Molecular evolution of cytochrome c from algae

Meatyyard, Barry T.

MOLECULAR EVOLUTION OF CYTOCHROME *c* FROM ALGAE

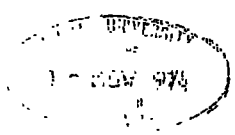
A Thesis
submitted in accordance
with the requirements of
the University of Durham
for the degree of
Doctor of Philosophy

By

Barry T. Meatyard

August 1974.

Department of Botany.



CONTENTS

		Page No.
Summary.		i
Acknowledgements.		iii
Abbreviations.		iv
INTRODUCTION.		1
MATERIALS AND METHODS.		
I.	General.	19
II.	Protein purification methods.	28
III.	Protein sequence determination methods.	35
IV.	Calculations based on amino acid sequence data.	43
RESULTS.		
1.	Purification of *Enteromorpha* cytochrome *c*.	45
2.	Purification of *Rhodymenia* cytochrome *c*.	47
3.	Purification of *Porphyra* cytochrome *c*.	48
4.	Extraction of cytochrome *c* from *Nitella*.	49
5.	Extraction of cytochrome *c* from brown algae.	50
6.	Extraction and purification of *Saprolegnia* cytochrome *c*.	50
7.	Purification of mushroom cytochrome *c*.	51
8.	The amino acid sequence of *Enteromorpha* cytochrome *c*.	51
9.	The amino acid sequence of *Rhodymenia* cytochrome *c*.	61
10.	The partial sequence of *Porphyra* cytochrome *c*.	68
11.	Preliminary sequence investigations of *Saprolegnia* cytochrome *c*.	71
12.	Cyanogen bromide cleavage of mushroom cytochrome *c*.	73

Page No.

DISCUSSION.

The distribution and characterization of *c*-type algal cytochromes. 74

Technical problems in the purification and sequencing of algal cytochromes *c*. 76

Sequence comparisons. 95

Evolutionary aspects. 105

BIBLIOGRAPHY. 137

APPENDICES.

SUMMARY

Cytochrome c has been extracted and purified from three species of algae, Enteromorpha intestinalis, Rhodymenia palmata and Porphyra umbilicalis, and from two fungi, mushroom (Psalliota campestris) and Saprolegnia sp. The complete amino acid sequence of the cytochrome c from Enteromorpha and partial sequences of the cytochromes c from Rhodymenia and Porphyra have been determined. Preliminary sequence investigations were carried out on the cytochromes c from mushroom and Saprolegnia. The sequence data obtained in this investigation showed that the cytochromes c from algae are clearly homologous with other eukaryotic mitochondrial cytochromes c. Sequence comparisons were used to examine the times of divergence of various taxonomic groups and to determine the phylogenetic relationships of the Chlorophyta and Rhodophyta. Phylogenetic trees were constructed by the ancestral sequence method and an overall minimum tree relating thirty eight species was obtained.

The results indicate that the Chlorophyta are on the direct line of descent of the higher plants, and that the Rhodophyta represent a more remotely diverging group on the same line of descent.

Comparisons of the algal sequences with the established sequences of Euglena and Crithidia cytochromes c placed the

Euglenophyta as being more closely related to the Protozoa than to the Chlorophyta, and this is held to support the division of living organisms into five kingdoms of equal taxonomic status.

When the ancestral sequence method was applied to the currently available sequence data for fungi, certain groupings became apparent. Neurospora and Humicola were joined to one branch node and Candida and Debaryomyces to another. Saccharomyces and Ustilago were placed on their own respective branches. The minimum topology obtained was not in accordance with current fungal taxonomy.

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Professor D. Boulter, for his continued help and guidance at all stages in this work, and for the use of the facilities of the Botany Department in the University of Durham.

My thanks are due to Drs. E. W. Thompson, J. A. M. Ramshaw and M. Richardson, for much helpful advice, and to Dr. D. Peacock for his assistance with the computing.

I am grateful to Mr. G. Bainbridge and the members of the technical staff of the Botany Department, who assisted in the collection of algae, and to Miss M. Berryman for typing this thesis.

I should also like to acknowledge the financial support received from the Science Research Council.

ABBREVIATIONS

The abbreviations used in this thesis are as recommended in "Instructions to Authors", Biochem. J. 121, 1, with the following additions:-

$CySO_3$	:	Cysteic acid
MeS	:	Methionine sulphone
Me_3Lys	:	ϵ-N-trimethyllysine
ox.	:	oxidised (ferricytochrome)
red.	:	reduced (ferrocytochrome)

INTRODUCTION

Common to all of Man's activities is a desire and need to classify. It is not surprising, therefore, that much attention has been given to the ways in which systems of classification are devised. From a practical standpoint biologists must have a universal code by which they recognise organisms by name. However, there is no major measure of agreement among taxonomists as to whether the main aim of a classificatory scheme is to reflect the evolutionary relationships of the organisms (phyletic classification), or to give groupings of organisms which give the maximum similarity of the characters used in the classification (phenetic classification) (see Davis and Heywood, 1963). Cronquist (1968) for example, is strongly in favour of the former view, Sokal (Sokal and Sneath, 1963) the latter. Whichever view is taken the classification of a given organism depends on the amount and quality of information with respect to that organism, that is available to the taxonomist. Turrill (1938) regarded the development of taxonomy as a gradient leading from 'alpha' to 'omega', the 'alpha taxonomy' being based on morphology, and the 'omega' using a much greater diversity of characters (see also the pioneer and encyclopaedic taxonomies of Davis and Heywood (1963)).

With regard to the algae, Dixon (1970) in a recent review, regards the present taxonomy as being very much in the alpha stage and that the classification of algae is designed for the identification of a particular species, and it is thus phenetic rather than phyletic. This situation may partly be due to the fact that at this point of time probably the only sure way of establishing phylogenetic relationships within a group, is a study of an adequate fossil record. This approach has been extensively used for members of the animal kingdom, particularly vertebrates, for which many lines of descent and divergence are now well documented (see Colbert, 1969).

The fossil record for algae extends back into the Pre-Cambrian era, although the greatest number of records is from the Cretaceous to the present time (Tilden, 1935; Round, 1965). Tilden (1935) suggested that the fossil algae can be grouped into five 'Algal Periods' which reflected the quality of illumination incident on the earth at the time, i.e. the groups are based on the evolution of different pigment systems. However, little evidence is available from the fossil record to satisfy the requirements of a phyletic system and the phylogenetic relationships of fossil algae are uncertain. In the absence of a fossil record for algae the manner in which the various classes

evolved must be conjectured on the basis of comparative studies of extant species. At one time the organisms referred to as 'algae' were grouped in one category. With increasing knowledge resulting especially from studies of biochemistry and physiology, it is now apparent that the group 'algae' is extremely artificial. The algae are generally considered to comprise several parallel lines of evolution that are only distantly related. Parallel evolution of certain morphological traits is evident in several classes of algae, and in several divisions variations seem to have occurred independently including rhizopodial, flagellated, palmelloid, coccoid, filamentous and siphonous forms. This parallelism is particularly well documented with respect to the Chlorophyceae, Xanthophyceae and Chrysophyceae (Smith, 1950; Fott, 1959; Rhodes and Stofan, 1967). Because of the high degree of parallelism and convergence in the groups of extant algae, a classification based on morphological characters is largely superfluous and in many groups taxonomic study is becoming more and more a comparative study of life histories (Bliding, 1963; Van den Hoek, 1964). This is an application of the ontogenic approach as defined by Davis and Heywood (1963).

The current conservative concept of the groups of organisms which constitute the 'algae' recognises the

following nine phyla:- (1) Chlorophyta; (2) Cyanophyta; (3) Chrysophyta; (4) Euglenophyta; (5) Pyrrophyta; (6) Rhodophyta; (7) Phaeophyta; (8) Cryptophyta; (9) Chloromonadophyta (Prescott, 1964). This sytem is based on the delineations of Pascher (1931) who, however, placed the Cyanophyta outside the general term 'alga'. The position of this group has also been questioned more recently by Van Oye (1961). Scagel et al. (1965) follow the same general pattern as Prescott (1964) but place the Cryptophyta as a class 'Cryptophyceae' in the Pyrrophyta and place the Chloromonadophyta as the class 'Chloromonadophyceae' in the phylum Xanthophyta. The possible interrelationships and phylogenetic arrangement of the classes of algae according to Scagel et al. (1965) are shown in App. 2 (see also Appendix 1). A broader approach is taken by Christensen (1966) whereby the eukaryotic algae are divided into two groups, the Aconta (Rhodophyta only) and the Contophora, which comprises all the other groups.

Traditionally living organisms have been divided into two kingdoms - Plants and Animals, but many single-celled algae and protozoa occupied uncertain positions in this scheme. Taking this into consideration, Copeland (1938, 1947, 1956) invoked a four-kingdom system (App. 5). However, the limitations of this scheme were realised by

Whittaker (1969), who revised Copeland's four kingdoms into a five-kingdom system (Appendix 5). Whittaker based his divisions on, (a) cell type, and (b) mode of nutrition. For the algae, this scheme is in good agreement with that of Scagel et al. (1965) (Appendix 2), but, in addition, attempts to indicate the relationships of algae with other groups. Although these various schemes differ in certain areas, there is a general agreement that the blue-green algae and the red algae represent the earlier evolved forms (see also Skuja, 1938; Tilden, 1935; Allsop, 1969). Fritsch (1945), Chapman (1962) and Boney (1966) are dubious of the strength of the phylogenetic relationship between these groups which had been proposed as early as 1867 (Cohn, 1867), although there is general acceptance that such a relationship exists (Tilden, 1935; Kylin, 1943; Papenfuss, 1955; Chadefaud, 1960; Christensen, 1962, 1964, 1966).

The controversy and uncertainties that exist in considering the classification of the algae have been underlined by Prescott (1964), whose view is that because of the uncertain phylogenetical relations in the algae it must be recognised that the classification system should be fluid and in the light of modern taxonomic data will change. Steps towards the 'omega' classification condition for algae are being taken and the relevant research activities have

been grouped into nine areas by Prescott (1964). These include physiology, ecology, microscopy and ontogeny, to which in the last few years more specialist additions can be made. Attempts to infer phyletic relationships between algae independently of fossil evidence, have more recently involved the use of comparative plant chemistry and biochemistry. If Cronquist's view that a classification scheme should be phyletic is taken, then the taxonomy of the algae must rely on sophisticated considerations such as the above. These considerations are largely biochemical and physiological and based on criteria such as pigment composition, structure of the cell wall, the type of storage product and the number and type of flagella present in the mobile stage of the life cycle. These characters are summarised in Appendix 1 which represents the scheme on which Scagel et al. (1965) have based their phylogeny (Appendix 2).

In using specialised biochemical data to establish a phylogenetic scheme, care must be taken to select suitable information. This is in order that the taxonomy aimed at does not have the same limitations as schemes based on morphological data. Thus, as in classical morphological taxonomy, biochemical characters can be considered as either having weak taxonomic significance, or strong taxonomic significance (see Davis and Heywood, 1963). Zuckerkandl

and Pauling (1965) distinguished between the semantides, DNA, RNA and proteins, and the episemantides, intermediate compounds of metabolism produced by the activity of the former, and discussed the strength of the taxonomic data that a study of each offered. A great deal of the biochemical information available belongs to the latter group of compounds and is considered to be of weaker taxonomic significance than information derived from the semantides.

The majority of episemantic molecules are of polygenic origin. That is to say that they are formed by biochemical pathways that are controlled by many enzymes. This can allow varying degrees of convergence of structure and function, which as previously discussed for the algae is a disadvantage of using purely morphological features to infer phylogeny. Attempts to use episemantides in taxonomy are also complicated by the fact that there may be more than one biochemical pathway leading to the formation of the particular metabolite, as is the case with certain amino acids, particularly lysine. This represents convergence of metabolic pathways to produce a common end product, and the common possession of, for example, lysine need not necessarily imply a phyletic relationship between two organisms (see Bartnicki-Garcia, 1970).

Of particular interest with respect to the Algae is a

proposed phylogenetic classification based on their lipid metabolism (Nichols, 1970). This scheme considers the significance of the fact that certain polyenoic acids can also be synthesized by at least two biochemical routes. A close relationship between the Euglenophyta and Chlorophyta is suggested and the Rhodophyta are placed as the ancestors of the Phaeophyta, Bacillariophyta and Xanthophyta (see Appendix 3). However, it is stressed that this scheme is more a classification of algae based on the lipid composition and metabolism of chloroplasts than a definition of evolutionary development. On the basis of studies of polyglucan metabolism an evolutionary progression from the Cyanophyta to the Rhodophyta to the Chlorophyta is suggested by Frederick (1968).

The occurrence of phycobilin pigments in the Cyanophyta and Rhodophyta and protoplasmic connections between cells of certain representatives may indicate the common origin of these two groups. Also, the simplicity of the chloroplast structure in the Rhodophyta may be significant in showing their close relationships to the Cyanophyta (Gibbs, 1970; Klein, 1970). However, other cytological and morphological details give the two groups an extremely divergent evolution (Chapman, 1962). The presence of phycobilins is also apparent in Cryptophyceae and may also indicate relationships

of the Pyrrophyta with the Cyanophyta and Rhodophyta. However, it appears that phycobilin pigments may have developed independently in the Cryptophyceae, since the latter, unlike the former two groups, do not have phycobilisomes (Stanier, 1974). Also, many of this group possess flagella which are lacking in the Cyanophyta and Rhodophyta (Scagel et al., 1965). The Cryptophyceae contain chlorophyll c in addition to chlorophyll a and the phycobilins. For this reason they may have evolved from a form that also gave rise to another line represented by the more typical Pyrrophyta, the Dinophyceae (dinoflagellates). This latter class which lacks phycobilins, has an abundance of carotenoid pigments as well as chlorophyll c, which also occurs in the Phaeophyta and some members of the Chrysophyta, particularly the diatoms (Bacillariophyceae). Also, in these two classes the carotenoid fucoxanthin is abundant, & the storage product is laminarin or the closely related chrysolaminarin. These differ from the starch-like carbohydrates of the Rhodophyta, Cyanophyta and Pyrrophyta in having β 1-3, 1-6 linkages between the glucose molecules instead of the α 1-4, 1-6 linkages of the latter groups (Parker, 1970).

Another similarity of the Phaeophyta and Chrysophyta is seen in the common occurrence of one whiplash flagellum and one tinsel flagellum (Bouck, 1969). Whether all of

these forms have evolved from a common ancestor is highly debatable, although at least two groupings exist.

The algae that lack masking pigments are also morphologically diverse, although possibly not as much as those classes considered previously. Chlorophyll a is common to all photosynthetic algae, and some only possess this one ubiquitous pigment (Xanthophyta). In addition, the Euglenophyta and Chlorophyta have chlorophyll b, (cf. Christensen, 1962). These algae may have had a common ancestry from which three distinct lines have resulted. The Xanthophyceae store chrysolaminarin and possibly diverged from the Chrysophyta-Phaeophyta line fairly early. The Euglenophyta have the same carbohydrate splitting enzymes as the Phaeophyta and the main storage product of the Euglenophyta is starch-like. Certain modern genera in the Chlorophyceae illustrate features to support the hypothesis that all green plants (vascular and non-vascular) evolved from some green algal or green algal-like precursor. For this reason it has been suggested that the Chlorophyta should be extended to include all green plants as one division which is then held to include the various classes of both green algae and telom plants (Fott, 1965). The Charophyceae represent a diverging evolutionary line from Chlorophycean ancestry and are considered by some to

represent the ancestors of the Bryophytes (see Scagel et al., 1965), although their evolutionary position is uncertain (see Fott, 1965). Whilst it can be confidently asserted that we know little of the affinities of the main algal groups, most botanists regard the Chlorophyta as the group from which higher plants have most likely originated (see Klein and Cronquist, 1967; Whittaker, 1969; Klein, 1970). However, recent ultrastructural considerations place the Charophyceae on the line of descent of the Bryophytes and higher plants, and suggest that the Chlorophyceae are a more divergent group (Pickett-Heaps and Marchant, 1972).

It is thus evident that the biochemical data available from studies of episemantides, provides as much conflicting taxonomic information as does the data derived from morphological and cytological data. This may be due to the fact that generally biochemical characters have been considered in isolation and that a more detailed numerical analysis of characters is required to clarify the situation. Numerical analysis of a large set of characters has been applied to members of the Chlorophyte genus Chlorococcum (McGuire, 1969), but difficulty still arises in the definition of typical operational taxonomic units (species) due to the weighting of characters. Such an analysis still does not infer phyletic relationships between the members of a group.

The semantides provide the primary genetic link between successive generations, and it is, therefore, within these molecules that the biochemical basis of evolution is documented. A distinction can be made between primary semantides - the DNA (and in some cases, RNA), that forms the primary blue-print for the cell, the secondary semantides - mostly messenger and transfer RNAs, that are formed according to the DNA master template, and the tertiary semantides - the proteins which are formed as a result of base sequences in the messenger RNAs. The detailed chemical analysis and characterization of primary and secondary semantides still remains a difficult task, although the base sequences of a limited number of relatively small tRNA molecules are known (see Dayhoff, 1972). DNA hybridization techniques have been applied to examine relationships between groups of organisms (Kohne, 1968), but difficulties arise both in the technique and interpretation of this method. Of particular relevance is the fact that during evolution, inversions, translocations and repetitions of DNA base sequences have occurred and these may make the results obtained from hybridization techniques difficult to interpret.

The implication of tertiary semantides, proteins, in taxonomic studies has fallen into three main areas of research. These are serology, comparative biochemistry and

analytical structural studies.

Serological methods are comparatively rapid and may be systematically valuable. The main advantage of the serological method is its ability to distinguish antigenic substances which are indistinguishable by other chemical means. The founder of serology as a tool in plant taxonomy, is held to be Mez, who pioneered work in this field in the 1920's (see Chester, 1937; Vaughan, 1968a). The application of modern serological techniques to certain taxonomic problems has proved successful (Vaughan, 1968a and b). However, the method does involve certain problems which could cause anomalies in results. Rabbits are commonly used to prepare antiserum and their ability to react to antigens to produce antibodies varies with their physiological condition. Also, there may exist more than one antigenic site on a given protein molecule, thus producing an antiserum of mixed antibodies. However, these problems are largely technical and the results of serological investigations testify to the technique as yielding much taxonomic information. For the higher plants Vaughan showed that serological methods gave results in accordance with classically established taxonomies for Solanum and Brassica species (Vaughan, 1968a) and Brassica and Sinapis species (Vaughan, 1968b). With respect to the algae, Bennett and Bogorad (1971) have

shown a serological relationship between the biliproteins of the Rhodophyta and Cyanophyta, but this is too limited to permit the analysis of evolutionary relationships (Glazer et al., 1971). Otherwise, it appears that there is very little evidence that serology has been applied in a similar way to other algal groups.

