brown band at the top of the pad was scraped off, suspended in 0.01 M-NaH₂PO₄-Na₂HPO₄, pH 7.0, and poured into a 2 cm-diam. column. The column was washed with 0.05 M-NaH₂PO₄-Na₂HPO₄, pH 7.0, and the cytochrome c-552 was eluted with 0.1 M-NaH₂PO₄-Na₂HPO₄, pH 7.0.

Solid (NH₄)₂SO₄ was added to the cytochrome eluate to a concentration of 60% (w/v) at 0°C. After 1 h the precipitate was centrifuged down and the supernatant was saturated with (NH₄)₂SO₄. After centrifugation the bright-pink cytochrome pellet was dissolved in a minimum volume of water and the (NH₄)₂SO₄ was removed by gel filtration on a column (2 cm × 15 cm) of Sephadex G-25 equilibrated with 0.01 M-NaCl

Table 1. *Amino acid composition of Euglena cytochrome c-552*

Cysteine was estimated as cysteic acid by amino acid analysis on oxidized apoprotein and expressed relative to 3 mol of leucine. No methylated lysine was detected on the 50 cm Locarte analyser column, equilibrated and run in 1 M-sodium citrate buffer, pH 6.5 (adjusted to the correct pH with conc. HCl). The spectrophotometric method of Goodwin & Morton (1946) was used to estimate tryptophan in the apoprotein. The extinctions at 280 nm and 294.4 nm of a solution were measured. The concentration of apoprotein in the solution used was determined by amino acid analysis after hydrolysis of a sample. Portions of the same solution were hydrolysed for 20 and 70 h and analysed for amino acids. The values are in nmol. The best value denotes the averages of the 20 and 70 h values excluding those for threonine and serine which were extrapolated to zero time of hydrolysis and those for isoleucine and valine for which the 70 h values were taken. For 'Content (residues/mol)' iron was determined on a portion of the solution used for amino acid analysis. It was calculated that 31.6 nmol of iron was present in each analysis sample. This value was used as a divisor to calculate residues/atom of iron. The sequence column denotes the totals of the amino acid residues identified in the sequence. The reference analysis column shows the amino acid analysis of cytochrome c-552 published by Cusanovich *et al.* (1971).

	Amino acids (nmol)			Content (residues/molecule)	Sequence	Reference analysis
	Hydrolysis for 20h	Hydrolysis for 70h	Best value			
Asp	331	317	324	10.22	10	14
Thr	154	135	162	5.12	5	7
Ser	151	124	162	5.12	5	9
Glu	289	280	285	9.00	9	15
Pro	67	67	67	2.12	2	4
Gly	378	361	370	11.70	12	14
Ala	320	314	317	10.0	10	15
Val	275	315	315	10.0	10	9
Met	31.3	30.3	30.8	0.98	1	1
Ile	101	124	124	3.92	4	5
Leu	95	96	96	3.04	3	7
Tyr	155	153	154	4.86	5	5
Phe	29.6	30.5	30.1	0.95	1	3
His	30.4	30	30.2	0.96	1	2
Lys	125	122	124	3.92	4	8
Arg	31.8	30.6	31.2	0.99	1	3
Trp				2.05	2	
Cys				1.54	2	2

(attempted gel filtration in water resulted in bad 'trailing' of the protein and gel filtration in 0.01M-acetic acid-sodium acetate buffer, pH4.0, caused heavy precipitation of white material in the column which distorted the chromatography).

The pH of the cytochrome eluate was adjusted to 4.0 with 0.1M-acetic acid and the solution was diluted until the conductivity was near that of 0.01M-acetic acid-sodium acetate buffer, pH4.0.

Cytochrome *c*-552 was adsorbed on to a column of CM-cellulose (Whatman CM-52; 2cm×10cm; equilibrated with 0.01M-acetic acid-sodium acetate, pH4.0) and chromatography was carried out in 0.02M-acetic acid-sodium acetate buffer, pH5.0. Four cytochrome fractions were reproducibly obtained. Two major fractions were partly resolved and were oxidized and reduced cytochrome *c*-552. (The cytochrome was almost completely reduced after extraction, but the exposure to low pH resulted in partial oxidation.)

