Phytochemistry 1974, Vol 13, pp 2783 to 2789 Pergamon Press Printed in England

PHYLOGENETIC IMPLICATIONS OF THE AMINO ACID SEQUENCE OF CYTOCHROME *c* FROM *ENTEROMORPHA INTESTINALIS*

JOHN A. M. RAMSHAW,* DEREK PEACOCK, BARRY T. MEATYARD and DONALD BOULTER†

Department of Botany, University of Durham, Durham City DH1 3LE

(*Received* 14 *June* 1974)

Key Word Index—*Enteromorpha intestinalis*; Chlorophyta, ulvales, green alga, cytochrome *c*, amino acid sequence, methylated lysine residues, protein phylogeny, origin of photosynthesis

Abstract—The amino acid sequence of *Enteromorpha* cytochrome *c* has been added to an affinity tree relating the cytochrome *c* sequences of animals, plants, fungi, protozoans and one bacterium, cytochrome c_2 from *Rhodospirillum* The *Enteromorpha* sequence lies on the line of descent of the higher plant sequences, it is not closely related to the cytochrome *c* of the photosynthetic protozoan, *Euglena* The distribution of ϵ-*N*-trimethyllysine in cytochrome *c* is discussed

INTRODUCTION

THE AMINO acid sequences of numerous higher plant cytochromes *c* have been determined and used to construct an affinity tree relating these sequences.[1,2] Recently, Meatyard and Boulter[3] have determined the amino acid sequence of *Enteromorpha* cytochrome *c*, a green alga. This paper reports on the evolutionary position of this sequence relative to those of other organisms of the main phyla.

RESULTS AND DISCUSSION

An affinity tree (Fig. 1) has been constructed using the ancestral amino acid sequence method of Dayhoff and Eck.[4] The sequences used were *Enteromorpha intestinalis* and a representative sample from animals, fungi, higher plants, the two published protozoan sequences, i.e. cytochrome c_{558} from *Euglena gracilis* and cytochrome c_{557} from *Crithidia oncopelti*, and that of cytochrome c_2 from *Rhodospirillum rubrum*.‡ The resulting tree agrees with the topology published by McLaughlin and Dayhoff,[5] the minor differences

* Present address Department of Inorganic Chemistry, University of Sydney, N S W 2006

† To whom reprint requests should be sent

‡ References to literature sources, see legend of Fig 1

[1] BOULTER, D., RAMSHAW, J. A. M., THOMPSON, E. W., RICHARDSON, M. and BROWN, R. H. (1972) *Proc. Roy. Soc. (London)* **B 181**, 441

[2] BOULTER, D. (1973) *Chemistry in Botanical Classification* (BENDZ, G. and SANTESSON, J., eds.), *Nobel Symp.* 25 Nobel Foundation, Stockholm Academic Press, New York

[3] MEATYARD, B. T. and BOULTER, D. (1974) *Phytochemistry* **13**, 2777

[4] DAYHOFF, M. O. and ECK, R. V. (1966) *Atlas of Protein Sequence and Structure*, Vol. 2 National Biomedical Research Foundation, Maryland