Comparative biochemical studies of enzymes and functional proteins provide another approach to taxonomic problems. It is advantageous in such studies that the test systems be those which are likely to have changed little during evolution, and consequently have retained the properties of these systems in ancestral organisms. In addition, it is preferable that the components involved in such systems should occur widely throughout the plant, animal, fungal and bacterial kingdoms. All aerobic organisms possess an oxidative electron transport system involving cytochromes. Cytochromes are also present in many anaerobic microorganisms. In either case cytochromes function as fundamental components in energy conversion mechanisms. Among the cytochromes, most attention has been paid to cytochrome c. This has the advantage of being relatively easy to purify, since it is a small molecule of molecular weight about 13,000, is highly basic, relatively stable, and is a coloured protein. Comparative studies using cytochrome c have involved two

major approaches. The first involves a comparison of the relative reactivities of the protein from different sources with cytochrome oxidases, and the second involves comparisons of the fine-structure of the molecule, also from different species. Yamanaka and his co-workers have established evolutionary relationships of many organisms, based on the relative reactivity of their cytochrome c with Pseudomonas and cow cytochrome oxidase (see Yamanaka, 1966). The advantage of this method is that it is applicable to very small amounts of cytochrome c and, therefore, can be used in cases where other studies involving larger quantities of protein are practically difficult or impossible. There is a distinct biological specificity in the reaction of cytochrome c with cytochrome oxidase (Yamanaka and Okunuki, 1963), and this specificity can be used to detect differences between cytochrome c molecules. These differences, manifest as different rates of reaction, form the basis on which Yamanaka (1966) established his relationships.

In the last ten years, it has become possible to establish the primary structure of many proteins. The "Atlas of Protein Sequence and Structure" for 1966 (Dayhoff and Eck, 1966) listed 184 complete sequences of proteins and related macromolecules. The same publication for 1972 (Dayhoff, 1972) lists 356 sequences, to which have been added some

150 more in a supplement for 1973 (Dayhoff, 1973).

Protein sequence information can be utilised in several ways. From the phylogenist's viewpoint, the elucidation of the primary structure of the same protein from different organisms can give information as to the history of the gene specifying that protein. If a virtually ubiquitous protein such as cytochrome c is considered, then a detailed study of the primary structures of cytochromes c from apparently distantly related extant species may give some indication of how the species have diverged during evolution. When the amino acid sequences of cytochromes c from different species are compared there is an apparent similarity of structure in the same molecule from different sources (see Appendix 8; Dayhoff, 1972). In attempting to establish phylogeny it is the differences in the molecules that are of greater interest, as these represent the degree to which the gene specifying the protein has changed during evolution.

The sequence of cytochrome c has been established for many animal, higher plant, bacterial and fungal species (see for example, Dayhoff, 1972), and to the list has

recently been added two Protozoan species (Pettigrew, 1972 and 1973). Despite the number of cytochrome sequences now known for many families of organisms, there are certain groups that have been largely neglected in terms of sequence data. These are particularly the 'lower' groups of organisms including the molluscs and invertebrates other than insects, protozoa, algae and the lower groups of plants such as Bryophytes and Pteridophytes. At the start of the present investigation no cytochrome c sequence from algae had been reported, although the sequence of ferredoxin from the green alga Scenedesmus, had been established (Sugeno and Matsubara, 1968). The Scenedesmus sequence represented the only protein sequence then known for algae, to which has recently been added the sequence of Chlorella fusca plastocyanin (Kelly and Ambler, 1973).

If a study of protein sequence data can in fact be used to determine the ancestry of the cytochrome c gene, then it is pertinent to examine groups which are considered to have a longer ancestry. These are the groups which classical taxonomists would consider to have diverged earlier in time, and are largely represented by those groups for which protein sequence data is lacking.

Dayhoff and Eck (1966) have demonstrated the use of computer analysis of homologous protein sequences as an

important indicator of taxonomic relationships in the animal kingdom. This approach was developed by Boulter *et al.* (1972) for the known cytochrome *c* sequences from higher plants and certain of the fungi. The present work demonstrates the application of these methods to the algae. The aim of the investigation was to isolate and determine the amino acid sequence of cytochrome *c* from algal sources in order to provide data from a wider range of the plant kingdom, on which to base a biochemical phylogeny. Although the cytochrome *c* sequence data for algae is non-existent, cytochrome *c* sequences for several members of the fungal kingdom are now known (see Dayhoff 1972, 1973). However, within the fungi sequence data is lacking for the Basidiomycetes. A preliminary investigation of the cytochrome *c* from a member of this group, and also from the Phycomycete fungus, *Saprolegnia*, was also undertaken.

MATERIALS AND METHODS

I. GENERAL

1. Biological Materials

Enteromorpha intestinalis (L.) was collected from a freshwater stream at Eyemouth Harbour, Berwickshire.

Porphyra umbilicalis (J. E. Agardh) thalli were collected at low tide from pure stands of the alga on rocks at St. Abbs Head, Berwickshire.

Rhodymenia palmata (Grev.) was collected firstly from St. Abbs and, subsequently, from Souter Point, Co. Durham; thalli having epiphytic Bryozoans were rejected.

Nitella sp. was collected from Lake Grasmere using a grapple from a boat as recommended by G. J. Thompson, Freshwater Biological Association, Ferry House, Ambleside.

Fucus serratus and *Laminaria* spp. were collected from Souter Point, Co. Durham.

Mushrooms (*Psalliota campestris*), in the 'cup stage', were obtained from Messrs. Crawfords Ltd., Durham City.

Saprolegnia sp. slope cultures were a kind gift from Dr. Hudson of University of Cambridge. The fungal mycelium was mass cultured in medical flats in the dark at 20°C. The mycelium was harvested after 70 h growth on two thicknesses of Whatman 3 MM filter paper on a Buchner funnel and sucked dry under mild vacuum. The mycelial cake was washed three

times with distilled water and stored at -20°C.

2. Chemicals and Reagents

With the exceptions listed below, all chemicals were obtained from British Drug Houses (BDH) Ltd., Poole, Dorset, and were of analytical reagent grade when available.

α Chymotrypsin, E.C.3.4.4.5. (three times recrystallized)

Trypsin, E.C.3.4.4.4. (twice recrystallized, salt free)

were obtained from the Worthington Biochemical Corp., Freehold, N.J., U.S.A.

Carboxypeptidase-A, E.C.3.4.2.1. (di-isopropyl-phosphorofluoridate-treated; crystalline suspension in water)

was obtained from the Sigma Chemical Co., St. Louis, Mo., U.S.A.

Sephadex G-10

Sephadex G-25

Sephadex G-50

CM-Sephadex C-50

DEAE Sephadex A-50

Blue Dextran 2000

were obtained from Pharmacia Ltd., Uppsala, Sweden.

Biogel P-30

was obtained from Bio-Rad Laboratories Ltd., London.

CM-Cellulose (CM-52)

was obtained from W. and R. Balston (Modified Cellulose) Ltd., Maidstone, Kent. (Whatman).

Amberlite CG-50 (100-200 mesh, type 1)

(Rohn and Haaw Co., Philadelphia, U.S.A.) was obtained through BDH Chemicals Ltd.

E.D.T.A. (Free acid)

Horse Heart Cytochrome c (Type VI)

were obtained from Sigma Chemical Co. Ltd., London.

Hydrazine sulphate

Tris

were obtained from Hopkin and Williams Ltd., Chadwell Heath, Essex.

Hydrazine (95%+)

was obtained from Eastman Chemicals Ltd.

Methyl Oxitol

was obtained from Shell Chemicals U.K. Ltd.

Thiodiglycol

was obtained from Koch-Light Laboratories Ltd., Colnbrook.

Ascorbic acid

was obtained from Roche Products (U.K.) Ltd., Manchester St., London.

All chemicals were used as supplied, except for phenyl-

isothiocyanate, which was vacuum distilled once before use.

3. Other Materials

Polyamide sheets were obtained from the Cheng Chin Trading Co. Ltd., Taipei, Taiwan.

Visking tubing was obtained from the Scientific Instrument Centre Ltd., Leeke St., London, W.C.1.

4. Preparation of Solutions

A. Buffer solutions:

(i) 0.05 M-Sodium phosphate, pH 8.0 for dialysis

7.80 g $NaH_2PO_4 2H_2O$

6.00 g NaH_2PO_4 anhydrous per litre was adjusted to pH 8.0 with 2 M-NaOH solution.

(ii) 300 mM-sodium phosphate, pH 7.2 for gradient elution

$Na_2HPO_4 .12H_2O$	38.70 g
$NaH_2PO_4 2H_2O$	6.55 g
Water	made to 500 ml.

(iii) 10 mM-sodium phosphate, pH 7.2 for gradient elution.

300 mM-sodium phosphate, pH 7.2, buffer was diluted 30-fold and the pH checked.

(iv) 0.2 M-sodium citrate, pH 3.0

0.2 M-citric acid (citric acid.H_2O, 42.0 g/litre) 16.4 ml

0.2 M-Na_3 citrate (Na_3 citrate. $2H_2O$, 58.8 g/litre) 3.6 ml

B. Chromatographic location reagents:

(i) Cadium - ninhydrin reagent (Heilman et al., 1957)

Solution A. 100 mg cadmium acetate
10 ml water
5 ml acetic acid
100 ml acetone

Solution A was used to prepare fresh 1% (w/v) ninhydrin solution through which the paper was passed. After air-drying the paper was heated at 60-80°C for 10 min. Coloured spots on a white background showed positive reaction. These were marked and the paper heated further to locate more slowly developing spots.

(ii) Ehrlich reagent:

Solution A: 2% (w/v) p-dimethylaminobenzaldehyde in 20% (v/v) HCl in acetone freshly prepared.

The paper was passed through Solution A. A positive reaction, given by tryptophan, showed as a purple colour. Greater sensitivity was obtained when this test followed ninhydrin staining (Easley, 1965). The pink ninhydrin spots turned colourless and a positive reaction purple colour appeared.

(iii) Platinic Iodide reagent:

Solution A: 0.002 M-platinochloric acid ($H_2PtCl_6.6H_2O$ 1 mg/ml in H_2O).

Solution B: 1.0 M-KI freshly made.

Solution C: 2 M-HCl.

The reagents were mixed just before use in the following order and proportions: 4 ml of Solution A, 0.25 ml of Solution B, 0.4 ml of Solution C and 76 ml of acetone. The paper was dipped and air-dried. A positive reaction, given by reducing sulphur compounds, showed as bleached areas on a purple background.

(iv) Sakaguchi reagent:

Solution A: 0.2% 8-hydroxyquinoline in absolute ethanol.

Solution B: 0.02 ml bromine was added to 20 ml 5% (w/v) KOH (aqueous) just before using.

The paper was sprayed well with Solution A several times without soaking, and allowed to dry. It was then sprayed with Solution B. A positive reaction, given by arginine, showed as a transient pink colour.

C. Electrophoresis marker solution:

A 0.1 M solution of arginylarginine in 1.0 M-$NaHCO_3$ was treated with an equal volume of 0.2 M-dansyl chloride in acetone. After 1 h at 37°, the mixture was diluted 1000-fold and ethanolic dansyl arginine added to a concentration of 0.1 mM.

D. Amino acid autoanalyser solutions:

(i) 0.05 M sodium citrate buffer, pH 2.875

Tri-sodium citrate. $2H_2O$	220.6 g
2.0 M-NaOH	375.0 ml
Thiodiglycol	75.0 ml
Brij 35 solution	150.0 ml
Water	13,600.0 ml

Adjusted to pH 2.875 with HCl, made up to 15 litres with deionised water, and stored at 2°C.

(ii) 0.05 M sodium citrate buffer, pH 2.875 with 10% (v/v) methanol

0.05 M sodium citrate buffer	9 parts
Methanol	1 part

(iii) 0.05 M sodium citrate buffer, pH 3.8 (Hamilton, 1962)

Tri-sodium citrate $2H_2O$	117.7 g
2.0 M-NaOH	200.0 ml
Thiodiglycol	40.0 ml
Brij 35 solution	80.0 ml
Water	7,200.0 ml

Adjusted to pH 3.8 with HCl, made up to 8 litres with deionised water, and stored at 2°C.

(iv) 0.8 M sodium citrate/chloride buffer, pH 5.0 (Hamilton, 1962)

Tri-sodium citrate. $2H_2O$	220.6 g
NaCl	526.0 g

2.0 M-NaOH	375.0 ml
Brij 35 solution	150.0 ml
Water	13,500.0 ml

Adjusted to pH 5.0 with HCl, made up to 15 litres with deionised water, and stored at 2°C.

(v) Brij 35 solution

Brij 35 (melted)	40.0 ml
Water	120.0 ml

(vi) 4.0 M-sodium acetate buffer, pH 5.5 (Technicon, 1963)

Sodium acetate (anhydrous)	328.1 g
Acetic acid	120.0 ml
Water	made to 1,200.0 ml

(vii) Ninhydrin (Technicon, 1968)

Ninhydrin	80.0 g
Methyl oxitol	6,000.0 ml
Acetic acid	340.0 ml
4 M-Sodium acetate buffer, pH 5.5	1,200.0 ml
Water	made to 12,000.0 ml

Prepared at least one day before use, and stored in a dark bottle under nitrogen. Acetic acid or NaOH used to adjust to pH 5.45-5.50.

(viii) Hydrazine sulphate solution (Technicon, 1968)

Hydrazine sulphate	3.12 g

Water	12,000 ml
Sulphuric acid	trace (approx. 10 drops)

Stored in borosilicate bottle under nitrogen.

E. Polyamide sheet chromatography solvents:

Solvent 1A - 1.5% (v/v) formic acid (Woods and Wang, 1967)

Solvent 2A - Benzene-acetic acid (9:1, v/v) (Woods and Wang, 1967)

Solvent 2B - Toluene-acetic acid (9:1, v/v)

Solvent 3A - Ethyl acetate-methanol-acetic acid (20:1:1, by vol.) (Crowshaw et al., 1967)

Solvent 3B - Butyl acetate-methanol-acetic acid (20:1:1, by vol.)

Solvent 3C - Butyl acetate-methanol-acetic acid (30:20:1, by vol.)

Solvent 3D - Ethanol-M-ammonia (1:1, by vol.)

F. Chromatography marker solution:

Dansyl-arginine, dansyl-glutamic acid, dansyl-glycine, dansyl-isoleucine, dansyl-phenylalanine, dansyl-proline and dansyl-serine, all 0.1 mg/ml in 95% (v/v) ethanol.

G. Culture medium for Saprolegnia:

Glucose	20 g
Yeast extract 'Difco'	4 g
K_2HPO_4	1 g
$MgSO_4.H_2O$	0.5 g

Water made up to 1 litre.

H. Extraction medium for Saprolegnia cytochrome c

(i) Solution A - 0.05 M-Sodium phosphate buffer, pH 7.2

$Na_2HPO_4.12H_2O$		50.0 g
$NaH_2PO_4.2H_2O$		9.4 g
Water	to	4,000 ml

(ii) Extraction medium

Solution A	1,000 ml
NaCl	60 g
Ethyl acetate	100 ml

II. PROTEIN PURIFICATION METHODS

1. Assay of Cytochrome c

(a) Qualitatively

The presence of cytochrome c in a solution was detected after reduction by a trace of ascorbic acid, by observation of the α band absorption at 550 nm with a low dispersion direct vision hand spectroscope (R. and J. Beck Ltd., London).

(b) Quantitatively

The quantity of cytochrome c was estimated spectrophotometrically using silica cells of 1 cm light path on either a Hilger & Watts Ultrascan, or a Perkin Elmer Model 402 recording spectrophotometer. Cytochrome c content was estimated from the α-absorption band by using the mammalian

cytochrome c extinction coefficient of 27.7 mM^{-1} cm^{-1}, assuming a molecular weight of 13,000 (Margoliash and Frohwirt, 1959).

2. Criteria of Purity of Cytochrome c

The purity of cytochrome c preparations was estimated spectrophotometrically from the ratios of the 280 nm, Soret (γ) and α-band absorptions of the protein in the oxidised and reduced states. These ratios were compared with those of cytochromes of known structure and purity (see Keilin, 1966; Margoliash and Schejter, 1966).

3. Dialysis

Unless otherwise stated, solutions of cytochrome c were dialysed against 10-20 vol. 0.05 M-sodium phosphate buffer, pH 8.0 for 16-40 h at 2-4°C.

4. Extraction of cytochrome c from plant material

(a) Algal thalli

Washed thalli were extracted in separate batches of 50-70 kg fresh weight. Each batch was blended in approximately 2-3 kg quantities in a Waring blender for 3 min with 1 litre of pre-chilled distilled water (2-4°C), 20 g ascorbic acid and 1 g EDTA. The pH of the total homogenate was adjusted to pH 8.0 with 1 M-tris. The homogenate (60-100 l.) was then filtered at 2-4°C through Terylene bags (Type 1481, Samuel Hill Ltd., Rochdale, Lancs.), in a 21" perforated basket centrifuge (Type 86, Thos. Broadbent and Sons Ltd., Huddersfield).

Alternatively, filtration was achieved on 27 cm Buchner funnels, using a mild vacuum (400 mm Hg) through Whatman No. 6 paper. In each case dry filter cakes were stirred with sufficient distilled water to produce a stiff slurry which was then re-filtered.

(b) Extraction of Saprolegnia cytochrome c

Approximately 100 g frozen weight of mycelium were minced in a kitchen mincer. The mycelium was then blended with sufficient extraction medium to give a final volume of approximately 350 ml. The homogenate was then stirred overnight at 2-4°C with a top drive macerator (Townson and Mercer Ltd., Croydon, England). The macerate was then filtered on a Buchner funnel through Whatman No. 6 filter paper and the cake was washed with 1 M-NaCl in 0.05 M-phosphate buffer pH 7.2. The filtrate was allowed to settle and the ethyl acetate layer removed in a separating funnel. A trace of ascorbic acid was added to the aqueous extract and the presence of cytochrome c determined by 550 nm absorption in the direct vision hand spectroscope.