Two minor fractions, which were eluted after the main peak, accounted for about 1-5% of the total cytochrome on the column. They were fully oxidized and had spectra indistinguishable from the protein in the major fractions. They may represent modified or aggregated forms, but have not been further investigated.

The main peak fractions containing reduced and oxidized cytochrome *c*-552 were pooled and the cytochrome was concentrated on top of a small DEAE-cellulose column (Whatman DE-52; 1cm×3cm; 0.01M-NaH₂PO₄-Na₂HPO₄ buffer, pH7.0) and eluted with 0.5M-NaCl. The concentrated solution was stored at 0-4°C. Pure protein (3-4μmol) was recovered from 1kg wet wt. of cells.

Purity of preparation. Polyacrylamide-gel electrophoresis [10% (w/v) acrylamide, 0.2% (w/v) *NN*-methylenebisacrylamide in 0.05M-Tris-HCl buffer, pH7.5, polymerized with 0.5ml of a 10% (v/v) ethanolic solution of *NNN'*-tetramethylethylenediamine and 0.5ml of 10% (w/v) ammonium persulphate for 50ml of monomer solution] was carried out on 3nmol and 0.1nmol of purified cytochrome *c*-552 in gel columns of 7cm×0.5cm run at 5mA/gel. Only two zones were detected by staining with Amido Black after electrophoresis and these represented reduced and oxidized cytochrome (distinguishable by their colour before staining). It was difficult to keep the cytochrome either fully reduced or fully oxidized during electrophoresis. About 0.1nmol was easily detectable.

The ratio of $E_{280}/E_{416(\text{reduced})}$ of 0.2 was very similar to that of Cusanovich *et al.* (1971). These authors, however, found a discrepancy between their molecular-weight estimation by gel filtration (10000) and their amino acid composition data based on haem content, which indicated that the preparation was only 70% pure. Their results are compared with

the amino acid composition of this preparation in Table 1. The good agreement between the residues located in the sequence and the composition data indicate that this preparation was quite homogeneous.

The comparable ratios of $E_{280}/E_{416(\text{reduced})}$ of this preparation and that of Cusanovich *et al.* (1971) indicate that the contaminating protein of their preparation may be low in aromatic amino acids and cautions against the uncritical use of such ratios as indications of purity. It is proposed that chromatography on CM-cellulose removes this impurity, which Cusanovich *et al.* (1971) could not remove by repeated chromatography on DEAE-cellulose. It is difficult to assess the crystalline preparation of Mitsui & Tsushima (1971) since these authors do not present amino acid composition values.

Investigation of the sequence

The haem group was removed from the protein by incubation with HgCl₂ in acid solution (Ambler & Wynn, 1973). Tryptic, chymotryptic and thermolytic digestions were performed on 1-2μmol of apoprotein and the peptide mixtures were examined by the purification methods described by Ambler & Wynn (1973) and Ambler (1963). A standard procedure was used involving gel filtration of the digestion mixture on a column (1cm×90cm) of Sephadex G-25 (superfine grade) in 5% (v/v) formic acid, and examination of the eluate by paper electrophoresis at pH6.5 of portions of each fraction. The resultant peptide 'map' contained all the peptides of the digestion mixture that were soluble in 5% (v/v) formic acid and, after staining, was used to decide which fractions to pool for further purification by paper electrophoresis and chromatography.

The sequences of purified peptides were determined by the dansyl-Edman method (Hartley, 1970; Gray, 1967) and secondary digestion with trypsin, chymotrypsin, thermolysin, carboxypeptidase A or aminopeptidase M where necessary. The proposed sequence is shown in Fig. 1. Some particular aspects of the derivation of the sequence which require emphasis are described below.