[5] MCLAUGHLIN, P. J. and DAYHOFF, M. O. (1973) *J. Mol. Evol.* **2**, 99

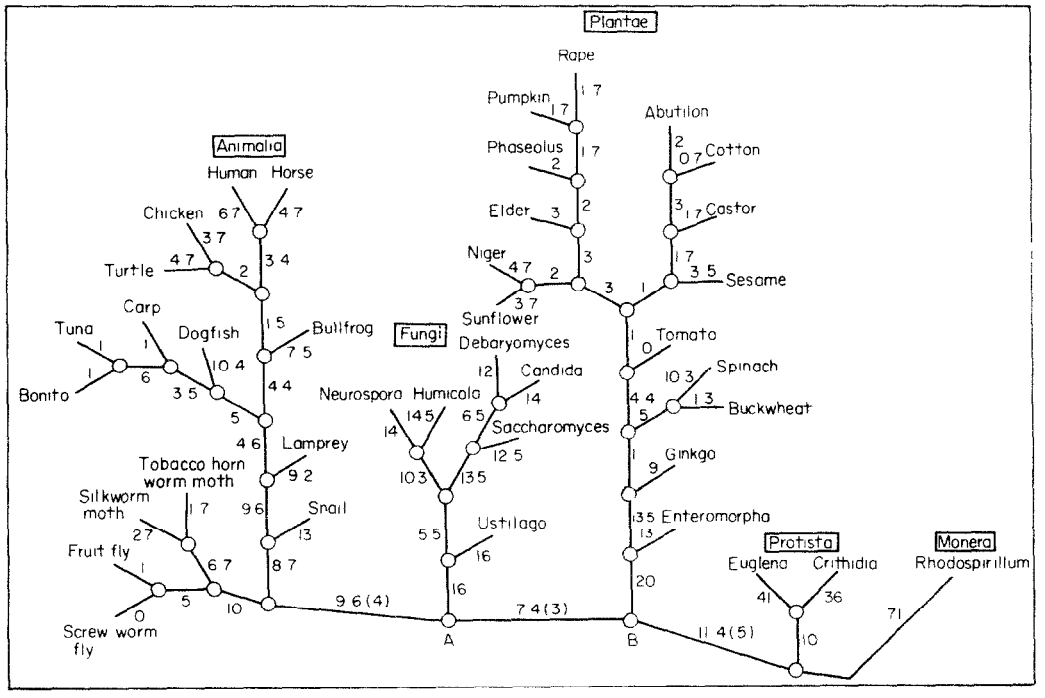

Fig 1 An affinity tree of cytochrome c

Tree constructed using the ancestral sequence method of Dayhoff and Eck [4] The lengths of the branches are in PAMs or Accepted Point Mutations estimated to have occurred on these branches Most of the sequence data are as from Dayhoff[32,33] In addition, are the recently reported sequences of *Euglena*[14] *Ustilago*,[29] snail,[34] spinach[35] elder,[36] Niger (J A M Ramshaw unpublished), correction of *Candida*[37] correction of baker's yeast (*Saccharomyces*)[38] correction of *Neurospora*,[39] and correction of *Humicola*[39]

between the two topologies are probably related to the selection of different equally probable solutions in that area of the topology

The data used to construct Fig 1 have been redrawn in Fig 2, where the branch lengths have been converted to elapsed time since divergence, assuming an approximately constant rate of evolution[6] The time-scale itself is based on the values given in Dickerson,[7] i e 20 million years for a change of 1 PAM The results in Fig 2 show that the time of divergence of *Enteromorpha* from the main plant line of descent was about 750 million years ago This finding, based upon calculations from the ancestral sequence method, is similar to the figure of 700 million years calculated from a direct comparison of the sequences themselves[8] The line to *Crithidia* and *Euglena* diverged 1300 million years ago at approximately the same time as the three eukaryotic phyla

Most botanists regard the Chlorophyta as the group which gave rise to higher plants.[9,10] However, recent ultrastructural considerations indicate that the Chlorophyceae

[6] Nolan, C and Margoliash, E (1968) *Ann Rev Biochem* **37**, 727
[7] Dickerson, R E (1971) *J Mol Evol* **1**, 26
[8] Ramshaw, J A M, Richardson, D L, Meatyard, B T, Brown, R H Richardson, M Thompson, E W and Boulter, D (1972) *New Phytologist* **71**, 773
[9] Klein, R M and Cronquist, A (1967) *Quart Rev Biol* **42**, 105
[10] Whittaker, R H (1969) *Science* **163**, 150