The dilute cytochrome c solution was then dialysed against 0.05 M-sodium phosphate buffer pH 8.0, for chromatography on CM-Sephadex.

(c) Extraction of Cytochrome c from Mushroom

The procedure for preparing cytochrome from algal sources was used (see para. 4a), with the following

modifications:

The fresh material (40 lb batches) was deep frozen at -20° and blended dry in a Hobart Blender (Type VCM 15-3, Hobart Manufacturing, London, England). The dry powder thus obtained was then blended with water, ascorbic acid and EDTA and adjusted to pH 8 prior to filtering in the Broadbent centrifuge. Subsequent steps were as in Fig. 1.

5. Ion Exchange Chromatography on Amberlite CG-50 resin

(a) Preparation of resin in the ammonium form

Resin was converted to the K^+ form by stirring in 3-4 vol. 2 M-KOH at 80°C for 3-4 h. The resin was washed with 10-15 vol. distilled water on a sintered funnel, and then resuspended in 3-4 vol. distilled water and taken to pH 1 with conc. H_2SO_4. After 1-2 h stirring, the resin was washed on a sintered funnel with 10-15 vol. distilled water, re-suspended in 3-4 vol. distilled water and adjusted to pH 10 with conc. ammonia solution. After 1 h stirring at constant pH, the resin was washed with 10-15 vol. distilled water and resuspended as a slurry ready for use.

(b) Adsorption

(i) Column: Filtrate at pH 8.0 was passed through columns (6.5 cm x 25 cm) of Amberlite CG-50, at a flow rate of 300-400 ml/h. Approximately 15-20 l. of filtrate was passed through each individual column. The

adsorption was carried out at 2-4°C.

(ii) Batchwise: Amberlite CG-50 resin was added to the filtrate at pH 8.0 (approx. 50 g wet resin/l.) and stirred at 2-4°C for 15-20 h.

(c) Elution

This was achieved by the batch method. The resin was removed from the column and washed with distilled water by stirring and decantation until the supernatant was colourless. Resin used as in b(ii) was similarly treated, the filtrate having first been decanted. It was then stirred with approximately an equal volume of 0.5 M-NaCl and 2 M-NaOH added to maintain the pH at 8.0. After stirring for 1 h at room temperature the suspended resin was poured back into the columns and the effluent collected. The resin was washed through with 0.5 M-NaCl until the washings showed no further absorption at 550 nm when reduced with a trace of ascorbic acid; absorption was determined with a direct vision low dispersion hand spectroscope.

6. Chromatography on CM-Sephadex

Following dialysis the cytochrome c solution was passed through a column (6.5 cm x 20 cm) of CM-50 Sephadex equilibrated with 0.05 M-sodium phosphate buffer, pH 8.0 at a flow rate of 150-250 ml/h. The column was then washed with 1-2 l. of 0.05 M-sodium phosphate buffer, pH 8.0,

containing 0.1 g/l. ascorbic acid. The red cytochrome band was eluted with 0.5 M-NaCl and collected as a single fraction (200 ml). Following dialysis of the eluate, chromatography was repeated using a 3 cm x 12 cm column of CM-50 Sephadex. The cytochrome c was eluted as a single fraction in approximately 30 ml 0.5 M-NaCl.

7. Ammonium Sulphate Fractionation

The cytochrome c was kept reduced by additions of a trace of ascorbic acid and the solution maintained at pH 8.0 with dilute ammonia solution. Solid ammonium sulphate was added slowly to give approximately 10% (w/v) increases in saturation. The degree of saturation was determined using the nomogram of Dixon (1953); the effect of any salt originally present in the solution, on the degree of saturation, was ignored. Any precipitate which was formed between additions of ammonium sulphate was removed by centrifuging at 10,000 x g for 10 min (M.S.E. High-Speed 18 Centrifuge). The pellet was resuspended in a minimum volume of water and examined for the presence of cytochrome c with the hand spectroscope; if cytochrome c was absent, it was discarded. All pellets containing cytochrome c were kept and dissolved in a minimum quantity of distilled water. The saturation of the solution was increased until all the cytochrome c had been precipitated.

8. Gel filtration

The column was prepared by pouring a slurry of either Biogel P-30 or Sephadex G-50 into the column and packing under pressure. The hydrostatic head used for filtration was 20-50 cm.

9. Chromatography on CM-52 Cellulose

The column was packed by pouring a slurry of resin, equilibrated in 10 mM-sodium phosphate buffer, pH 7.2, into the column and packing by pumping through starting buffer, 10 mM-sodium phosphate, pH 7.2 (DCL Micro Pump, F. A. Hughes and Co. Ltd., Epsom). The cytochrome c sample solution, either salt-free or dialysed against the starting buffer, was then pumped onto the column followed by 2-5 ml of 1 mM-$K_3Fe(CN)_6$ in starting buffer. The column was washed by pumping through 25-50 ml buffer and then developed with a linear gradient from 10 mM-300 mM-sodium phosphate, pH 7.2, formed by using the apparatus described by Bock and Ling (1954).

10. Desalting by Gel-filtration

Pure cytochrome c preparations were concentrated on a rotary evaporator and desalted by passage through a 1 cm x 10 cm column of either Sephadex G-25 or G-10. The cytochrome c solution was then lyophilized and stored at -20°C.

A summary of the extraction and purification procedure used for the preparation of algal cytochrome c is given in Fig. 1.

FIGURE 1.

Generalized scheme for the preparation of Cytochrome c from Algae

	Fresh thalli
Extraction	Blended with ascorbic acid and EDTA
	Adjusted to pH 8.0
	Filtered
	Chromatography on Amberlite CG-50 resin
	Eluted with 0.5 M-NaCl
Concentration	Dialysed
	Concentrated on CM-Sephadex resin
	Eluted with 0.5 M-NaCl
	$(NH_4)_2SO_4$ fractionation
Purification	Gel-filtration on Bio-gel P-30
	Chromatography on CM-52 Cellulose
	Desalted on G-25 Sephadex
	Pure Cytochrome Lyophilised.

N.B. Impure cytochrome preparations were routinely recycled to improve the purity.

III. PROTEIN SEQUENCE DETERMINATION METHODS

1. Denaturation of Cytochrome c

Cytochrome c was denatured with ethanol, based on the method of Margoliash et al. (1962). Cytochrome was dissolved in water (10 mg/ml) and oxidised by the addition of 1 µl saturated $K_3Fe(CN)_6$ solution. The solution was made to 80% (v/v) ethanol by the addition of absolute ethanol, and left at room temperature overnight. The denatured protein precipitate was removed by centrifugation, washed three times by suspending it in absolute ethanol and re-centrifuging, and dried in vacuo.

2. Proteolytic Digestion of Cytochrome c by Chymotrypsin or Trypsin

Denatured cytochrome c was resuspended in water and adjusted to pH 8.0 with 25 mM-NaOH on a Radio meter TTTIC autotitrator fitted with a temperature compensator.

For chymotryptic digestion, α chymotrypsin was dissolved in water; and for tryptic digestion, trypsin was dissolved in 1 mM-HCl, immediately prior to their addition. Digestion was carried out at 37°C, the rate recorded by the titrator (Radiometer SBR2c), and digestion terminated by adjusting to pH 4·5 with acetic acid and freezing at -20°C.

3. Cyanogen bromide Cleavage of Cytochrome c

Freeze-dried cytochrome c was taken up in 70% (v/v) formic acid (5 mg/ml). An approximately 50-fold excess (w/w)

of crystalline cyanogen bromide (Eastman Organic Chemicals) was added, the reaction vessel sealed and the reaction allowed to proceed for 16 h in the dark. After cyanogen bromide cleavage, the reaction mixture was diluted approximately 10-fold with distilled water, shell frozen and lyophilized to complete dryness.

4. Peptide Purification Methods

(a) Electrophoretic separation

Peptides were separated by high-voltage paper electrophoresis at pH 6.5 (pyridine-acetic acid-water, 25:1:225 by vol.) on Whatman 3 MM paper (15 cm) in a flat plate apparatus (107 cm x 15 cm, The Locarte Co., London SE3, U.K.) at 90 V/cm for 120-150 min. Peptides requiring further separation were purified by electrophoresis at pH 1.9 (acetic acid-formic acid-water, 4:1:45 by vol.).

(b) Location of peptides on paper electrophoretograms

Guide strips cut from papers were treated with the detection reagents in the order and combinations recommended by Easley (1965).

(c) Electrophoretic mobilities of peptides

At pH 6.5 peptide mobilities were measured from dansyl-arginine relative to dansyl-arginylarginine, and at pH 1.9 from 1-dimethylaminophthalene-5-sulphonic acid relative to dansyl-arginine.

(d) Elution of peptides

Peptides were eluted with water into Pyrex screw-cap tubes (10 mm x 50 mm) and freeze-dried.

(e) Separation of haem peptides

The majority of the haem peptide was not eluted by water from the electrophoretograms, but was eluted with 20% (v/v) pyridine from the dried, water-washed paper.

(f) Column chromatography of peptides

(i) Sephadex G-50 chromatography: Sephadex G-50 resin was prepared by stirring in 70% (v/v) formic acid. The column was prepared by pouring a slurry of resin into the column and packing under a hydrostatic head of approximately 30 cm.

(ii) Sephadex G-10 chromatography: Sephadex G-10 resin was prepared by stirring in 10% (v/v) acetic acid. The column was prepared by pouring a slurry of resin into the column and packing under a hydrostatic pressure of approximately 30 cm.

5. Quantitative Amino Acid Composition of Proteins

Quantitative amino acid analyses of proteins were made using a Technicon automatic amino acid analyser, with a 133 cm column of Chromobead type A resin. The nine chamber Autograd 18L elution system was used and the recorder had a scale expander fitted. Protein (0.2-1.0 mg) was hydrolysed

at 110°C in evacuated tubes with constant boiling 5.7 M-HCl (Moore and Stein, 1963).

6. Semi-quantitative Amino Acid Composition of Peptides

An aliquot of peptide, 1-10 nmol, was dried *in vacuo* in an acid-washed Durham tube (30 x 6 mm, A. Gallenkamp Ltd., London, EC2). 50 µl of constant boiling 5.7 M-HCl was added and the tube sealed and heated at 105°C for 20 h. The acid was removed *in vacuo* over NaOH and the free peptide amino acids labelled by the dansyl method (see section 7c) but without final hydrolysis. The dansyl derivatives of the amino acid were identified by chromatography on polyamide layers (see Figs. 2 and 3).

7. Peptide Sequencing Methods

(a) N-terminal method

The dansyl-Edman procedure of Gray and Hartley (1963a) was used. Approximately 10% of the peptide material was used to identify the N-terminal amino acid at each step of the Edman degradation.

(b) C-terminal method

An aliquot of peptide was digested with carboxy-peptidase-A and the amino acids liberated then identified as their dansyl derivatives.

(c) The dansyl procedure (Gray and Hartley, 1963b)

A 1-10 nmol peptide sample was transferred to a

FIGURE 2.

Chromatography of dansyl-amino acids on polyamide thin layers.

Development was by Solvent 1A in the first dimension and Solvent 2A in the second dimension.

FIG. 2

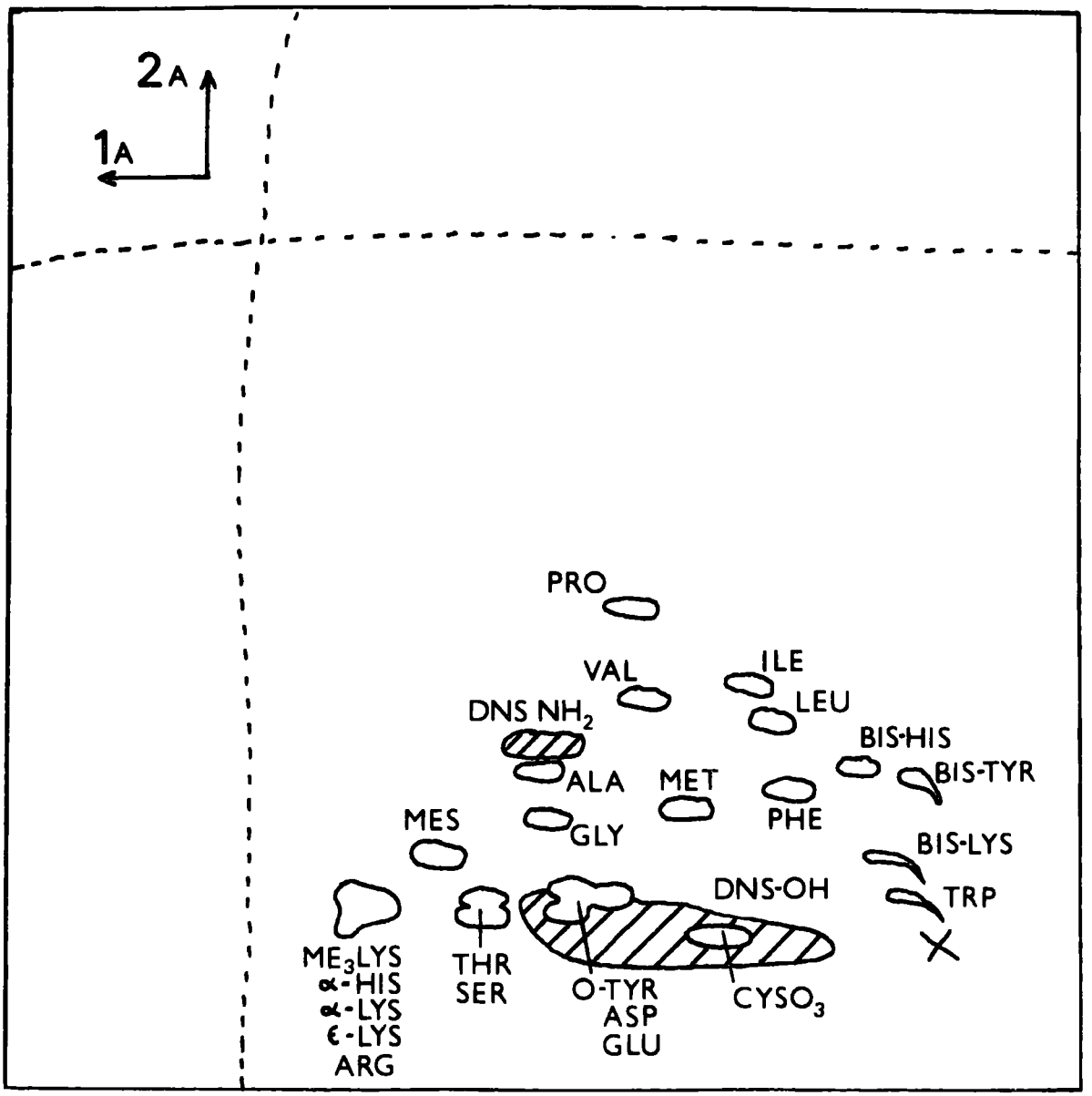
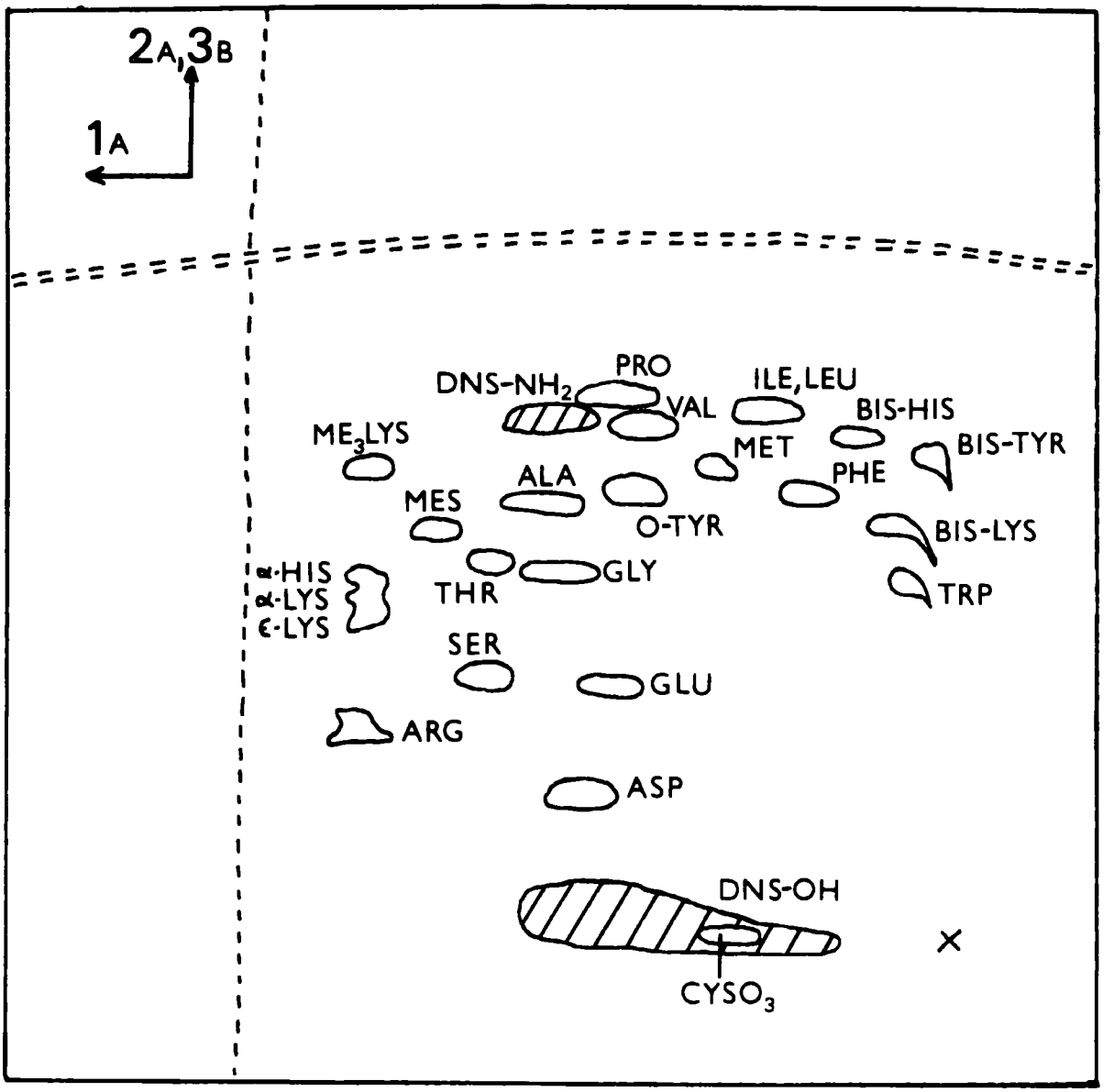

2A
1A
PRO
VAL
ILE
LEU
DNS NH2
BIS-HIS
BIS-TYR
ALA
MET
PHE
MES
GLY
BIS-LYS
DNS-OH
TRP
ME3LYS
α-HIS
α-LYS
ε-LYS
ARG
THR
SER
O-TYR
ASP
GLU
CYSO3

FIGURE 3.