Aminoethylation. The tryptic 'haem peptide' (Gly-1-Lys-25) was aminoethylated and then digested with trypsin (Cole, 1967). The peptide was dissolved in 0.05ml of EDTA solution (10mg/ml in 0.5M-Tris-HCl buffer, pH8.7) and 0.2ml of dithiothreitol was added (15mg/ml in 0.5M-Tris-HCl, pH8.7). The solution was gassed with N₂ and incubated at 25°C for 5h. Ethyleneimine (5μl; 100μmol) was added and then after 10min a further addition of 2μl in 0.2ml of 0.5M-Tris-HCl buffer, pH8.7, was made. After a further 10min, the

reagents were removed by gel filtration on a column of Sephadex G-25 (fine grade, 1cm×90cm) in 5% (v/v) formic acid and peptides were located with ninhydrin after separation of portions of each fraction of the column eluate by paper electrophoresis at pH6.5.

Cleavage with hydroxylamine. Two asparagine residues in the asparagine-glycine bonds Asn-16-Gly-17 and Asn-51-Gly-52 appeared to be susceptible to deamidation. Chemical and enzymic sequencing procedures were blocked and it is presumed that β -aspartyl bonds had been formed (Ambler, 1963). To overcome this, hydroxylamine cleavage was attempted before digestion of the apoprotein with thermolysin (Balian *et al.*, 1971).

Apoprotein (20mg) was incubated with 5ml of

2M-hydroxylamine-0.2M-potassium carbonate, pH 9.0, at 45°C for 2h and the reagents were removed by gel filtration on a column of Sephadex G-25 (1.5cm×10cm, fine grade) in 5% (v/v) formic acid.

Gel filtration in 5% (v/v) formic acid on a column (1.5cm×90cm) of Sephadex G-50 gave two main peak fractions. One contained a peptide with the amino acid composition of the region Gly-1-Asn-16, but from which carboxypeptidase A released no amino acids. The other peak was higher-molecular-weight material and presumably contained a mixture of large peptides owing to partial cleavage at the two asparagine-glycine bonds. This mixture was not characterized, but was directly digested with thermolysin and the peptides were fractionated.

Fig. 1. Amino acid sequence of *Euglena* cytochrome *c*-552

Peptides derived from tryptic digestion of the protein are shown above the sequence and are numbered sequentially from the *N*-terminus (T1-T6). Peptides derived by digestion with chymotrypsin (C) and hydroxylamine (H) and thermolysin (Th) are shown below the sequence and are similarly numbered. Full lines indicate quantitative analysis. Broken lines indicate qualitative analysis. —, indicates end groups and subsequent residues revealed by degradation with phenyl isothiocyanate and identified as Dns-amino acids. When a Dns-dipeptide was identified as well as a Dns-amino acid, the arrows are linked i.e. $\curvearrowright$. Fast and slow release of amino acids with carboxypeptidase A are denoted by $\leftarrow$ and $\leftarrow\leftarrow$ respectively. Fast and slow release of amino acids with aminopeptidase M are denoted by $\overline{\text{AP}}$ and $\underline{\text{AP}}$ respectively. The letters AE above cysteine in the sequence denote aminoethylcysteine. $\rightarrow\rightarrow$, indicates ambiguity or uncertainty in the identification of the Dns-amino acid. These few cases of uncertainty are individually discussed in the Supplementary Publication (SUP 50027). Secondary digestion of peptides is denoted by an arrow leading to the peptide fragments produced and labelled with the enzyme used for digestion. Thus thermolysin digestion of peptide T6 yielded peptides T6Th1 to T6Th10.

Peptides Gly-17-Asn-19 and Gly-52-Pro-57 were isolated, but also peptides Val-15-Ser-22 and Val-49-Pro-57. The latter two peptides had electrophoretic mobilities at pH6.5 which indicated that Asn-16 and Asn-51 were deaminated.

The cleavage with hydroxylamine was therefore only partly successful and this was probably due to

prior exposure to alkaline conditions (0.1M-NH₃) during the gel filtration separation of the apoprotein from the haem group. Bornstein (1970) reported that prior exposure to alkaline conditions causes saponification of the proposed cyclic imide to give a mixture of α and β aspartyl bonds which are stable to hydroxylamine attack.