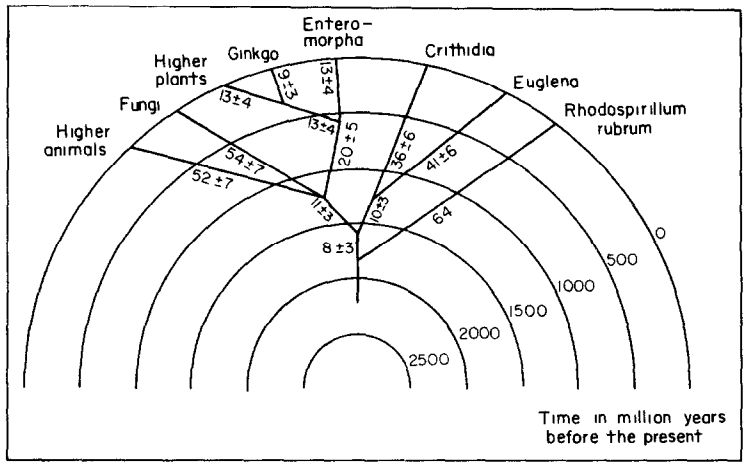

FIG 2 CLADOGRAM OF CYTOCHROME c SEQUENCES

Data as in Fig 1 redrawn to a time-scale based on Dickerson[7] For each kingdom, the average branch lengths from the kingdom node to each taxa in that kingdom were calculated (these values are shown in PAMs with their standard errors) The mean distance in PAMs that the kingdoms have evolved from the lowest node (i e Monera/Protista) was calculated from these values This value is represented in the Figure as the present day or 0 time line The distance of each point of divergence from this datum line is converted into elapsed time since divergence, using the estimate of Dickerson[7] (20 million per PAM corrected to 28 million to account for the fact that on average the distances on the ancestral tree are 1 4 times greater than the directly measured amino acid changes between sequences)

are a more divergent group than had previously been thought,[11] and it would appear that of the main groups or close relatives of the Chlorophyta, the Charophyceae (Charophyta) are on the direct line of descent of the green plants.[11] In Fig. 1, *Enteromorpha* appears as the most "primitive" sequence on the green plant line of descent, which is in accord with the other evidence mentioned above.

The Euglenophyceae (Euglenophyta) are considered by many authors to be closely related to the Chlorophyta, e.g Scagel *et al*.,[12] since both groups, in common with all land plants (Bryophytes to Angiosperms), contain both chlorophyll a and b The cytochrome data, however, do not support this suggestion and place *Euglena* closer to *Crithidia*, a non-photosynthetic protozoan Also, both the protozoan cytochromes have in common an unusual haem linkage in which the haem is attached through only a single cysteine residue [13,14] Thus, the sequence data suggest that the protozoans are "primitive" and well separated from the other main eukaryotic phyla, i.e. fungi, animals and plants. A similar placement was given to *Euglena* by Fitch[15] using the matrix method for tree construction.[16]

This placement would imply that photosynthesis may have arisen more than once in the course of evolution, a suggestion not incompatible with the endo-symbiont hypothesis

[11] PICKETT-HEAPS, J D and MARCHANT, H J (1972) *Cytobios* **6**, 255
[12] SCAGEL, R F, BANDONI, R J, ROUSE, G E, SCHOFIELD, W B, STEIN, J R and TAYLOR, T M C (1965) *An Evolutionary Survey of the Plant Kingdom* Blackie, London
[13] PETTIGREW, G W (1972) *FEBS Letters* **22**, 64
[14] PETTIGREW, G W (1973) *Nature* **241**, 531
[15] FITCH, W M (1973) *J Mol Evol* **2**, 123
[16] FITCH, W M and MARGOLIASH, E (1967) *Science* **155**, 279

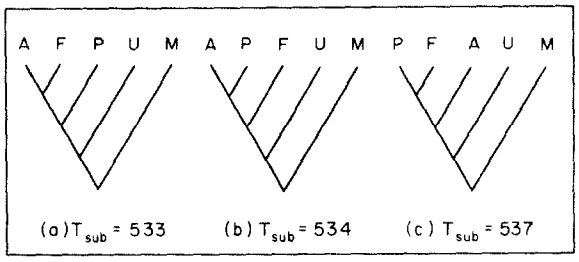

Fig 3 The three most parsimonious configurations of the five kingdoms'