Chromatography of dansyl-amino acids on polyamide thin layers.

Development was by Solvent 1A in the first dimension and Solvent 2A followed by Solvent 3B in the second dimension.

FIG. 3


2A,3B
1A
DNS-NH2
PRO
VAL
ILE,LEU
BIS-HIS
ME3LYS
MET
BIS-TYR
ALA
PHE
MES
O-TYR
BIS-LYS
α-HIS
α-LYS
ε-LYS
THR
GLY
TRP
SER
GLU
ARG
ASP
DNS-OH
CYSO3

clean Durham tube and dried *in vacuo* over NaOH and conc. H_2SO_4. (Durham tubes were cleaned prior to use by washing or by baking at 550°C (Gray and Smith, 1970)). The residue was dissolved in 10 µl of 0.1 M-$NaHCO_3$ and dried again. 5 µl of water and 5 µl of a solution containing 2.5 mg dansyl chloride/ml acetone were pre-mixed and added, the tube sealed by 'Parafilm' and the reaction stopped after 1 h at 37°C by drying *in vacuo* over NaOH and conc. H_2SO_4. Then 50 µl of constant boiling 5.7 M-HCl was added, the tube sealed and the dansyl peptide hydrolysed at 105°C for 6-14 h and the hydrolysate dried *in vacuo* over NaOH.

(d) Chromatography of dansyl derivatives

Dansyl derivatives were identified by chromatography on polyamide sheets (Woods and Wang, 1967). The sample was dissolved first in 5 µl 95% (v/v) ethanol, and then in 5 µl M-ammonia solution. The sample (1 µl aliquots, Drummond Microcaps, Shandon Scientific Co. Ltd., Willesden, London) was spotted on both sides of the sheet in a 4:1 ratio and dried under a hot air draught. 1 µl of chromatography marker solution was applied to the reverse side of the sheet. Frames of the type described by Smith (1958) were used to support up to five sheets for simultaneous chromatography.

Chromatograms were initially developed in two dimensions. Solvent 1A was run for 40 min and the sheets dried for 20-30 min in a hot air draught. After cooling they were run in

either solvent 2A or 2B for 40 min. The results of the two-dimensional separation were then recorded. Poorly resolved groups of derivatives were identified by chromatography in a third solvent either 3A, 3B or 3C, in the same direction as the second dimension for 40 min. The separations achieved are shown in Figs. 2 and 3. The chromatography of samples and standards was used to help in the identification of close pairs of derivatives.

Chromatograms were run for 15 min in solvent 3D to resolve DNS-$CySO_3$ from the DNS-OH.

Polyamide sheets were reused after washing for 1 h in acetone-M-ammonia solution (1:1, by vol.).

(e) The Edman degradation procedure (Edman, 1956; Blomback et al., 1966)

The peptide (0.1-0.5 μmol) was dissolved in 150 μl of 5% (v/v) redistilled phenylisothiocyanate in pyridine. The tube was flushed with oxygen-free nitrogen (Ilse and Edman, 1963), sealed with 'Parafilm' and heated at 45°C for 1 h. The excess reagents and volatile reaction by-products were then removed by drying *in vacuo* over NaOH and P_2O_5 at 60°C. When completely dry, 200 μl of anhydrous trifluoroacetic acid was added (Elmore and Tcseland, 1956), the tube resealed with 'Parafilm' and heated at 45°C for 30 min. Excess reagent was removed by drying *in vacuo* over NaOH at 60°C. The degraded peptide was dissolved in 200 μl

of distilled water and extracted three times with 1.5 ml of butyl acetate (Gray, 1967) and finally dried *in vacuo* over conc. H_2SO_4 and NaOH.

(f) *Proteolytic digestion of peptides*

(i) Chymotryptic digestion: Peptides were digested with 5% (w/w) α-chymotrypsin in 0.2 M-NH_4HCO_3 buffer, pH 8.4 at 37°C. Reaction was stopped by freezing and freeze-drying.

(ii) Tryptic digestion: Peptides were digested with 5% (w/w) trypsin in 0.2 M-NH_4HCO_3 buffer, pH 8.4 at 37°C. Reaction was stopped by freezing and freeze-drying.

(iii) Carboxypeptidase-A digestion: Peptides were digested with 0.2-0.5 mg of carboxypeptidase-A/mol in 0.2 M-NH_4HCO_3 buffer, pH 8.4 at 37°C (Potts *et al.*, 1962). Reaction was stopped by drying *in vacuo*.

(g) *Removal of the haem moiety from haem peptides for sequence analysis.*

The direct performic oxidation method of Nolan and Margoliash (1966) was used.

The haem peptide was taken up in 200 µl of 90% (v/v) formic acid, and either two or three 5 µl portions of 30% (w/v) H_2O_2 were added at 15 min intervals to performic oxidise the thioether link. The reaction proceeded at room temperature and was stopped, after all the red colour had been discharged, by freezing and freeze-drying.

8. Determination of Amide Residues

Amide residues were determined where possible from peptide mobilities at pH 6.5 using the method of Offord (1966). A graph was constructed giving peptide charge from mobilities relative to the dansyl markers used (Fig. 4).

9. Determination of Acetyl Group

N-terminal acetyl groups were determined as 1-acetyl-2-dansyl-hydrazine derivatives (Schmer and Kriel, 1969). The sample was dried in a Durham tube. 25 µl of 0.1 M-HCl were added and dried *in vacuo* over NaOH. 20 µl 95%+ hydrazine were added, the tube sealed and then heated at 105°C for 16 h. The sample was dried *in vacuo* and dissolved in 5 µl 0.2 M-sodium citrate buffer, pH 3.0. 5 µl of dansyl chloride (2.5 mg / ml in acetone) were added and the tube sealed by 'Parafilm' and heated at 35°C for 16 h. After drying the 1-acetyl-2-dansyl hydrazine derivative was identified by chromatography on polyamide sheets.

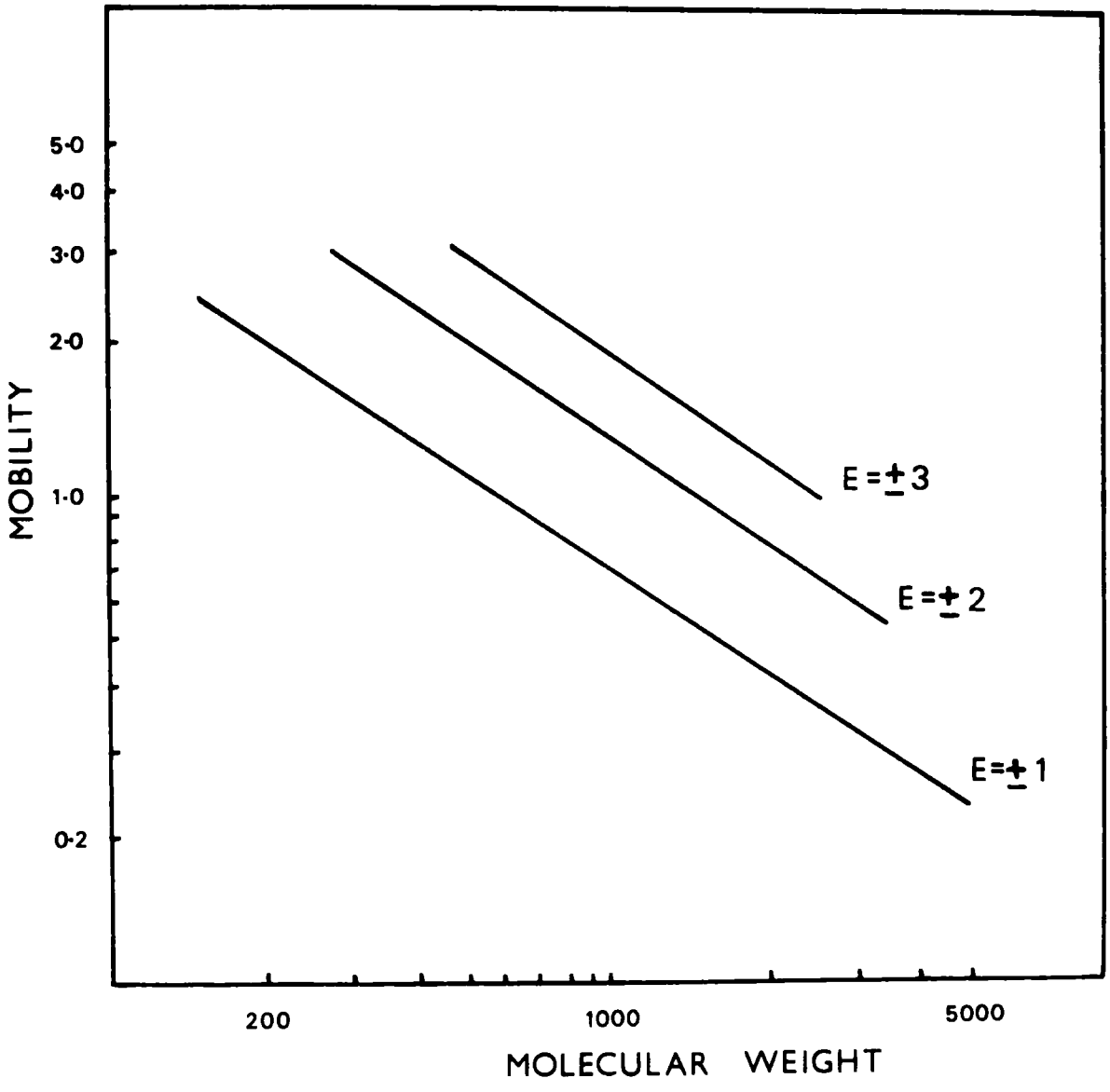
10. Nomenclature used to Describe Sequence Analysis Data

Peptides are numbered on the basis of their occurrence in the complete sequence starting at the N-terminus. Peptides prefixed by C refer to chymotryptic peptides, and those by T to tryptic peptides. Cyanogen bromide fragments are prefixed CNBr. Peptides derived by further cleavage are given a subscript to the parent peptide. All residue numbering is

FIGURE 4.

The electrophoretic mobilities of peptides relative to dansyl-Arg-Arg at pH 6.5 is plotted against their molecular weight and charge (E) at pH 6.5. Peptides containing histidine or cysteic acid do not conform to this diagram.

FIG. 4


5·0
4·0
3·0
2·0
1·0
0·2
MOBILITY
E = ± 3
E = ± 2
E = ± 1
200
1000
5000
MOLECULAR WEIGHT

given in the appropriate Figures.

Arrows (⟶) indicate positions confirmed by dansyl-Edman analysis; arrows (⇢) indicate positions tentatively assigned by dansyl-Edman analysis; and arrows (⟵) indicate positions confirmed by carboxypeptidase digestion followed by dansylation. Asterisks (*) indicate that no assignment could be made for the position during the dansyl-Edman analysis.

11. Protein Sequencer Method

A Beckman 890c sequencer, using the quadrol fast protein double cleavage program, as recommended in the Beckman (1972) operation manual, was used. PTH-amino acid derivatives were identified by gas chromatography on 10% SP400 AW Chromosorb W 100-120 mesh (Pisano et al., 1972; see also Beckman Manual, 1972), and by thin-layer chromatography on silica plates incorporating a fluorescent indicator and located by their fluorescent quenching (Jeppsson and Sjöquist, 1967). Alternatively, the hydriodic acid procedure of Inglis et al. (1971) was used to regenerate the parent amino acid from the PTH derivative which was then identified by conversion to the dansyl derivative and subsequent chromatography on polyamide sheets.

IV. CALCULATIONS BASED ON AMINO ACID SEQUENCE DATA

1. Matrix of Different Construction

(a) Amino acid differences

Sequence alignments were made relative to the

two cysteinyl residues, to which the haem group is attached. Differences due to deletions were considered as single changes.

(b) Minimum mutation differences

The method of Fitch and Margoliash (1967a) was used. Minimum mutation distances were calculated from the genetic code given by Dayhoff (1969).

2. Phylogenetic Tree Construction

Phylogenetic trees were constructed by the ancestral sequence method of Dayhoff and Eck (see Dayhoff and Eck, 1966; Dayhoff, 1969 and 1972) using a program written by A. A. Young, Computing Laboratory, University of Durham, in PL 1. Jobs were executed on the Cambridge IBM 370 computer system.

RESULTS

1. The purification of Enteromorpha Cytochrome c

(a) Preliminary Investigations

The method used initially for preparing cytochrome c from Enteromorpha was modified from that used by Laycock (1968) to extract cytochrome from mung bean seeds. Subsequently, the homogenate was adjusted directly to pH 8.0 before filtration, prior to adsorption on Amberlite resin.

After batchwise elution from the Amberlite resin and dialysis, the crude cytochrome c was concentrated on CM-Sephadex and eluted in 80 ml 0.5 M-NaCl. This was dialysed against 10 mm sodium phosphate buffer pH 7.2 for chromatography of CM-52 cellulose (Fig. 5). The cytochrome formed a dense red band at the top of the resin preceded by a lighter band and a yellowish band.

Fractions which showed an $E_{410}(ox)/E_{280}(ox)$ ratio greater than 3 were pooled. All fractions containing cytochrome c of ratio $E_{410}(ox)/E_{280}(ox)$ 1.5 to 3 were also pooled. Other fractions containing cytochrome c were pooled and designated 'tails'. The gel filtration method of Flatmark (1964) using Sephadex G-50 was then used to improve the purity ratios (Fig. 6). However, the highest $E_{410}(ox)/E_{280}(ox)$ ratios obtained were still only 3, showing purification of the cytochrome to be incomplete. Further

FIGURE 5.

Chromatography of impure Enteromorpha cytochrome c on CM-52 Cellulose.

A 2 cm x 14 cm column was used. The gradient was formed using 500 ml of each buffer. The flow rate was 25 ml/h.

_____ Absorption at 410 nm (E_{410}).

----- Absorption at 280 nm (E_{280}).

__ __ Buffer concentration.

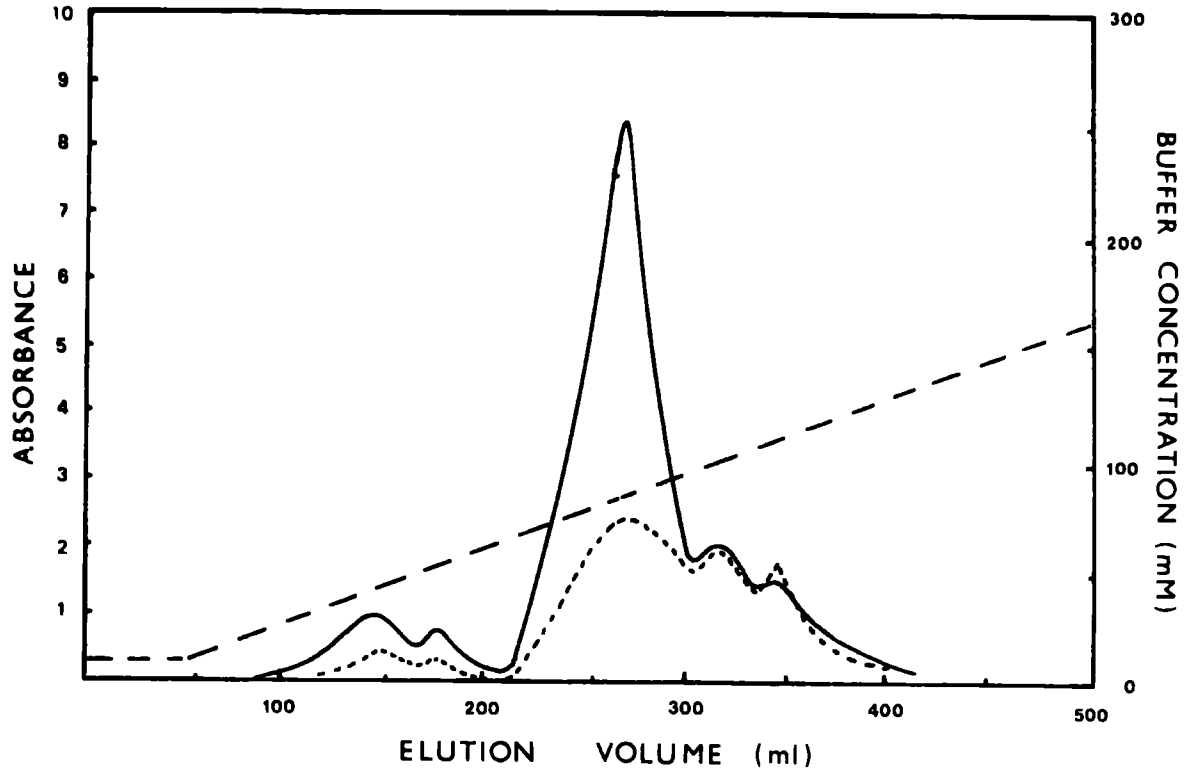
Fractions with extinction values too high to be measured directly were diluted, and values assigned pro rata.