Determination of tryptophan and location in the sequence. Some 2.07 residues of tryptophan and 4.7 residues of tyrosine per molecule were determined spectrophotometrically in the apoprotein (Goodwin & Morton, 1946). No Ehrlich-positive peptides soluble in 5% (v/v) formic acid were found in the tryptic digestion mixture, but the large insoluble C-terminal tryptic peptide (T6) contained tyrosine and tryptophan in the ratio of 1:1.13 (Goodwin & Morton, 1946). Two Ehrlich-positive peptides were isolated from both the thermolysin and chymotrypsin secondary digestions of peptide T6, and the tryptophan residues could be located on the basis of exopeptidase digestion of these sub-digestion peptides. Ehrlich-positive peptides from chymotryptic and thermolysin digestions of the whole protein confirmed these tryptophan assignments.

Assignment of amide groups. Amides were located by examination of the mobilities of peptides and secondary digestion peptides [no assignments depend exclusively on plots of electrophoretic mobility against molecular weight (Offord, 1966)]. Asp-8 and Asn-9 could not be identified in this way and the position of the amide group was determined during a run on the Beckman automatic sequencer.

Supplementary information

Detailed evidence for the amino acid sequence of *Euglena* cytochrome *c*-552 has been deposited in the British Library, Lending Division (formerly the National Lending Library for Science and Technology), Boston Spa, Yorks. LS23 7QB, U.K., for storage on microfiche (Supplementary Publication SUP 50027). The information comprises:

(1) A description and definition of the conventions and abbreviations used in recording the information.

(2) The results of carboxypeptidase A hydrolysis of whole apoprotein.

(3) Electrophoretic 'maps' of the tryptic, chymotryptic and thermolytic digestion mixtures.

(4) Tables summarizing the amino acid compositions of peptides and certain parameters of purification, i.e. the electrophoretic mobility at pH 6.5, the relative elution volume on gel filtration through Sephadex G-25 and the final yield.

(5) The descriptions of peptides from each enzymic digestion. This includes conditions of digestion, dansyl-Edman evidence, secondary digestion of peptides and assignment of amide groups.

Discussion

Deduction of sequence

The totals of sequence residues show good correspondence with the amino acid composition of whole protein (Table 1). The complete sequence could be

assembled from the tryptic and chymotryptic peptides, but the thermolysin peptides provided a useful confirmation of several doubtful areas.

The location of peptide T5 (Asn-Gly-Lys) was not based on overlapping dansyl-Edman evidence, since the chymotryptic peptides C5 and C5a could not be degraded beyond Asn-51. Peptide T5 was placed between peptides T4 and T6 because the section of sequence covered by peptides T1, T2, T3 and T4 was confirmed by overlapping chymotryptic peptides and peptide T6 was C-terminal (from the carboxypeptidase A hydrolysis of peptide T6 and whole protein). The thermolysin peptides Th11 and Th11a confirmed this location of peptide T5.

The three tryptic peptides of the aminoethylated peptide T1 could be unambiguously ordered and the sequence of peptide T1Tc was determined as far as Gly-17. This tenuous overlap into peptide C2c was supported by the composition of peptide Th3 and the compatibility of the amino acid compositions of peptides T1Tc, C2, C2a and C2c.

Some parts of the region Gly-54-Ser-87 could also not be positioned on overlapping dansyl-Edman evidence. Dansyl-Edman degradation was performed on peptide T6 as far as Leu-63 and the peptide T6Th3 extended this known sequence to Glu-67. The composition of peptide T6Cl indicated that peptides T6Th4 and T6Th5 complete the region 68-74, and this was confirmed by carboxypeptidase A digestion of peptide C5. Peptides T6Th4 and T6Th5 were ordered on the basis of expected chymotryptic cleavage at tyrosine in peptide T6Cl. The only peptide to contain the bond Tyr-74-Val-75 was peptide T6. Carboxypeptidase A digestion of peptide T6 gave results that were interpreted as major hydrolysis into the region covered by peptide C7 (or T6C3) and partial hydrolysis after Gly-81. Thus, of the four peptides produced by secondary chymotryptic digestion of peptide T6, peptide T6Cl is N-terminal and peptides T6C3 and T6C4 are adjacent with peptide T6C4 being C-terminal. Peptide T6C2 can therefore only fit between peptides T6Cl and T6C3.