All of the fifteen possible configurations of the five kingdoms were examined F Fungi A Animalia, P Plantae, U Protista (unicellular), M Monera, T_{sub}, total number of amino acid substitutions

of the origin of the chloroplast [17] If this were so, the results of photosynthetic experiments, in which mixed components from both *Euglena* and higher green plants are used, would need to be interpreted with care

McLaughlin and Dayhoff[5] derived a tree in which the order of derivation of the main kingdoms differs from that of Fitch [15] In both trees the branch lengths separating the junctions of the main kingdoms were large. In a previous paper,[8] it was calculated that the main eukaryotic phyla had originated more or less simultaneously. We have now repeated the calculations, including *Euglena*, amongst the lower organism group (see Fig 3), and find that the results are substantially the same as before. This finding can be interpreted to mean that the branch lengths between the junctions of the main phyla are small, and that the large branch lengths found by McLaughlin and Dayhoff[5] and by Fitch[15] may be artefacts due to the methods of calculation which they used In the Fitch[15] method, numerical residuals accumulate in this area of the dendrogram and in the McLaughlin and Dayhoff[5] method, since there are a considerable number of blanks at the level of the ancestral nodes, substitutions are averaged over the branches equally In Fig 1, therefore, for the nodes between the phyla, the known substitutions are given as a minimum number (in brackets), and unassigned substitutions, which have not been included in these values, have been averaged and added to give the larger values. When additional data reduce the number of blanks in the ancestral nodes associated with the junctions between animals

[17] Margulis L (1970) *Origin of Eukaryotic Cells* Yale University Press New Haven

Table 1 Sequences around the ε-N-

	72	75	80
Enteromorpha[3]	-Leu-Tyr-Asp-Tyr-Leu-Leu-Asn-Pro-TML-Lys-Tyr-Ile-Pro-		
Higher plants[1]	-Leu-Tyr-$\frac{Asp}{Glu}$-Tyr-Leu-Leu-Asn-Pro-TML-Lys-Tyr-Ile-Pro-		
Crithidia[13]	-Leu-Asp-Val-Tyr-Leu-Glu-Asn-Pro-TML-Lys-Phe-Met-Pro-		
Euglena[14]	-Leu-His-Lys-Phe-Leu-Glu-Asn-Pro-Lys-Lys-Tyr-Val-Pro-		
Neurospora[22,28]	-Leu-Phe-Glu-Tyr-Leu-Glu-Asn-Pro-TML-Lys-Tyr-Ile-Pro-		
Ustilago[29]	-Phe-Leu-Glu-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr-Ile-Pro-		
Humicola[24]	-Leu-Phe-Glu-Tyr-Leu-Glu-Asn-Pro-TML-Lys-Phe-Ile-Pro-		
Debaryomyces[30]	-Leu-Ser-Asp-Tyr-Leu-Glu-Asn-Pro-TML-Lys-Tyr-Ile-Pro-		
Saccharomyces[18,23,31]	-Met-Ser-Glu-Tyr-Leu-Thr-Asn-Pro-TML-Lys-Tyr-Ile-Pro-		

TML = ε-N-trimethyllysine

and fungi (A) and plants and lower organisms (B), the minimum value given between these nodes will probably not increase greatly, since it is expected that the unassigned substitutions will be distributed along the lines of descent.

The distribution of methylated lysine residues in cytochrome *c* is of interest (see Table 1). All higher plants contain two residues of ϵ-*N*-trimethyllysine in positions 80 and 94. *Enteromorpha* contains only one residue of ϵ-*N*-trimethyllysine in position 80. Animals, on the other hand, do not contain any methylated lysines. The situation in fungi is even more complex, since several fungi have one ϵ-*N*-trimethyllysine residue in the position equivalent to 80, while *Ustilago* does not contain any methylated lysines, and *Humicola* has a ϵ-*N*-trimethylated residue in the position equivalent to 94 and a dimethylated residue in the position equivalent to 80. In *Saccharomyces*, Foucher *et al* [18] have shown that a small quantity, 1–2%, of the *iso*-1-cytochrome is also present in a non-trimethylated form. *Crithidia* contains two residues of ϵ-*N*-trimethyllysine, one in the position equivalent to 80, the other at the *N*-terminal region of the protein in position 2 *Euglena* contains only one residue of ϵ-*N*-trimethyllysine in the position equivalent to 94

In *Neurospora* cytochrome *c*, Scott and Mitchell[19] have shown that the methylation of the lysine residue occurs after the initial synthesis of the protein and crystallographic studies with horse heart cytochrome *c*[20] have shown that both the lysine positions involved are on the surface of the molecule and to that extent available.