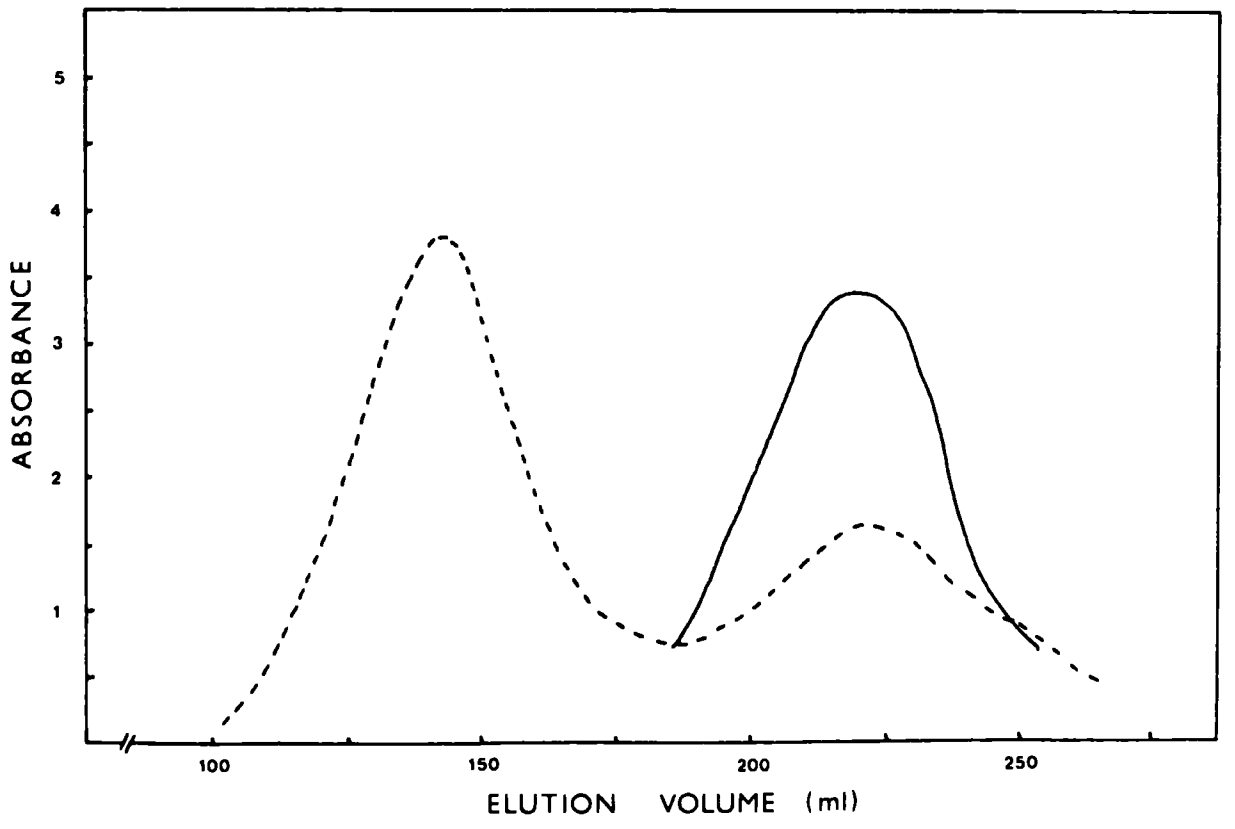
FIGURE 6.

Chromatography of partially pure Enteromorpha cytochrome c on G-50 Sephadex.

A 2 cm x 95 cm column was used. The resin was equilibrated and the column washed with 0.1 M-sodium phosphate buffer pH 7.2. The void volume (V_o) was 100 ml. Flow rate was 180 ml/h.

(See Fig. 5 for notation used).


ABSORBANCE
BUFFER CONCENTRATION (mM)
ELUTION VOLUME (ml)
10
9
8
7
6
5
4
3
2
1
300
200
100
0
100
200
300
400
500


ABSORBANCE
ELUTION VOLUME (ml)
5
4
3
2
1
100
150
200
250

chromatography on CM-52 cellulose gave purity ratios of above 5, and pooled eluates of this purity were exhaustively dialysed against distilled water.

(b) Subsequent Purification Procedure

An alternative method using ammonium sulphate fractionation was developed. After concentration on CM-Sephadex to give an approximately 0.5 mg/ml cytochrome c solution, ammonium sulphate fractionation was used to further concentrate and purify the cytochrome. Cytochrome c was present in the 65% saturation pellet and the bulk of the cytochrome was precipitated at 75% saturation. The supernatant remained pale pink up to 100% saturation and from this solution cytochrome c slowly precipitated overnight at 2-4° C. The cytochrome residues were dissolved in a minimum quantity of water and desalted and further purified on a column of Biogel P-30 (see Fig. 7). All fractions with an $E_{410}(ox)/E_{280}(ox)$ ratio of greater than 2.5 were pooled and adsorbed onto a column of CM-52 cellulose. Approximately 2 ml of 1 mM-$K_3Fe(CN)_6$ was pumped through the column to completely oxidise the cytochrome. Fractions with an $E_{410}(ox)/E_{280}(ox)$ ratio greater than 5.5 were pooled, the volume reduced by rotary evaporation, and the cytochrome desalted using Sephadex G-25. During desalting the cytochrome solution typically became cloudy and traces of the protein

FIGURE 7.

Chromatography of partially purified Enteromorpha cytochrome c on Biogel P-30.

A 3 cm x 95 cm column was used; it was equilibrated and eluted by water. The flow rate was 60 ml/h.

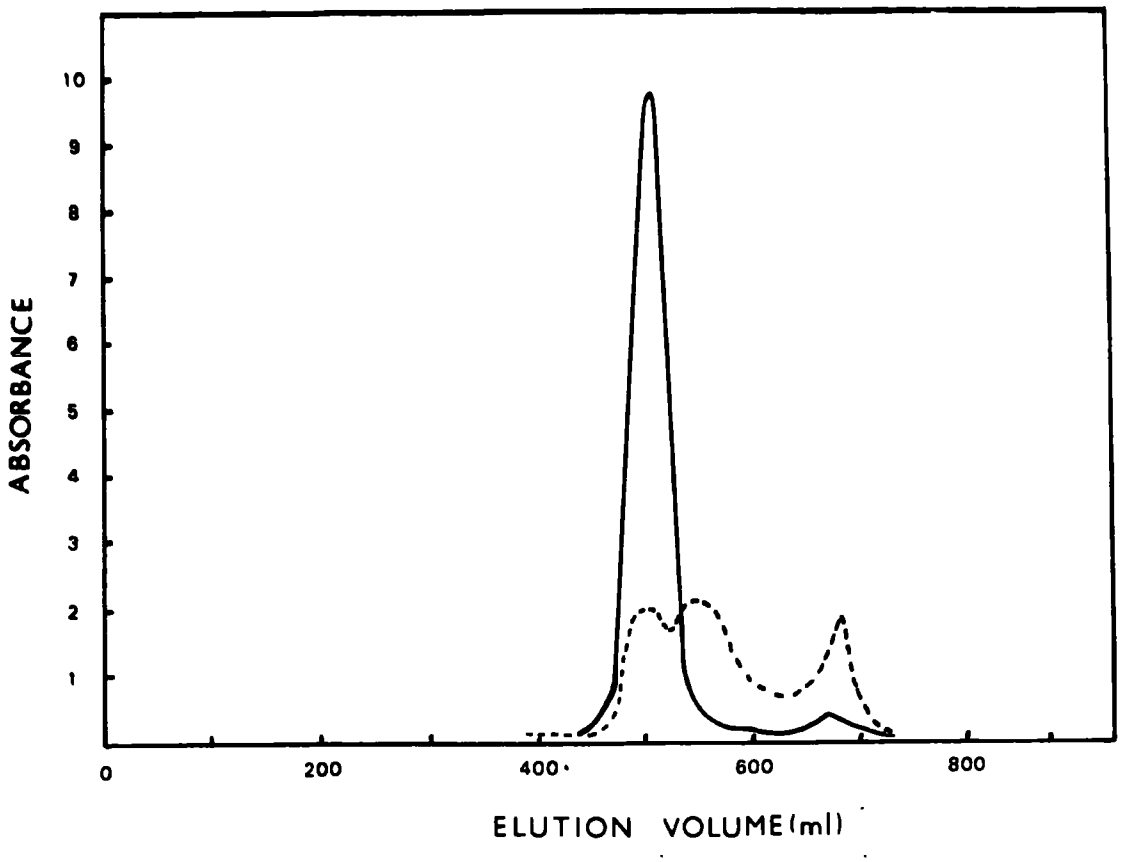
(See Fig. 5 for the notation used).

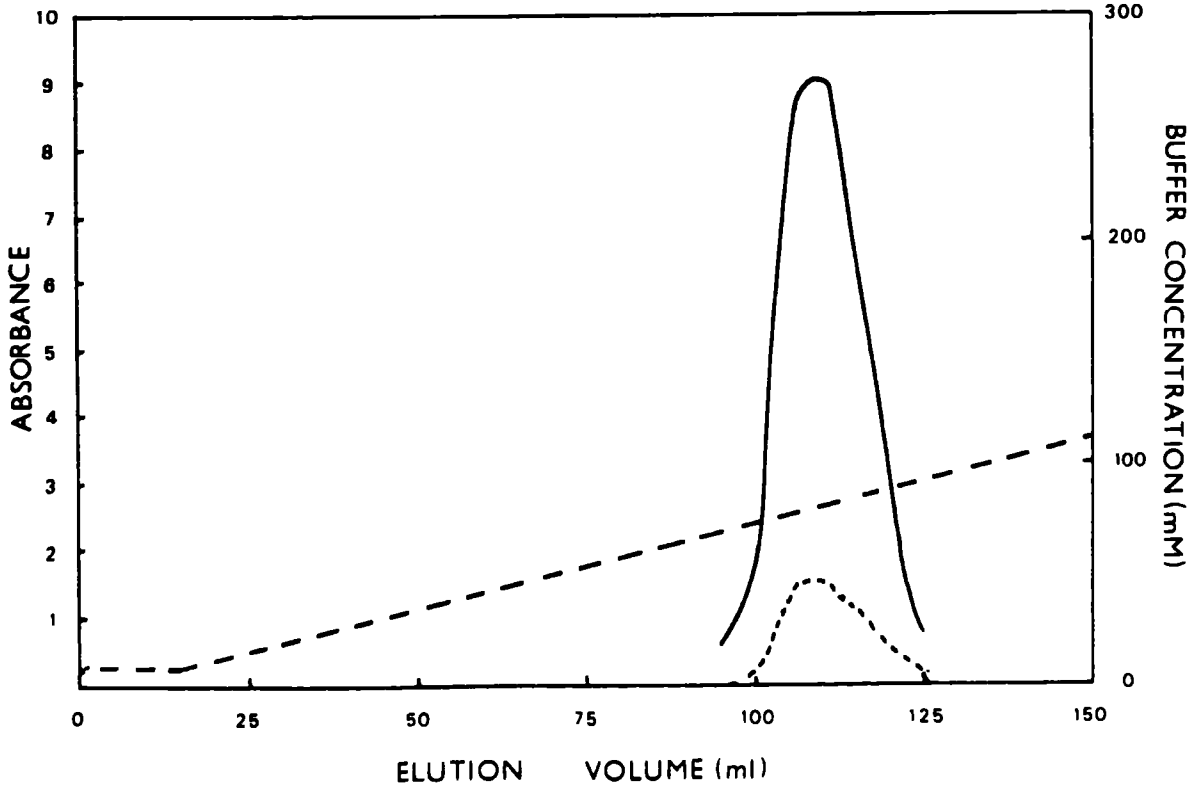
FIGURE 8.

Chromatography of partially purified Enteromorpha cytochrome c on CM-52 cellulose.

A 2 cm x 13 cm column was used. The gradient was formed using 200 ml of each buffer. The flow rate was 30 ml/h.

(See Fig. 5 for the notation used).


ABSORBANCE
10
9
8
7
6
5
4
3
2
1
0
200
400
600
800
ELUTION VOLUME(ml)


ABSORBANCE
10
9
8
7
6
5
4
3
2
1
0
25
50
75
100
125
150
ELUTION VOLUME (ml)
300
200
100
0
BUFFER CONCENTRATION (mM)

were left adhering to the sinter. These were dissolved and eluted with the salt band. The $E_{280}(ox)/E_{550}(red)$ ratio of the clear cytochrome solution was typically 1.2. Cytochrome c thus prepared was frozen in screw-cap tubes and lyophilized. A summary of the procedure developed for the extraction and purification of Enteromorpha cytochrome c is given in Fig. 1 of Methods. This method formed the basis of the procedure adopted for the preparation of cytochrome from other algal sources.

The yields and purities of Enteromorpha cytochrome c at the various steps in the procedure are given in Table 1.

Impure cytochrome fractions were routinely recycled through the appropriate stages to increase the purity.

2. The purification of Rhodymenia palmata cytochrome c

After dialysis of the Amberlite CG50 eluate the cytochrome was further concentrated by chromatography on CM-Sephadex. The columns were washed with 3 litres of buffer which caused the cytochrome band to move down the column. Before the cytochrome was eluted a white floccular precipitate that had formed at the top of the resin was removed by stirring and siphoning. Washing of the column with 0.5 M-NaCl effected partial separation between the cytochrome band and a yellowish impurity whichwas eluted separately and absorbed strongly at 280 nm. The cytochrome fraction (50 ml, 1 mg/ml

TABLE 1.

The Purification of Enteromorpha Cytochrome c

Step	Yield mg cytochrome/150 kg* Fresh weight starting material	Purity ratio $E_{410}(ox)/E_{280}(ox)$
Amberlite CG-50		1.0
CM-50 Sephadex	45 mg	1.2-2.0
$(NH_4)_2SO_4$ fractionation	Not calculated	Not measured
Biogel P30	30 mg	3.0-4.0
CM-52 Cellulose	22.5 mg	5.6-6.1
Desalting (G-25 Sephadex)	20 mg	5.6-6.1

*Approximately 15 kg dry weight.

cytochrome) was fractionated with ammonium sulphate. A copious white precipitate formed at 50% saturation and traces of cytochrome were precipitated at 55% saturation. The bulk of the cytochrome precipitated at 70% saturation and a clear cytochrome-free supernatant was obtained at 100% saturation. The 70% and 100% residues were pooled in a minimum volume of water and chromatographed on Biogel P-30 (Fig. 9). Fractions of ratio $E_{410}(ox)/E_{280}(ox)$ of greater than 4.0 were pooled and concentrated on a rotary evaporator for CM-52 cellulose chromatography (Fig. 10). Column fractions with $E_{410}(ox)/E_{280}(ox)$ ratios of greater than 4.5 were pooled and concentrated by lyophilization. The residue was redissolved in a minimum quantity of water and desalted on G-25 Sephadex. The cytochrome thus prepared had an $E_{280}(ox)/E_{550}(red)$ ratio of 1.0 and was freeze-dried in screw-cap tubes.

The yields and purities of Rhodymenia cytochrome c at the various stages are given in Table 2. The table also shows the seasonal variation in the yield of the protein.

3. Purification of Porphyra umbilicalis cytochrome c

On elution of the crude cytochrome extract from CM-Sephadex by 0.5 M-NaCl the red cytochrome band was preceded by a bright yellow impurity which was eluted from the column as a separate fraction to the cytochrome. This fraction adsorbed strongly at 280 nm. The cytochrome fraction

FIGURE 9.

Chromatography of partially purified Rhodymenia cytochrome c on Biogel P-30 resin.

A column 2 cm x 93 cm was used; it was equilibrated and eluted with water. Flow rate was 60 ml/h and the void volume (V_o) of the resin was 90 ml.

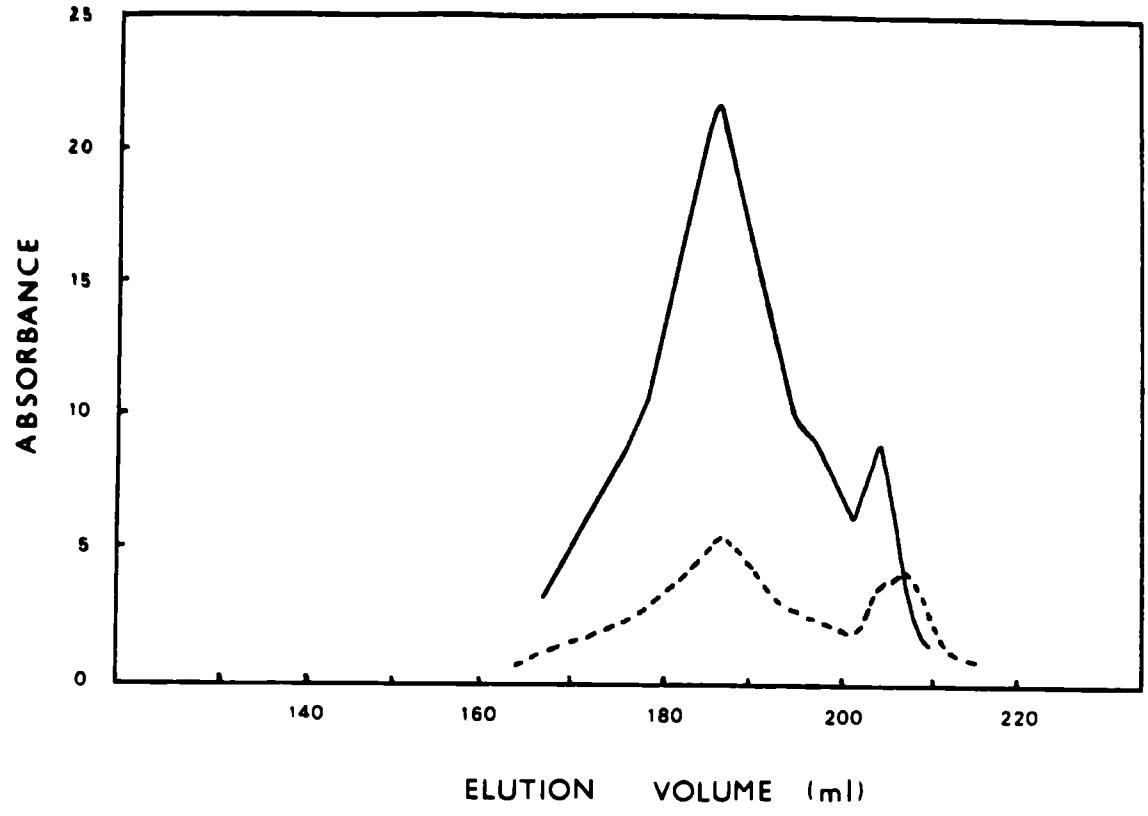
(See Fig. 5 for the notation used).

FIGURE 10.