It is proposed that two bonds in the protein are susceptible to deamidation and formation of β -aspartyl bonds. These are Asn-16-Gly-17 and Asn-51-Gly-52. The evidence on which this interpretation is based includes electrophoretic mobilities of peptides, apparently blocked dansyl-Edman degradations and partially successful cleavage of these bonds with hydroxylamine. The small tryptic peptide T5 was apparently not susceptible to this deamidation.

Out of the 86 peptide bonds in the protein 50 have been shown to be susceptible to enzymic hydrolysis. This is a considerably lower fraction than, for example, the very complete study by Ambler & Brown (1967). This is partly due to the lower quantities of starting material used, so that peptides were not

isolated in sufficient quantities to study them extensively by using enzymic methods. Less than 50mg of protein was used to establish the proposed sequence. This consideration of amount of starting material also precludes any comments on micro-heterogeneity, since peptides present in quantities less than 30nmol would probably not be detected and purified. In a digest of 1 μ mol this represents 3%.

Comparison of *c*-type cytochrome sequences

Fig. 2 includes the sequences of *Euglena* cytochrome *c*-552, *Monochrysis* *c*-553 (Laycock, 1972) and one of the two cytochromes of green sulphur bacteria reported by Van Beeumen & Ambler (1973), i.e. cytochrome *c*-555 from *Chlorobium thiosulphatophilum*. The variations encountered in the terminal regions of mitochondrial cytochromes *c* and the terminal sequences of *Rhodospirillum* cytochrome *c*₂ (Dus *et al.*, 1968) are also included.

The primary structures of *Euglena* cytochrome *c*-552, *Monochrysis* cytochrome *c*-553 and *Chlorobium* cytochrome *c*-555 have some similar features. The region preceding the characteristic *c*-type haem-attachment site (Cys-X-Y-Cys-His) contains glycine at position 6 and an aromatic residue at position 10 (the numbering is that of *Monochrysis* cytochrome *c*-553). By analogy with cytochrome *c* and cytochrome *c*₂ the conserved methionine at position 60 probably functions as the 6th iron ligand in the two cytochromes *f* and *Chlorobium* cytochrome *c*-555. Several other positions near Met-60 are conserved. There is a pronounced acidic region in the two algal cytochromes *f* between residues 68 and 72 and the aromatic residue 78 is conserved (although it is not found in *Chlorobium* cytochrome *c*-555).

If the primary structures of the two cytochromes *f* and *Chlorobium* cytochrome *c*-555 are compared with those of cytochrome *c* and cytochrome *c*₂ then many of these features are found to be common to the different classes of cytochrome. It is therefore possible that the following features of tertiary structure may be common to cytochrome *c*, cytochrome *c*₂ and cytochrome *f*.

Cytochrome *c* (Dickerson *et al.*, 1971) and cytochrome *c*₂ (Salemme *et al.*, 1973) have a structure known as the 'right channel', two sides of which are formed by the *N*- and *C*-terminal helices and the third side by a loop of residues immediately after the haem-attachment site. The two helices form a cross structure with Gly-6 positioned between the helices at the crossing point and in a space that could not accommodate a large side chain (Dickerson *et al.*, 1971). In cytochrome *c*, two aromatic residues Phe-10 and Tyr-97 lie close together and parallel to each other in the middle of the channel. In cytochrome *c*₂ (Salemme *et al.*, 1973) there is no aromatic residue at position 10 but Phe-20, located on the loop after

the haem-attachment site, takes its place as the coplanar partner to Tyr-107. An acidic patch of residues (especially pronounced in cytochrome *c*₂, residues 98–102) is located at the crossing point of the two terminal helices.

In the absence of information on tertiary structure for cytochrome *f* one might speculate therefore that the sequence similarities including the pattern of residues preceding the haem attachment site, the aromatic residues at positions 10 and 78 and the acidic patch of residues indicate that cytochrome *f* may have a structure comparable with the 'right channel' of cytochrome *c* and *c*₂.