The sequence around the ϵ-*N*-trimethyllysine residue in position 80 in *Enteromorpha* is identical with that of the higher plants. In the case of position 94 in which lysine is methylated in the higher plants but not in *Enteromorpha*, there is a substitution of alanine in *Enteromorpha* for proline in position 91 (see Table 1). Since proline is important structurally,[21] it is possible that this change could have affected the conformation required for the specificity of the methylating enzymes.

In the case of animals where no methylated lysines occur,[22,23] this could be due to either the absence of the enzymes, or to conformational changes resulting from amino acid sub-

[18] Foucher, M., Verdiere, J., Lederer, F and Slominski, P P (1972) *European J Biochem* **31,** 139
[19] Scott, W A and Mitchell, H K (1969) *Biochemistry* **8,** 4282
[20] Dickerson, R E, Takano, T, Eisenberg, D, Kallai, O B, Samson, L, Cooper, A and Margoliash, E (1971) *J Biol Chem* **246,** 1511
[21] Dickerson, R E and Geis, I (1969) *The Structure and Action of Proteins* Harper & Row, New York
[22] DeLange, R J, Glazer, A N and Smith, E L (1969) *J Biol Chem* **244,** 1385
[23] DeLange, R J, Glazer, A N and Smith, E L (1970) *J Biol Chem* **245,** 3325

TRIMETHYLLYSINE RESIDUES IN CYTOCHROME *c*

85	90	95	100
Gly-Thr-Lys-Met-Val-Phe-Ala-Gly-Leu-Lys-Lys-Pro-Glx-Asx-Arg-Ala-Asp-Leu-			
Gly-Thr-Lys-Met-Val-Phe-Pro-Gly-Leu-TML-Lys-Pro-Glx-Asx/Glx-Arg-Ala-Asp-Leu-			
Gly-Thr-Lys-Met-Ser-Phe-Ala-Gly-Ile-Lys-Lys-Pro-Gln-Glu-Arg-Ala-Asp-Leu-			
Gly-Thr-Lys-Met-Ala-Phe-Ala-Gly-Ile-TML-Ala-Lys-Lys-Asp-Arg-Gln-Asp-Ile-			
Gly-Thr-Lys-Met-Ala-Phe-Gly-Gly-Leu-Lys-Lys-Asp-Lys-Asp-Arg-Asn-Asp-Ile-			
Gly-Thr-Lys-Met-Ala-Phe-Gly-Gly-Leu-Lys-Lys-Glu-Lys-Asp-Arg-Asn-Asp-Leu-			
Gly-Thr-Lys-Met-Ala-Phe-Gly-Gly-Leu-TML-Lys-Asn-Lys-Asp-Arg-Asn-Asp-Leu-			
Gly-Thr-Lys-Met-Ala-Phe-Gly-Gly-Leu-Lys-Lys-Ala-Lys-Asp-Arg-Asn-Asp-Leu-			
Gly-Thr-Lys-Met-Ala-Phe-Gly-Gly-Leu-Lys-Lys-Ala-Lys-Asp-Arg-Asn-Asp-Leu-			

stitutions, affecting enzyme specificity However, as pointed out by Delange *et al*.,[22] the latter possibility is not indicated, as *Neurospora crassa* cytochrome *c* is trimethylated in position 80, yet the sequence around this residue is identical with that of the animal sequences. In the protozoans[13,14] the Phe-Met sequence around position 80 in *Crithidia* is substituted by a Tyr-Val sequence in *Euglena*, whereas the sequence around position 94 is the same in both protozoans. Yet Lys-80 in *Euglena* and Lys-94 in *Crithidia* are non-methylated, whereas Lys-80 in *Crithidia* and Lys-94 in *Euglena* are trimethylated