Chromatography of partially purified Rhodymenia cytochrome c on CM-52 cellulose.

A column 2 cm x 12 cm was used; the gradient was formed using 200 ml of each buffer. The flow rate was 45 ml/h.

(See Fig. 5 for the notation used).


25
20
15
10
5
0
ABSORBANCE
140
160
180
200
220
ELUTION VOLUME (ml)

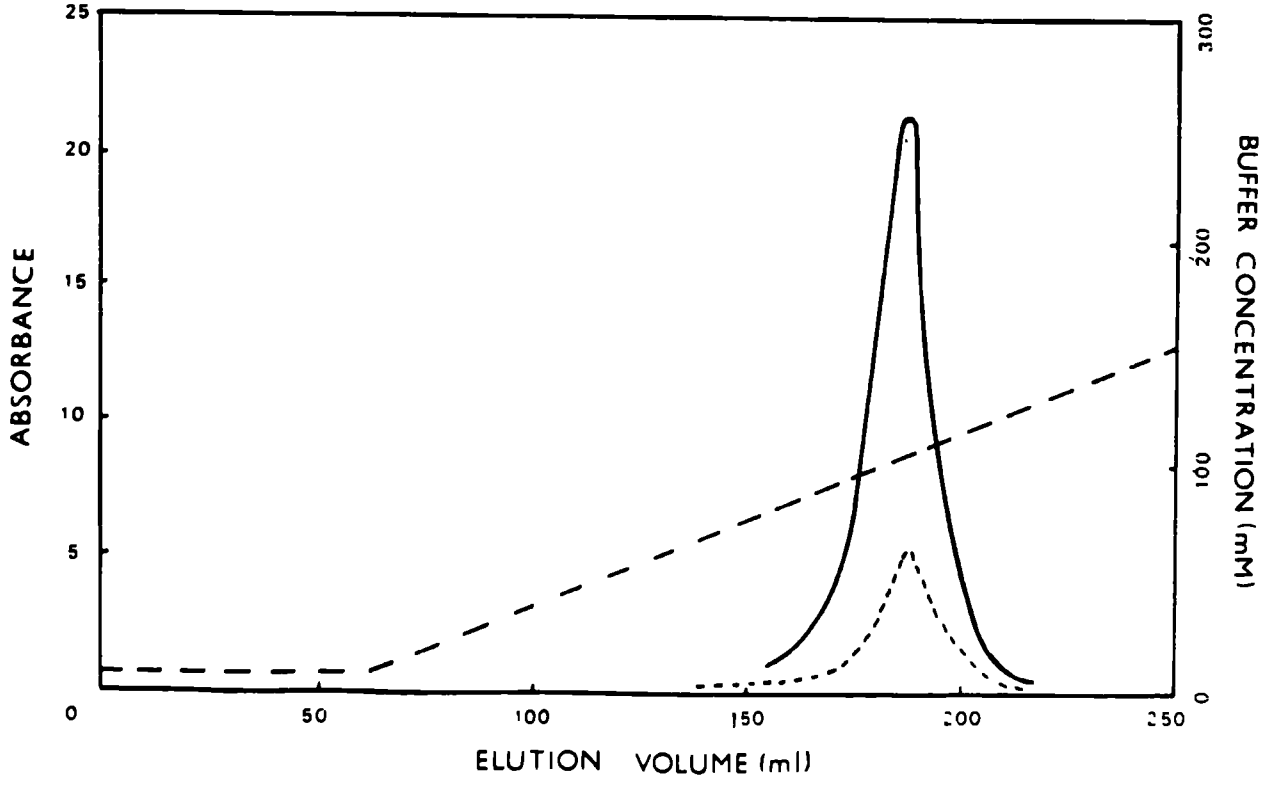

25
20
15
10
5
0
ABSORBANCE
50
100
150
200
250
ELUTION VOLUME (ml)
300
200
100
0
BUFFER CONCENTRATION (mM)

TABLE 2.

The Purification of Rhodymenia Cytochrome c

Step	Yield mg/10 kg fresh weight algae*		Purity $E_{410}(ox)/E_{280}(ox)$
	A	B	
Amberlite CG50	-	16.5	1.2-1.5
CM-Sephadex	5	15	1.6-1.8
$(NH_4)_2SO_4$ fractionation	-	-	-
Biogel P-30	4.0	12.6	3.4-3.6
CM-52 Cellulose	3.2	10	4.5-5.0
Desalting (G-25 Sephadex)	2.4	8	5.0

A - Thalli collected in May.

B - Thalli collected in September.

*Approximately 1.4 kg dry weight.

(approx. 1 mg/ml cytochrome in 0.5 M-NaCl) was fractionated with ammonium sulphate. Copious white precipitates were obtained at saturations of up to 55%. The bulk of the cytochrome was precipitated at 70% saturation. Cytochrome c remaining in solution was precipitated by adding ammonium sulphate to 100% saturation and leaving overnight at 2-4° C. Cytochrome residues were redissolved in a minimum quantity of water and chromatographed on Biogel P-30 (Fig. 11). Fractions with $E_{280}(ox)/E_{410}(ox)$ ratios of greater than 1 were pooled and the total volume reduced by lyophilization. The impure cytochrome was redissolved in a minimum volume of 10 mM-sodium phosphate buffer pH 7.2 and adsorbed on to a CM-52 cellulose column (Fig. 12). Fractions with an $E_{410}(ox)/E_{280}(ox)$ ratio of greater than 4.6 were pooled, frozen and lyophilized. The residue was taken up in a minimum volume of water and desalted on G-25 Sephadex. The $E_{550}(red)/E_{280}(ox)$ ratio was 0.93. Cytochrome c of this purity was freeze-dried in screw-cap tubes.

The yields and purities at the various stages in preparation are given in Table 3.

4. Extraction of Cytochrome c from Nitella sp.

Approximately 60 kg fresh weight of washed Nitella were blended and filtered in the usual way and the extract passed through Amberlite CG-50 resin. Elution of the resin with

FIGURE 11.

Chromatography of partially purified *Porphyra* cytochrome *c* on Biogel P-30.

A 2 cm x 92 cm column was used; it was equilibrated and eluted with water. The flow rate was 70 ml/h.

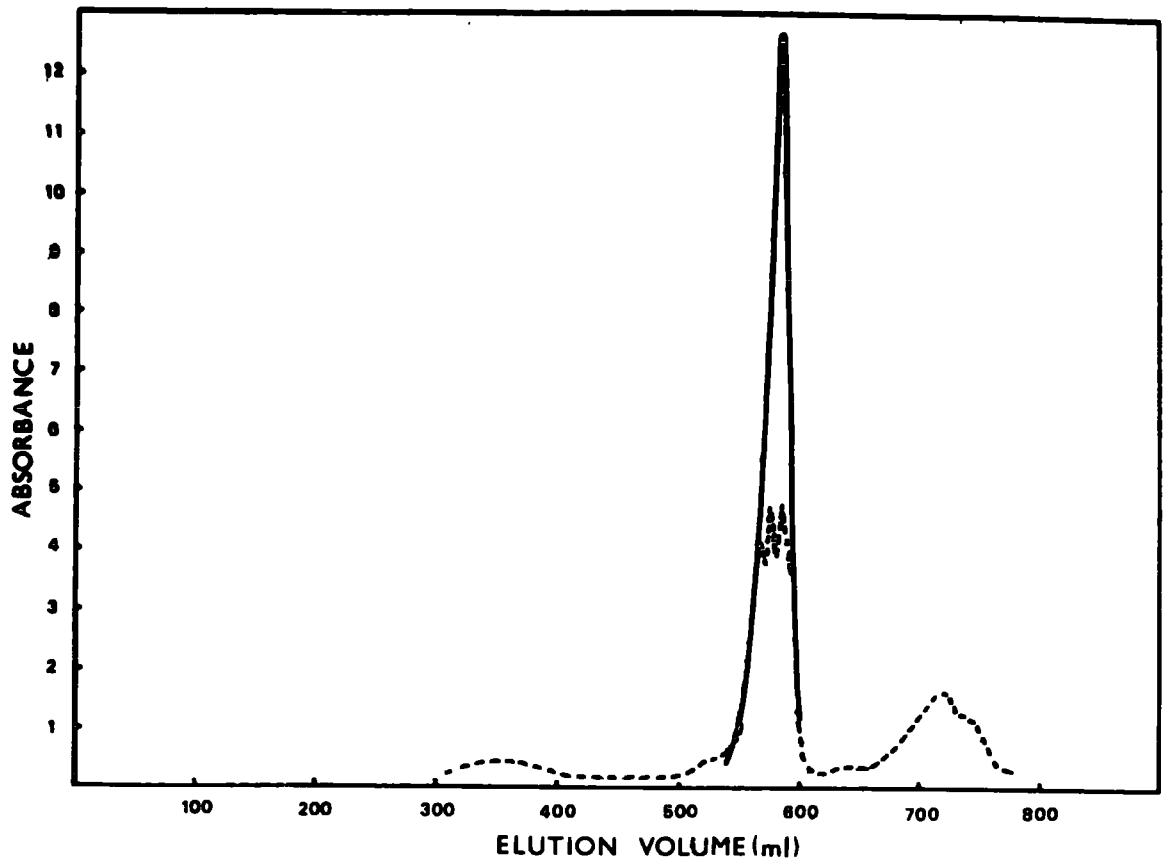
(See Fig. 5 for the notation used).

FIGURE 12.

Chromatography of partially purified *Porphyra* cytochrome *c* on CM-52 cellulose.

A column 2 cm x 12 cm was used. The gradient was formed using 500 ml of each buffer. The flow rate was 25 ml/h.

(See Fig. 5 for the notation used).


ABSORBANCE
12
11
10
9
8
7
6
5
4
3
2
1
100
200
300
400
500
600
700
800
ELUTION VOLUME (ml)

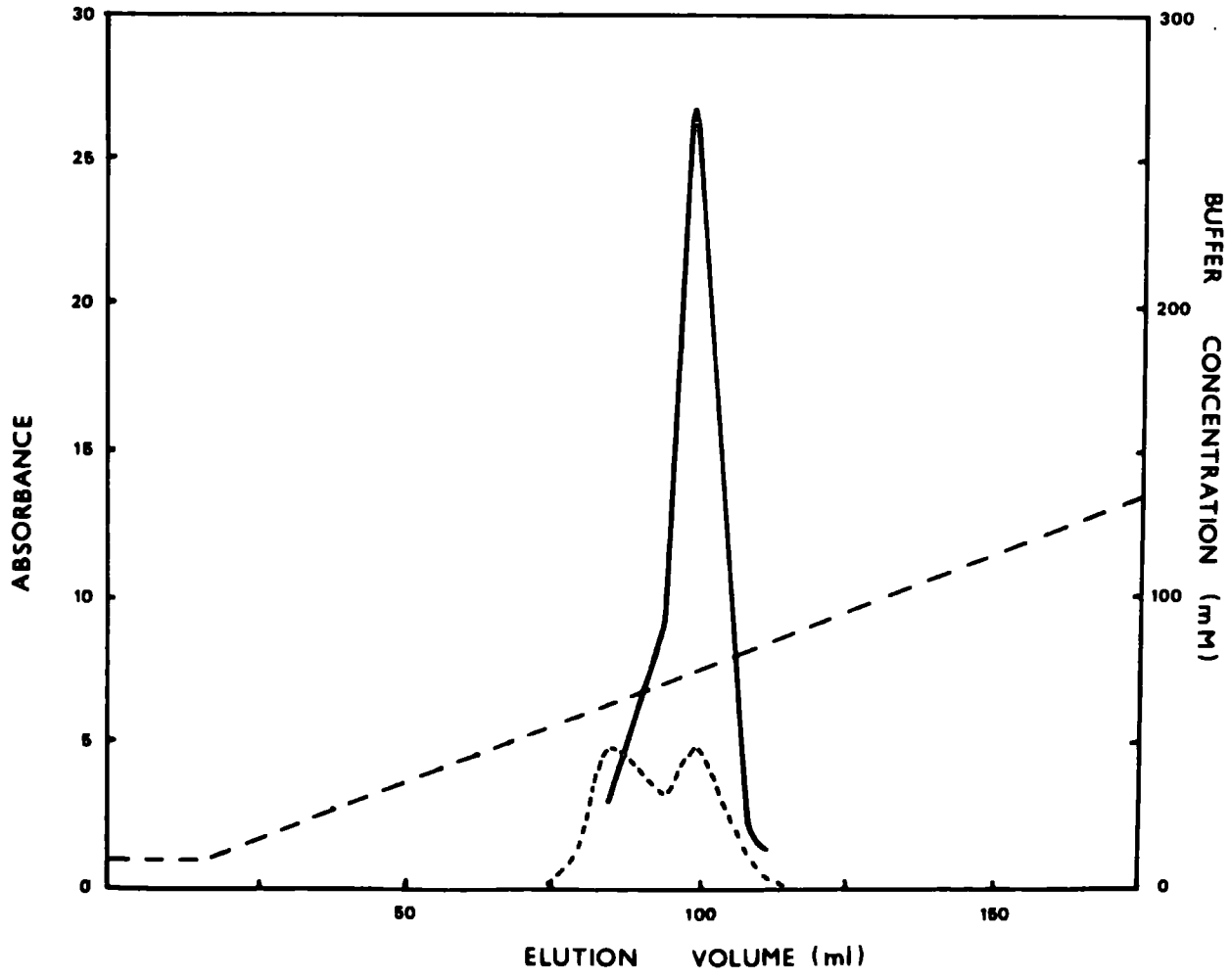

ABSORBANCE
30
25
20
15
10
5
0
BUFFER CONCENTRATION (mM)
300
200
100
0
50
100
150
ELUTION VOLUME (ml)

TABLE 3.

The Purification of Porphyra Cytochrome c

Step	Yield mg/10 kg fresh weight algae*	Purity $E_{410}(ox)/E_{280}(ox)$
Amberlite CG-50	4.5	0.1
CM-Sephadex	4.0	0.2-0.75
$(NH_4)_2SO_4$ fractionation	-	-
Biogel P-30	2.6-3.3	1.0-2.5
CM-52 Cellulose	1.06	4.7
Desalting (G-25 Sephadex)	1.0	4.7

*Approximately 1.5 kg dry weight.

Total cytochrome extracted from 150 kg fresh material was 18 mg.

0.5 M-NaCl gave an eluate which showed no detectable 550 nm absorption in the direct vision spectroscope. Further concentration of the solution on CM-Sephadex also gave no detectable cytochrome in the 0.5 M-NaCl eluate using the direct vision spectroscope.

5. Extraction of Cytochrome c from Brown Algae

Pilot extraction procedures were carried out on Fucus serratus and Laminaria hyperborea thalli. In each case much interference with slime was encountered and the Amberlite eluates gave no detectable 550 nm absorption. The Amberlite eluate of the Fucus preparation was dialysed and passed through a CM-Sephadex column. No 550 nm absorption was detected with the direct vision spectroscope at the top of the resin where typically the red cytochrome band was concentrated.

6. Extraction and Purification of Saprolegnia Cytochrome c

Pilot extractions were carried out at pH 6.0, pH 7.2 and pH 8.5, with and without the addition of 10% ethyl acetate, on fresh, frozen and acetone powdered fungal mycelium. The yields obtained from the various extractions are shown in Table 4. In each case the cytochrome preparation contained an impurity which absorbed strongly at 264 nm.

TABLE 4.

Yields of Saprolegnia* Cytochrome c

pH	Buffer	Concentration of NaCl	Yield per 100g fresh wt. +10% Ethyl Acetate	Yield per 100g fresh wt. - Ethyl Acetate	$E_{410}(ox)/E_{280}(ox)$
6.0	0.05 M-sodium phosphate	1M	0.5 mg	trace	-
7.2	0.05 M Tris-HCl	1M	1.2 mg	trace	1.0
7.2	0.05 M sodium phosphate	1M	2.5 mg	0.5 mg	1.5
8.6	0.05 M Tris-HCl	1M	0.3 mg	trace	-

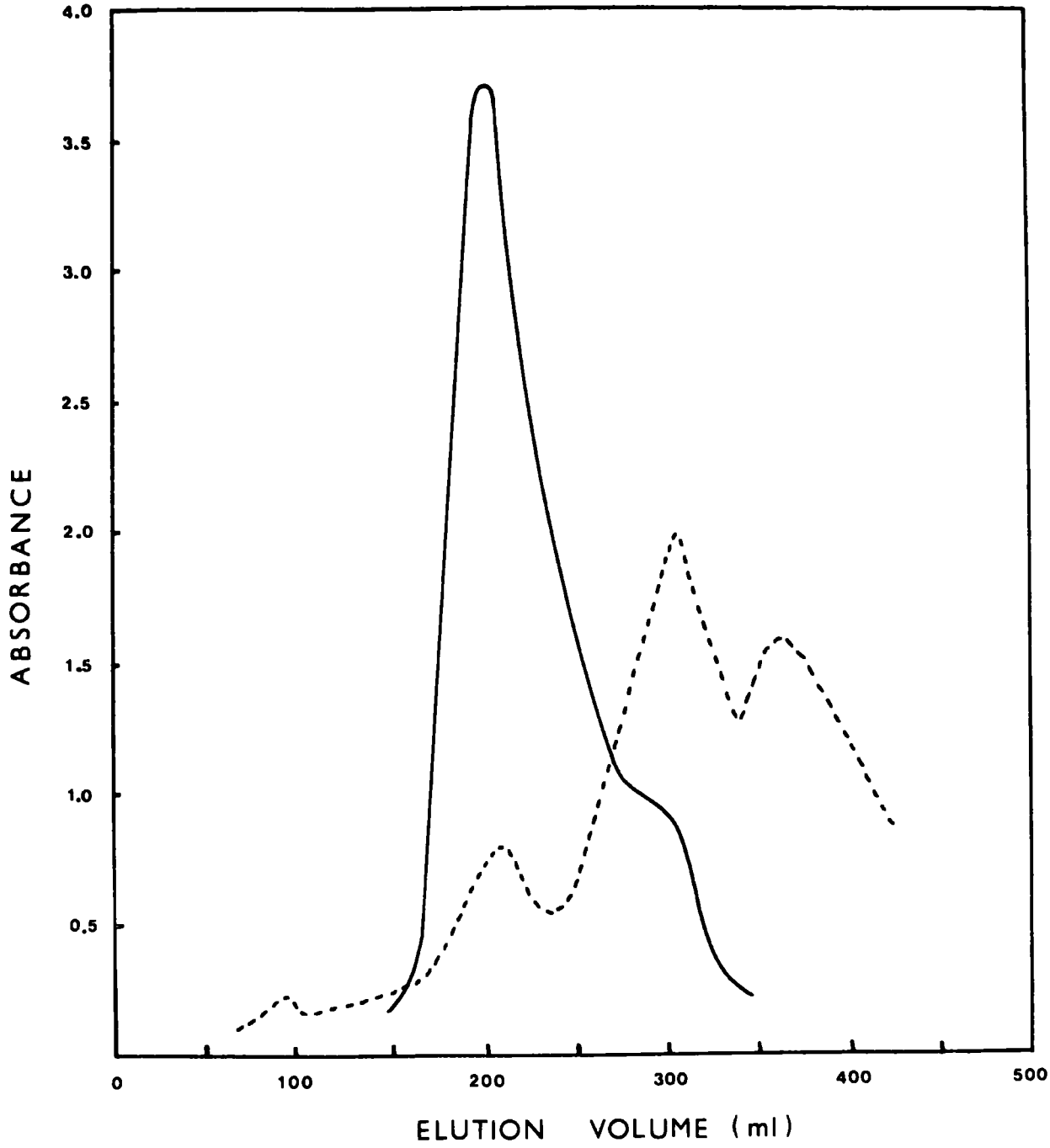
*Minced frozen mycelium.