The amino acid sequences of the cytochromes *f* between the haem-attachment site and Met-60 bear little relationship to the corresponding regions of cytochrome *c* or *c*₂. There is conservation of structure around Met-60 in the cytochromes *f* but this does not correspond to the highly conserved sequence preceding Met-80 of cytochrome *c*. The only aromatic residue which is conserved in the region between the haem-attachment site and Met-60 of the two cytochromes *f* is Tyr-51 (which is not found in the *Chlorobium* sequence). This contrasts sharply with the conservation of aromatic residues in the structures of cytochrome *c* (Dickerson *et al.*, 1971), cytochrome *c*₂ (Salemme *et al.*, 1973) and *Micrococcus* cytochrome *c*-550 (Timkovich & Dickerson, 1973). In particular, Tyr-67, Tyr-74 and Trp-59 of cytochrome *c* form what has been described as the 'left channel' and this structure has been implicated in the recently postulated mechanism of reduction of cytochrome *c* by Takano *et al.* (1973).

It is possible, therefore that the 'left side' of the cytochrome fold (which in cytochrome *c* and *c*₂ includes the 'left channel' and the residues surrounding the 6th iron ligand) is not conserved in cytochrome *f* and that any similarity that does exist with cytochrome *c* or *c*₂ is restricted to the 'right side'.

Why is the 'right channel' apparently conserved in these cytochromes? It is possible that the conservation reflects some common and critical functional role. Dickerson *et al.* (1971) suggested a role as a possible binding site for the oxidase. If the photosynthetic and respiratory electron transport chains have a common origin as now seems likely (Timkovich & Dickerson, 1973) then similarities might be detected in the interactions of cytochrome *c*, cytochrome *c*₂ or cytochrome *f* with their reductases rather than with their oxidases, which have very different functions. On this slender argument it is proposed that the structural similarity between these three classes of *c*-type cytochrome, i.e. the 'right channel', may be involved in interaction with the reductase.

An alternative explanation for the conservation around the haem-attachment site is that it may be

Fig. 2. Comparison of the amino acid sequences of *Euglena* cytochrome *c*-552, *Monochrysis* cytochrome *c*-553, *Chlorobium* cytochrome *c*-555, *Rhodospirillum* cytochrome *c*₂ and cytochrome *c*

Sequences are numbered from Gly-1 of vertebrate cytochrome *c* so that *Euglena* cytochrome *c*-552 begins at 5. The numbering is therefore different from that of Fig. 1 and is used in the Discussion section in comparing the different cytochromes. Conserved residues are boxed. The cytochrome *c* data included in the alignment summarizes the variation that has been found at different positions. The complete sequences of *Rhodospirillum* cytochrome *c*₂ and cytochrome *c* are not included in the alignment. The missing sections between Pro-30 and the region preceding the sixth haem ligand (Met-91 and Met-80 respectively) are much longer than the corresponding regions of the two cytochromes *f* and *Chlorobium* cytochrome *c*-555 and show no obvious relationship to them. The one-letter notation for amino acids is: A, Ala; B, Asp or Asn; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; X, *ε*-N-monomethyl-lysine; Y, Tyr; Z, Glu or Gln. Deletion is —.

Fig. 3. Features of the sequence of *Euglena* cytochrome *c*-552

Particular features that have been discussed in the text are emphasized. The numbering and one-letter notation for amino acids are from Fig. 2. Residues common to *Euglena* cytochrome *c*-552 and *Monochrysis* cytochrome *c*-553 are boxed. The proposed *N*- and *C*-terminal α -helices, homologous to those of cytochrome *c* and cytochrome *c*₂ are marked by a heavy line. Various features of the sequence that are homologous to features in the primary structure of cytochrome *c* are interpreted in terms of possible similarities in tertiary structure. These interpretations are highly speculative and must be tested by future X-ray-crystallographic studies.

necessary to hold the cysteine residues in a particular orientation so that the haem can be attached to the newly synthesized apoprotein. Most of the α -helix prediction methods discussed by Dickerson *et al.* (1971) predict α -helix in the region 9–21 of cytochrome *c* and, in fact, residues 9–13 and 14–18 do resemble α -helical turns in the

tertiary structure of cytochrome *c*. The possibility that this region is α -helical in the apoprotein, but that the organizing effects of the haem group distort the helix in the native protein, has been made unlikely by the work of Stellwagen & Rysavy (1972) and Fisher *et al.* (1973). These authors detected no organized structure in the apoprotein by using optical

rotatory dispersion and circular dichroism. However, there is presumably some specificity of attachment of the haem group whether this be enzyme recognition of the extended or of the folded polypeptide chain.