Morgan *et al*.[24] have isolated two proteins from *Humicola* The major component contained one residue of ϵ-*N*-dimethyllysine and one residue of ϵ-*N*-trimethyllysine, whereas the minor protein contained one residue of ϵ-*N*-dimethyllysine and trace amounts of ϵ-*N*-trimethyllysine, both contained trace amounts of ϵ-*N*-monomethyllysine. These authors suggest that the trimethylation may be accompanied by a significant conformational change as the two components, both with the same overall charge, are readily separated by ion-exchange chromatography A similar situation is thought to exist in the two components of *iso*-1-cytochrome from *Saccharomyces*[18] Morgan *et al*.[24] consider the heterogeneity of the methylated lysine population in *Humicola* to have arisen, in part, by incomplete methylation of the lysine residues However, various ϵ-*N*-mono- and -dimethyllysines have been reported in other proteins, e g histones[25,26] and flagellar proteins,[27] which suggest that their occurrence is not due to incomplete methylation in all cases Since Scott and Mitchell[19] did not find ϵ-*N*-mono- or -demethyllysine derivatives in *Neurospora*, the possibility of dimethylation occurring during extraction of *Humicola* cytochrome *c* has to be considered Whilst no pattern can be observed in the primary sequences which might suggest the specificity of the methylating enzymes, it must be remembered that comparisons between related parts of sequences disregard more distant parts of the sequence which may nevertheless be adjacent in the three-dimensional structure, and important for conformation

The actual location of the methylating enzymes relative to the protein to be methylated, may also be important Thus, whilst the functional significance of methylation is not clear, Scott and Mitchell[19] have suggested that in *Neurospora* it may be involved in the binding

[24] MORGAN W T, HENSLEY C P JR and RIFFM, J P (1972) *J Biol Chem* **247**, 6555
[25] HNILICA L S (1967) *Progr Nucl Acid Res Mol Biol* **7**, 25
[26] PAIK, W K and KIM, S (1967) *Biochem Biophys Res Commun* **27**, 479
[27] AMBLER R P and REES M W (1959) *Nature* **184**, 56
[28] HELLER, J and SMITH, E L (1966) *J Biol Chem* **241**, 3165
[29] BITAR, K G VINOGRADOV S N, NOLAN, C, WEISS, L J and MARGOLIASH E (1972) *Biochem J* **129**, 561
[30] SUGENO K, NARITA, K and TITANI, K (1971) *J Biochem (Tokyo)* **70**, 659
[31] NARITA K and TITANI, K (1969) *J Biochem (Tokyo)* **65**, 259
[32] DAYHOFF M O (1972) *Atlas of Protein Sequence and Structure*, Vol 5 National Biomedical Research Foundation Maryland
[33] DAYHOFF M O (1973) *Atlas of Protein Sequence and Structure*, Vol 5 suppl 1 National Biomedical Research Foundation, Maryland
[34] BROWN R H RICHARDSON M BOULTER D, RAMSHAW, J A M and JEFFERIES R P S (1972) *Biochem J* **128**, 971
[35] BROWN, R H, RICHARDSON M SCOGIN R and BOULTER, D (1973) *Biochem J* **131**, 253
[36] BROWN, R H and BOULTER, D (1974) *Biochem J* **137**, 93
[37] LEDERER F (1972) *European J Biochem* **31**, 144
[38] LEDERER F SIMON A M and VERDIERI J (1972) *Biochem Biophys Res Commun* **47**, 55
[39] LEDERER F and SIMON A M (1974) *Biochem Biophys Res Commun* **56**, 317

of the protein to the mitochondrial matrix. It is possible, therefore, that the methylating enzyme itself may be part of the mitochondrial membrane, and as a consequence only some lysines are methylated.

Acknowledgements—We thank Mr G Bainbridge, Mrs E Innes, Mrs L Mayo and Mrs L Colpin for invaluable technical assistance, and the Nuffield Foundation and Science Research Council for financial support