FIGURE 13.

Chromatography of partially purified Saprolegnia cytochrome c on Biogel P-30.

A 2 cm x 94 cm column was used; it was equilibrated and eluted with water. The flow rate was 65 ml/h.

(See Fig. 5 for notation used).


4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
ABSORBANCE
0
100
200
300
400
500
ELUTION VOLUME (ml)

7. Purification of Mushroom cytochrome c

The eluates of the CM-Sephadex columns obtained from 160 lb. of fresh starting material were pooled and concentrated to give 50 ml of a 0.75 mg/ml cytochrome solution. On ammonium sulphate fractionation a reddish brown precipitate was obtained at 50% saturation. Traces of cytochrome were detected in the 55% saturation residue and the bulk of the cytochrome was precipitated at 65% saturation. The cytochrome remaining in solution was precipitated by taking to 100% saturation and leaving overnight at 2-4° C. Cytochrome residues were taken up in a minimum volume of distilled water (5 ml). Further purification was achieved by chromatography on Biogel P-30 (Fig. 14). Three cytochrome peaks were obtained. Peak 1 was further purified on CM-52 cellulose (Fig. 15). Fractions with $E_{410}(ox)/E_{280}(ox)$ ratios of greater than 3.5 were pooled and the cytochrome thus obtained subjected to cyanogen bromide cleavage.

8. The Amino Acid Sequence of Enteromorpha Cytochrome c

The amino acid composition of Enteromorpha cytochrome c is given in Table 5. The values obtained were in reasonable agreement with those calculated from the complete sequence given in Fig. 16, although the values for aspartic acid and alanine were both low, while the value for glutamic

FIGURE 14.

Chromatography of partially purified Mushroom cytochrome c on Biogel P-30.

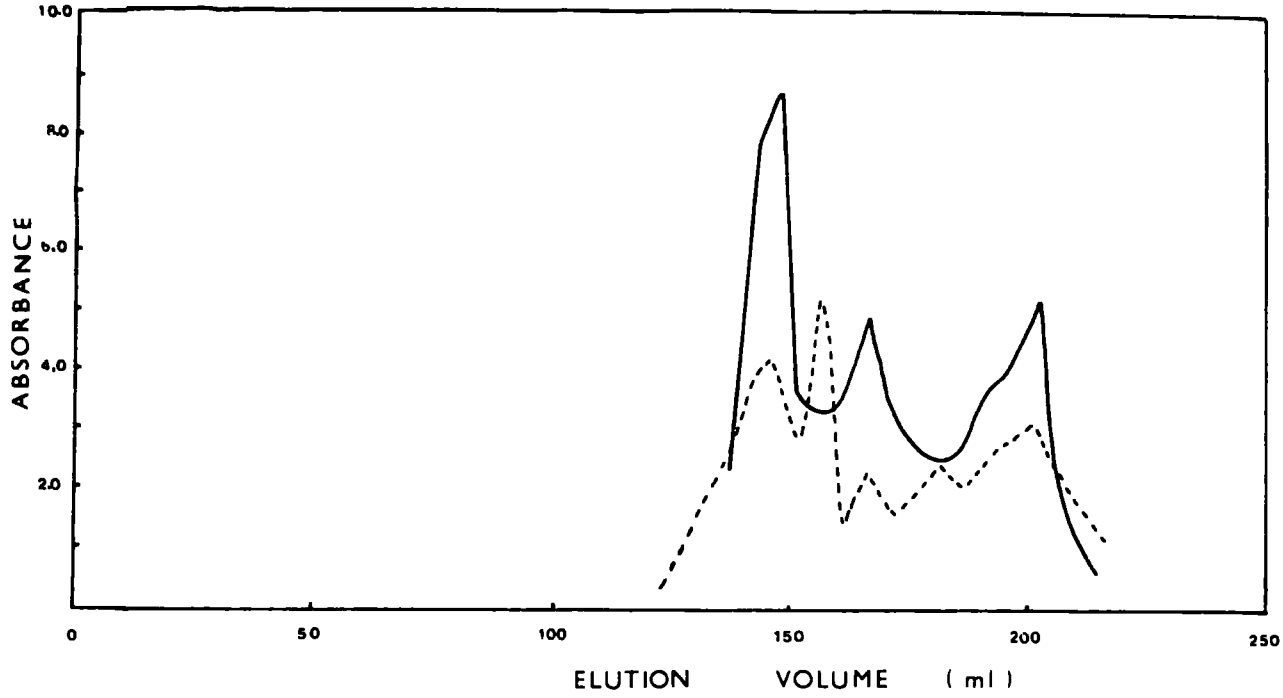
A 2 cm x 95 cm column was used; it was equilibrated and eluted with water. The flow rate was 65 ml/h. V_o for the column was 95 ml.

(See Fig. 5 for notation used).

FIGURE 15.

Chromatography of peak 1 of Mushroom cytochrome c from the Biogel P-30 column on CM-52 cellulose. A 2 cm x 11 cm column was used. The gradient was formed using 200 ml of each buffer. The flow rate was 25 ml/h.

(See Fig. 5 for notation used).


10.0
8.0
6.0
4.0
2.0
ABSORBANCE
0
50
100
150
200
250
ELUTION VOLUME (ml)

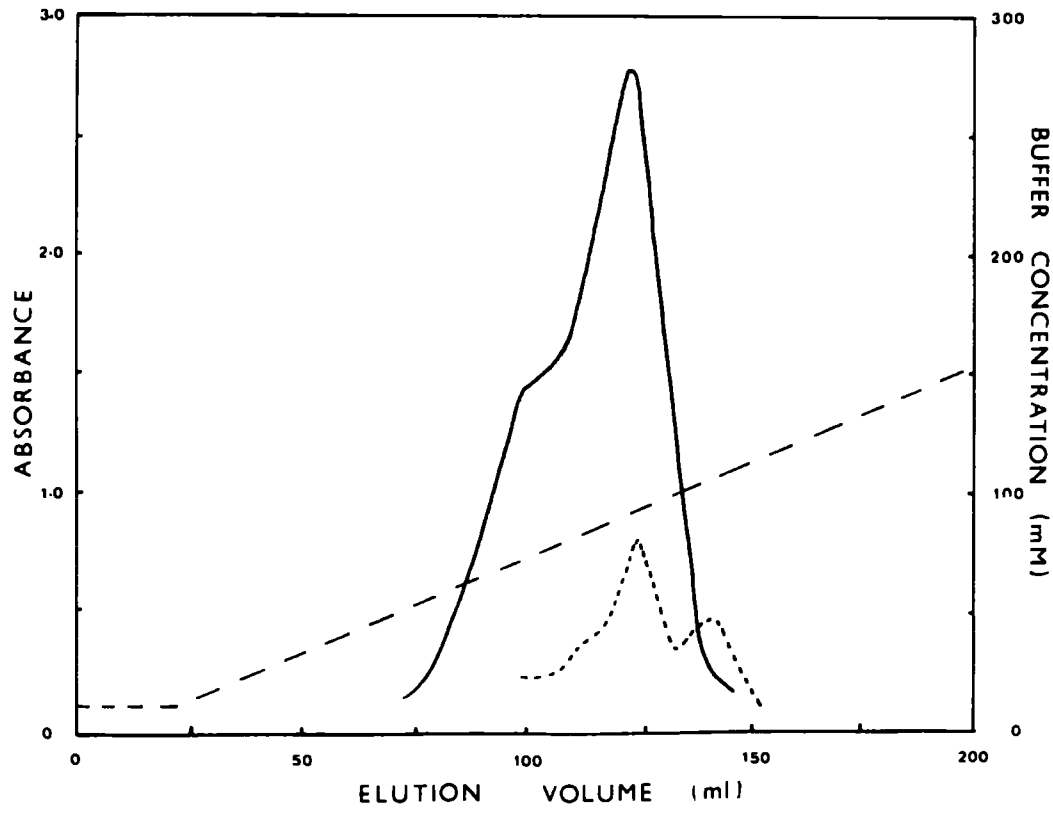

3·0
2·0
1·0
0
ABSORBANCE
300
200
100
0
BUFFER CONCENTRATION (mM)
0
50
100
150
200
ELUTION VOLUME (ml)

TABLE 5.

Amino Acid Analysis of Enteromorpha Cytochrome c

Amino Acid	Amino Acid Analysis (Integral Values)	Sequence Value
Asp	11	14
Thr	8	8
Ser	4	4
Glu	5	4
Pro	6	7
Gly	12	11
Ala	14	18
Val	2	2
Cysteine	2	2
Met	1	1
Ile	3	3
Leu	7	7
Tyr	4	4
Phe	6	6
Me_3Lys	1	1
Lys	11	12
His	2	2
Arg	2	2
Trp	*	1

The amino acid analysis was based on one 20 h hydrolysis. The integral values were calculated using the value for leucine as a standard.

*Tryptophan was not determined; the best spectral ratios obtained during the preparation indicate one residue is present.

acid is high.

The protein was readily denatured by ethanol. For chymotryptic digest approximately 6 mg of denatured cytochrome was equilibrated at 40° C; 2% (w/w) enzyme was added at zero time, an additional 1% (w/w) after 60 min. and the digestion terminated after 120 min. For tryptic digestion 5.0 mg denatured cytochrome was equilibrated at 37° C. 2% (w/w) enzyme was added at zero time, an additional 1% after 90 min., and the digestion terminated after 150 mins.

The result of the sequence analysis on chymotryptic peptides are shown in Table 6, and those of tryptic peptides in Table 7.

Peptide C1 (Acetyl-Ser-Thr-Phe)

This peptide was ninhydrin negative and was detected by the starch iodide method between peptides 11(a) and 11(b). It was purified as the only neutral peptide at pH 1.9, suggesting that the N-terminal amino group was blocked. Total hydrolysis and subsequent dansylation gave the composition (Ser, Thr, Phe). Carboxypeptidase-A digestion released phenylalanine after 2 h and phenylalanine and threonine after overnight digestion.

Peptide C2 (Ala-Asp-Ala-Pro-Pro-Gly-Asp-Pro-Ala(Lys,Gly, Ala,Lys,Ile,Phe)

This peptide was recovered only in small yield, probably due to incomplete cleavage at phenylalanine-3. The

TABLE 6.

Chymotryptic peptides of *Enteromorpha* Cytochrome *c*

Peptide	Mobility at pH 6.5	Mobility at pH 1.9	Dansyl-Edman Results
C1	-1.50	0	Does not react.
C2	0	-	Ala-Asx-Ala-Pro-Pro-Gly-Asx-Pro-Ala (see text)
C3	0	-	Haem peptide (see text)
C4	0.80	0.75	Lys-Glx-Gly-Pro-Asx-Leu-Asx
C5	0	0.60	Gly-Ala-Phe
C6	1.5	-	Gly-Arg-Thr-Ser-Gly-Thr-Ala-Ala-Gly-Phe
C7	0	0.65	Ser-Tyr
C8	0	0.7	Ser-Ala-Ala-Asx-Lys-Asn
C9	0	0.45	Lys-Thr-Ala-Asx-Trp
C10	-1.2	0.28	Asx-Glx-Asx-Thr-Leu-Tyr
C11a	-1.65	0.46	Asp-Tyr
C11b	-1.4	0.43	Asp-Tyr-Leu
C11c	1.68		Leu-Asx-Pro-Me_3Lys-Lys-Tyr
C12	1.1	0.61	Ile-Pro-Gly-Thr-Lys-Met
C13	0	0.70	Val-Phe
C14	0	0.68	Ala-Gly-Leu
C15	0.7	1.00	Lys-Lys-Pro-Glx-Asx-Arg-Ala-Asp-Leu
C16	0	0.58	Ile-Ala-Phe
C17a	0	0.82	Leu-Lys-Asp-Ala-Thr-Ala
C17b	0	0.45	Ile-Ala-Phe-Leu-Lys-Asp-Ala-Thr-Ala

TABLE 7.

Tryptic peptides of Enteromorpha Cytochrome c

Peptide	Mobility at pH 6.5	Mobility at pH 1.9	Dansyl-Edman Results
T1	-1.30	-	Does not react. Digested with chymotrypsin.
T1C1			See text.
T1C2		0.40	Ala-Asx-Ala-Pro-Pro-Gly-Asx-Pro-Ala-Lys
T2	2.2	-	Gly-Ala-Lys
T3	1.57	1.03	Ile-Phe-Lys
T4	2.50	-	Ala-Lys
T5	0	0.15	$CySO_3$-Ala-Glx-$CySO_3$-His-Thr-Val-Asx-Ala-Gly-Ala-Gly-His-Lys
T6	0.22		See text.
T7	0		See text.
T8	0		See text.
T9	1.01	0.75	Tyr-Ile-Pro-Gly-Thr-Lys
T10	0.70	0.68	Met-Val-Phe-Ala-Gly-Leu-Lys
T11	0.70	1.15	Lys-Pro-Glx-Asx-Arg
T12	0	0.55	Ala-Asx-Leu-Ile-Ala-Phe-Leu-Lys
T13	-1.7	-	Asx-Ala-Thr-Ala

positions in brackets were placed on the results of total hydrolysis, but were confirmed by their tryptic counterparts.

Peptide C3 (Lys-Ala-Lys-$CySO_3$-Ala-Glx-$CySO_3$-His-Thr-Val-Asx(Ala,Gly,Ala)Gly-His)

This was obtained by elution from the electrophoretogram by 20% (v/v) pyridine. The haem prosthetic group was removed by performic oxidation. Total hydrolysis for semi-quantitative composition analysis and N-terminal determination were carried out before dehaeming. Carboxypeptidase-A yielded histidine on short digestion and histidine with glycine on longer digestion. The glycine and alanine detected by total hydrolysis were placed by homology in the brackets as shown. Digestion of chymotryptic haem peptides with chymotrypsin after dehaeming and subsequent electrophoresis at pH 1.9 gave only weak peptide zones which gave inconclusive dansyl-Edman results. Electrophoretic mobility could not be used directly to give information on amide content because of the histidine and cysteic acid residues present.

Peptide C4 (Lys-Glx-Gly-Pro-Asx-Leu-Asx)

The electrophoretic mobility at pH 6.5 indicated that positions 36, 39 and 41 were all amides.

Peptide C5 (Gly-Ala-Phe)

This peptide gave a yellow colour with cadmium-ninhydrin reagent, suggesting the N-terminal, glycine.

Peptide C6 (Gly-Arg-Thr-Ser-Gly-Thr-Ala-Ala-Gly-Phe)

This peptide was purified from a separate digest and was purified from peptides C11(c) and C12, with which it was usually found contaminated. It gave a yellow colour with cadmium-ninhydrin reagent and was Sakaguchi positive.

Peptide C7 (Ser-Tyr)

This peptide was recovered in high yields from all chymotryptic digests.

Peptide C8 (Ser-Ala-Ala-Asx-Lys-Asn)

The mobility of this peptide indicated that position 60 was aspartic acid, assuming chymotryptic specificity for cleavage at asparagine, position 62.

Peptide C9 (Lys-Thr-Ala-Asp-Trp)

This was neutral at pH 6.5 indicating aspartic acid in position 66. The peptide was Ehrlich positive and carboxypeptidase-A digestion revealed tryptophan as the C-terminus.

Peptide C10 (Asx-Glx-Asx-Thr-Leu-Tyr)

The mobility of this peptide at pH 6.5 indicates a charge of -1 which indicated that two of positions 68, 69 and 70 were amides.

Peptide C11

This peptide had the N-terminus aspartic acid as determined by the mobilities of peptides 11(a) and 11(b).

Asparagine in position 78 is indicated by the mobility of peptides 11(a), 11(b) and 11(c).

Peptide C11(a) and 11(b) (Asp-Tyr-Leu) and (Asp-Tyr)

The mobilities of both peptides 11(a) and 11(b) indicated that position 74 was aspartic acid.

Peptide C11(c) (Leu-Asn-Pro-Me_3Lys-Lys-Tyr)

The mobility of this peptide at pH 6.5 indicated that position 78 was asparagine. The presence of ε-N-trimethyl-lysine was observed in both total hydrolysis of the peptide and during dansyl-Edman analysis.

Peptide C12 (Ile-Pro-Gly-Thr-Lys-Met)

This platinic iodide positive peptide occurred in high yields in all chymotryptic digests.

Peptide C13 (Val-Phe)

Peptide C14 (Ala-Gly-Leu)

Peptide C15 (Lys-Lys-Pro-Glx-Asx-Arg-Ala-Asp-Leu)

This peptide was Sakaguchi positive. The mobility at pH 6.5 indicated a charge of +1, showing that two of positions 97, 98 and 101 were acidic.