Repeating features in the sequence

There are some internal similarities in the *Euglena* cytochrome *c*-552 sequence, although one is reluctant to place great emphasis on these in view of the analysis of McLachlan (1972) who demolished most of the claims for internal homologies in proteins. The two regions

Ser-Lys-Thr-Ala-Ile-Glu-Glu-Tyr
28 35

and

Thr-Lys-Glu-Ala-Ile-Glu-Tyr
41 47

have four identities and one 'conservative' replacement in seven residues. These regions are not found in *Monochrysis* cytochrome *c*-553.

Evolutionary considerations

The similarities between the algal cytochromes *f* and mitochondrial cytochrome *c* discussed above are not extensive enough to exclude the possibility of convergent evolution. The most likely explanation for the similarities, however, is that cytochrome *f*, cytochrome *c*-555, cytochrome *c*₂ and cytochrome *c* have a common origin. It is more uncertain whether other bacterial cytochromes such as *Pseudomonas* *c*-551 are also homologous proteins (Ambler & Wynn, 1973).

Euglena cytochrome *c*-552 and *Chlorobium* cytochrome *c*-555 can be aligned with no deletions. *Monochrysis* cytochrome *c*-553 requires a single residue deletion. The two cytochromes *f* have 30 common positions out of the 84 compared (the gap being counted as a 21st amino acid). *Euglena* cytochrome *c*-552 and *Chlorobium* cytochrome *c*-555 have 14 common positions out of the 80 compared. These figures should be compared with the 28 residues common to all the cytochromes *c* studied. Distant pairs of cytochromes *c* have almost 50% of their residues the same. Cytochrome *c*-555 from *Chloropseudomonas ethylica* and *Chlorobium thiosulphatophilum* have about 50 common residues (Van Beeumen & Ambler, 1973).

The similarities between the sequences of cytochrome *c*-555 and *Euglena* cytochrome *c*-552 support the observations of Yamanaka & Okunuki (1968) who emphasized the similarities between these two types of *c*-type cytochrome in their reaction with oxidase systems and their Soret/ α -peak ratios.

If numerical comparisons between the cytochrome *f* and the cytochrome *c*-555 sequences were to yield any information about evolutionary distances in the way that the data for cytochrome *c* promises to (Dickerson, 1971) it would be necessary to assume a constant rate of change in the history of these proteins. This assumption is unlikely to be justified since the large difference in redox potential indicates that these proteins are not functionally equivalent. The similarities in the sequences do indicate a remote connection between the green sulphur bacteria and the eukaryotic algae. Blue-green-algal cytochromes *f* may be more conclusive homologues to the chloroplast proteins. Good homology would indicate that, either blue-green algae gave rise to the chloroplast or that they gave rise to the whole eukaryotic cell. From other evidence the former alternative is more likely and could be tested by comparing the phylogenies of different proteins associated with different organelles.

The sequence similarities between the two algal cytochromes *f* argue against a polyphyletic origin for eukaryotic photosynthesis (eg. in Raven, 1970). More cytochrome *f* sequences should greatly facilitate the construction of an algal phylogeny.

I am indebted to Dr. R. P. Ambler for his guidance and Dr. T. Meyer for helpful discussion, and I thank Miss S. Murray and Dr. D. Pritchard for running the protein on the Beckman sequencer.

References

- Ambler, R. P. (1963) *Biochem. J.* **89**, 349-378
- Ambler, R. P. (1971) in *Recent Developments in the Chemical Study of Protein Structures* (Previero, A., Pechere, J. F. & Coletti-Previero, M. A., eds.), pp. 289-305, Inserm, Paris
- Ambler, R. P. & Brown, L. H. (1967) *Biochem. J.* **104**, 784-825
- Ambler, R. P. & Wynn, M. (1973) *Biochem. J.* **131**, 485-498
- Balian, G., Click, E. M. & Bornstein, P. (1971) *Biochemistry* **10**, 4470-4475
- Boardman, N. K. (1968) *Advan. Enzymol. Relat. Areas Mol. Biol.* **30**, 1-79
- Bornstein, P. (1970) *Biochemistry* **9**, 2408-2420
- Boulter, D. & Ramshaw, J. A. M. (1972) *Phytochemistry* **11**, 553-561
- Cole, R. D. (1967) *Methods Enzymol.* **11**, 315-317
- Cusanovich, M. A., Meyer, T., Tedro, S. M. & Kamen, M. D. (1971) *Proc. Nat. Acad. Sci. U.S.A.* **68**, 629-631
- Davenport, H. E. & Hill, R. (1951) *Proc. Roy. Soc. Ser. B* **139**, 327-345
- Dickerson, R. E. (1971) *J. Mol. Evolution* **1**, 26-45
- Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A. & Margoliash, E. (1971) *J. Biol. Chem.* **246**, 1511-1536
- Dus, K., Sletten, K. & Kamen, M. D. (1968) *J. Biol. Chem.* **243**, 5507-5518

- Fisher, W. R., Taniuchi, H. & Anfinsen, C. B. (1973) *J. Biol. Chem.* **248**, 3188–3195
- Goodwin, T. W. & Morton, R. A. (1946) *Biochem. J.* **40**, 628–632
- Gray, W. R. (1967) *Methods Enzymol.* **11**, 139–151
- Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822
- Hutner, S. H., Zahalsky, A. C., Aaronson, S., Baker, H. & Frank, O. (1966) in *Methods in Cell Physiology* (Prescott, D. M., ed.), vol. II, pp. 217–228, Academic Press, London
- Kato, S. & San Pietro, A. (1967) *Arch. Biochem. Biophys.* **118**, 488–496
- Kamen, M. D. & Horio, T. (1970) *Annu. Rev. Biochem.* **39**, 673–700
- Laycock, M. V. (1972) *Can. J. Biochem.* **50**, 1311–1325
- Lee, R. E. (1972) *Nature (London)* **237**, 44–46
- Margoliash, E. (1970) *Harvey Lectures (Series 66)*, 177–247, Academic Press, London
- McLachlan, A. D. (1972) *J. Mol. Biol.* **64**, 417–437
- Mitsui, A. & Tsushima, K. (1971) *Methods Enzymol.* **23**, 368–371
- Nelson, N. & Racker, E. (1972) *J. Biol. Chem.* **247**, 3848–3853
- Offord, R. E. (1966) *Nature (London)* **211**, 591–593
- Perini, F., Kamen, M. D. & Schiff, J. A. (1964a) *Biochim. Biophys. Acta* **88**, 74–90
- Perini, F., Kamen, M. D. & Schiff, J. A. (1964b) *Biochim. Biophys. Acta* **88**, 91–98
- Raven, P. H. (1970) *Science* **169**, 641–646
- Salemme, F. R., Freer, S. T., Xuong, Ng. H., Alden, R. A. & Kraut, J. (1973) *J. Biol. Chem.* **248**, 3910–3921
- Stellwagen, E. & Rysavy, R. (1972) *J. Biol. Chem.* **247**, 8074–8077
- Sugeno, K., Narita, K. & Titani, K. (1971) *J. Biochem. (Tokyo)* **70**, 659–682
- Takano, T., Kallai, O. B., Swanson, R. & Dickerson, R. E. (1973) *J. Biol. Chem.* **248**, 5234–5255
- Timkovich, R. & Dickerson, R. E. (1973) *J. Mol. Biol.* **79**, 39–56
- Van Beeumen, J. J. & Ambler, R. P. (1973) *Antonie van Leeuwenhoek; J. Microbiol. Serol.* **39**, 355–356
- Yakushiji, E. (1971) *Methods Enzymol.* **23**, 364–368
- Yamanaka, K. & Okunuki, K. (1968) *J. Biochem. (Tokyo)* **63**, 341–346