Peptide C16 (Ile-Ala-Phe)

Peptides C17(a) and 17(b) (Leu-Lys-Asp-Ala-Thr-Ala)

As both peptide C17(a) and C17(b) were neutral at pH 6.5 position 108 was aspartic acid. The observed C-terminal

amino acid, alanine, confirmed by dansylation without subsequent hydrolysis at Edman position -5 for peptide 11(a) and -4 for peptide 11(b), is not normally susceptible to chymotryptic cleavage, indicating that this was the C-terminal peptide for the protein.

Peptide T1

This ninhydrin positive peptide gave no N-terminus by the dansyl method. It was digested with chymotrypsin and the two fragments T1C1 and T1C2 separated at pH 1.9.

Peptide T1C1

Elution of the neutral zone gave this peptide which was ninhydrin negative. No N-terminus was given by the dansyl method but total hydrolysis revealed threonine, serine and phenylalanine. Carboxypeptidase-A digestion yielded phenylalanine on short digestion, and phenylalanine with threonine on longer digestion. Hydrazinolysis and subsequent identification of the dansyl-acetyl-hydrazide confirmed the N-terminal blocking group as acetyl.

Peptide T1C2

The mobility at pH 1.9 and of the parent peptide at pH 6.5 indicated that positions 5 and 10 were both aspartic acid.

Peptide T2 (Gly-Ala-Lys)

This peptide gave a yellow colour with cadmium-ninhydrin reagent.

Peptide T3 (Ile-Phe-Lys)

Peptide T4 (Ala-Lys)

Peptide T5 ($CySO_3$-Ala-Glx-$CySO_3$-His-Thr-Val-Asx-Ala-Gly-Ala-Gly-His-Lys)

The haem peptide was recovered in low yield after performic oxidation to remove the haem prosthetic group. The amino acid composition was determined before dehaeming. The amide content could not be determined by the electrophoretic mobility because of the cysteic acid and histidine content.

Peptide T6 (Glx-Gly-Asx-Leu-Asx-Gly-Ala-Phe-Gly-Arg)

This peptide gave the N-terminus as glutamyl and an aliquot taken for Edman degradation revealed glycine as the second residue. Digestion of the peptide with chymotrypsin gave daughter peptides which were recovered in very low yields and gave inconclusive dansyl-Edman results. The sequence was based on the amino acid composition as determined by total hydrolysis and dansylation, and comparison with the relevant chymotryptic peptides. The positions 36, 39 and 41 were all amides.

Peptide T7 (Thr-Ser-Gly-Thr-Ala-Ala-Gly-Phe-Ser-Tyr-Ser-Ala-Ala-Asn-Lys-Asp-Lys)

This peptide was obtained in low yield and was incompletely

purified at pH 1.9. Dansyl-Edman analysis gave the first two amino acids as threonine and serine but further analysis gave very weak fluorescent spots on chromatograms. No information was obtained from chymotryptic digest of this peptide and the sequence was given by evidence from the corresponding chymotryptic peptides of the whole protein.

Peptide T8 (Thr-Ala-Asx-Trp-Asx-Glx-Thr-Leu-Tyr-Asp-Tyr-Leu-Leu-Asx-Pro-Me_3Lys-Lys)

A fluorescent Ehrlich positive zone was located in the neutral region at pH 6.5. On subsequent electrophoresis at pH 1.9, no Ehrlich positive zone was observed. Dansyl-Edman analysis gave the sequence Thr-Ala-Asx- and ϵ-N-trimethyllysine was present in the total hydrolysis. The sequence was assigned by the evidence from chymotryptic peptides. The electrophoretic mobility indicated that three of positions 66, 68, 69, 70, 74 and 78 were amides.

Peptide T9 (Tyr-Ile-Pro-Gly-Thr-Lys)

Peptide T10 (Met-Val-Phe-Ala-Gly-Leu-Lys)

This peptide was platinic iodide positive and gave the N-terminus, methionine.

Peptide T11 (Lys-Pro-Glx-Asx-Arg)

The electrophoretic mobility of this peptide indicated that either position 97 or 98 was an amide.

Peptide T12 (Ala-Asx-Leu-Ile-Ala-Phe-Leu-Lys)

This peptide was neutral at pH 6.5, thus indicating that position 101 was aspartic acid.

Peptide T13 (Asx-Ala-Thr-Ala)

This peptide was acidic, the mobility indicating aspartic acid in position 108. No ϵ-dansyl-lysine was observed during dansyl analysis and dansylation without hydrolysis with subsequent chromatography after three Edman degradations revealed alanine as the C-terminus.

The overlapping chymotryptic and tryptic peptides give the sequence of Enteromorpha cytochrome c shown in Fig. 16. All residues were identified during the sequence analysis except for cysteic acid in positions 22 and 25, which were identified by peptide compositions and placed by homology with other plant and animal cytochromes c (Dayhoff, 1969).

Determination of amide positions

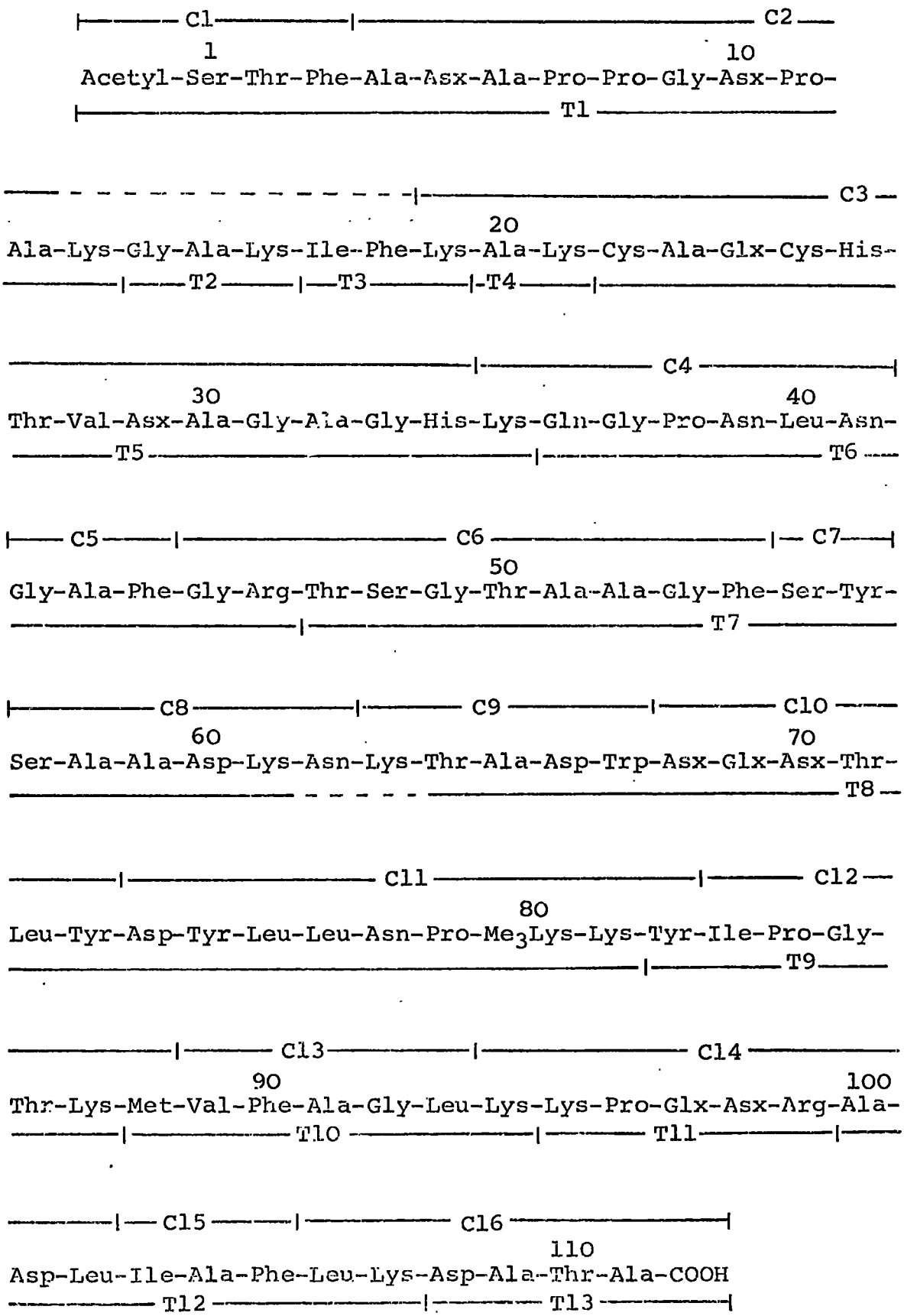
Where possible the amide content was determined by the electrophoretic mobilities of the peptides at pH 6.5, using the method of Offord (1966). The presence of nine amide residues is indicated in the sequence. Of these, four can be placed directly; these are asparagine in positions 39, 41 and 78, and glutamine in position 36.

The neutral chymotryptic haem peptide (C3) indicated one amide and one acidic residue in positions 24 and 29. Lederer et al. (1972) has suggested that the glutamyl residue in

FIGURE 16. The amino acid sequence of *Enteromorpha* cytochrome *c*.

FIGURE 16.

The Amino Acid Sequence of *Enteromorpha* Cytochrome c



position 24 is an invariant glutamine, thus indicating that position 29 is aspartic acid. Comparison of the mobility of this peptide with the basic chymotryptic haem peptide of horse heart cytochrome c (Margoliash et al., 1961), of which the amide content has been determined, gives further evidence for the substitution of alanine for lysine in position 39 as is also the case in wheat and some animal cytochromes c (see Dayhoff, 1969).

Peptide C8 indicated that either position 60 or 62 is an amide and this may represent deamidation of one of these residues as reported by Margoliash and Schejter (1966). Two amides in positions 68, 69 and 70 are indicated by the mobility of peptide C10. This peptide gave a yellow-brown colour with cadmium-ninhydrin reagent suggesting an asparagine N-terminus in position 68. An asparagine N-terminus may also enhance the mobility of a peptide by approximately 10% (Offord, 1966) and this is in agreement with the observed mobility compared to the calculated mobility of 1.0 assuming a charge of -1 for a peptide of this molecular weight. An asparagine residue in position 78 and an aspartic acid residue in position 74 account for the mobilities of peptides C11, C11(a), C11(b) and C11(c). One amide in position 97 or 98 is indicated by the mobilities of peptides T11 and T15. The observed mobility of peptide C2 was not in agreement with the proposed sequence and would

be expected to have been neutral at pH 6.5 if positions 4 and 10 were both aspartic acid. The sequence Lys-Gly-Ala-Lys-Ile-Phe-Lys (positions 13-19) is indicated by peptides T2 and T3. Peptide C2 did not therefore give information which could be used to assign amide positions.

In certain pH 6.5 electrophoretograms a great deal of streaking of the peptide zones was observed, possibly due to the presence of traces of buffer salt in the digest, and this may have caused the apparent anomalies in certain of the mobilities.

9. The Amino Acid Sequence of Rhodymenia Cytochrome c

The amino acid composition of Rhodymenia cytochrome c is given in Table 8. The values obtained are compared with those calculated from the known fragments of the sequence.

For both tryptic and chymotryptic digests 7 mg protein was used in each case. After denaturing, the samples were transferred to the digest vessel and equilibrated at 36° at pH 8.0. 2% (w/w) enzyme was added at zero time and another 1% after 90 min. Digests were terminated after 120 min by the addition of 5% (v/v) acetic acid to bring the pH to 4.0. The results of the sequence analysis on chymotryptic peptides are shown in Table 9 and those of the analysis of tryptic peptides in Table 10.

TABLE 8.

The Amino Acid Composition of Rhodymenia Cytochrome c

	Mean values 20 h hydrolysis	Mean values 70 h hydrolysis	Amino acid analysis (integral values)	Sequence values
Asp	12.5	9.4	11	8
Thr	9.3	4.6	7	7
Ser	3.5	3.1	3	2
Glu	6.1	5.8	6	5
Pro	8.0	8.6	8	9
Gly	9.4	11.4	10	10
Ala	17.3	11.8	15	14
Val	4.6	4.2	4	3
Cysteine	1.5	2.2	2	(2)
Met	2.8	1.2	2	1
Ile	7.4	4.0	5	5
Leu	8.7	5.3	6	6
Tyr	5.4	4.2	5	4
Phe	5.0	6.5	6	6
Me_3Lys	-	-	-	-
Lys	14.0	11.0	12	12
His	0.8	0.7	1	1
Arg	2.3	2.7	2	2
Trp*	-	-	-	(2)

The 'mean values' were obtained from analysis of two samples of 0.1 mg cytochrome at each hydrolysis time. The integral values are obtained as an average of the four determinations. The 'sequence values' refer only to amino acids definitely assigned to the sequence and do not include the projected sequence of the haem region.

*Trp was not determined. The best spectral ratios of the cytochrome indicate two residues are present.

TABLE 9.

Chymotryptic Peptides of Rhodymenia Cytochrome c

Peptide	Mobility at pH 6.5	Mobility at pH 1.9	Dansyl-Edman Results
C1	0	0.45	Ala-Pro-Ala-Ala-Ala-Tyr
C2	0	0.73	Ala-Asx-Leu-Lys-Gly-Asx-Pro-Thr-Lys-Gly-Ala
C3	basic	0.3	Lys-Glx-Gly-Pro-Asx-Val(Trp)
C4	0.7	-	Gly-Arg-Thr-Ser-Gly-Thr-Val-Pro-Gly-Phe
C5	0	0.65	Ala-Tyr
C6	0.75		Lys-Ala-Gly-Pro-Trp
C7	0.92		Thr-Glx-Ser-Asx-Leu-Phe
C8	-1.4		Asp-Tyr-Leu
C9	1.66		Leu-Asn-Pro-Lys-Lys-Tyr
C10	0.94	0.65	Ile-Pro-Gly-Thr-Lys-Met-Val-Phe
C11	0	0.70	Val-Phe
C12			Not detected - identified from cyanogen bromide fragment.
C13	0	0.73	Ile-Lys-Ala-Asx-Thr-Thr-Glu

TABLE 10.

Tryptic Peptides of Rhodymenia Cytochrome c

Peptide	Mobility at pH 6.5	Mobility at pH 1.9	Dansyl-Edman Results
T1	0	0.61	Ala-Pro-Ala-Ala-Ala-Tyr-Ala-Asx-Leu-Lys
T2	1.35	0.95	Gly-Asx-Pro-Thr-Lys
T3	1.83	1.27	Gly-Ala-Lys
T4	1.75		Ile-Phe-Lys
T5	0.51	0.47	Glx-Gly-Pro-Asx-Val-(*)-Gly-Leu-Phe-Gly-Arg
T6	0	0.47	Ala-Gly-Pro - digested with chymotrypsin
T6C1		0.5	Ala-Gly-Pro-Trp
T6C2		0.35	Thr-Glx-Ser-Asx-Leu-Phe
T6C3		0.8	Asx-Tyr-Leu-Leu-Asx-Pro-Lys
T7(a)	1.6		Lys-Tyr-Ile-Pro-Gly-Thr-Lys
T7(b)	0.96	0.79	Tyr-Ile-Pro-Gly-Thr-Lys
T8	0.75		Met-Val-Phe-Ala-Gly-Leu-Ile-Lys
T9	1.00	1.15	Lys-Pro-Glx-Arg
T10	0	0.70	Ala-Asx-Leu-Ile-Ala-Phe-Ala-Lys
T10(c)	-1.02		(Ala-Asx-Leu-Ile-Ala-Phe)
T11	-1.03		Ala-Asx-Thr-Thr-Glx

TABLE 11.

Cyanogen Bromide Fragments of *Rhodymenia* Cytochrome *c*

Fragment	G-50 Elution vol.	Dansyl-Edman Results
1	60 ml	Ala- (Digested with trypsin)
2	73 ml	Val-Phe-Ala-Gly-Ile-Lys-Lys-Pro-Glx-Glx-Arg-Ala

Cyanogen bromide cleavage of Rhodymenia cytochrome c

6 mg protein were used for cleavage. The freeze-dried residue was taken up in 1 ml 70% (v/v) formic acid and chromatographed on a 1 cm x 110 cm Biogel G-50 column. The flow rate was 8 ml/h and fractions of volume 3 ml were collected. 50 ml aliquots were taken for N-terminal amino acid analysis by dansylation. Two major fragments were identified (Table 11), one with the N-terminus alanine, at elution volume = 60 ml and one with N-terminus valine, eluting after 78 ml.

The second fragment was subjected directly to dansyl-Edman analysis and gave the sequence Val-Phe-Ala-Gly-Ile-Lys-Lys-Pro-Gly-Glx-Arg-(-), placing it as position 27 et seq. of fragment III of the sequence (Fig. 17).

The first fragment (N-terminus alanine) was digested with 5% (w/w) trypsin. Resulting tryptic fragments were separated by paper chromatography.

Weakly cadmium-ninhydrin positive zones were eluted but these gave no clear sequence data on dansyl-Edman analysis.

Sufficient information was obtained from overlapping chymotryptic, tryptic and cyanogen bromide peptides to give three major fragments representing approximately 90% of the total sequence. These are shown in Fig. 17.

The N-terminal region up to position 21 of Fragment I was confirmed by data obtained from a Beckman Automatic Sequencer, and this also gave position

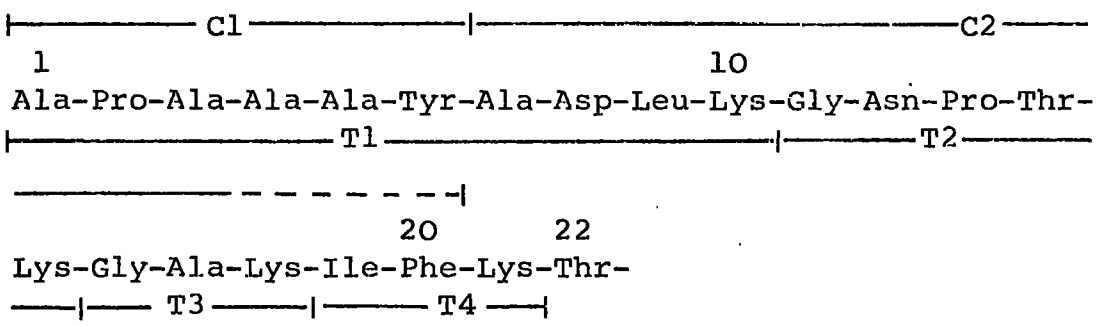
FIGURE 17.

The partial amino acid sequence of Rhodymenia cytochrome c.

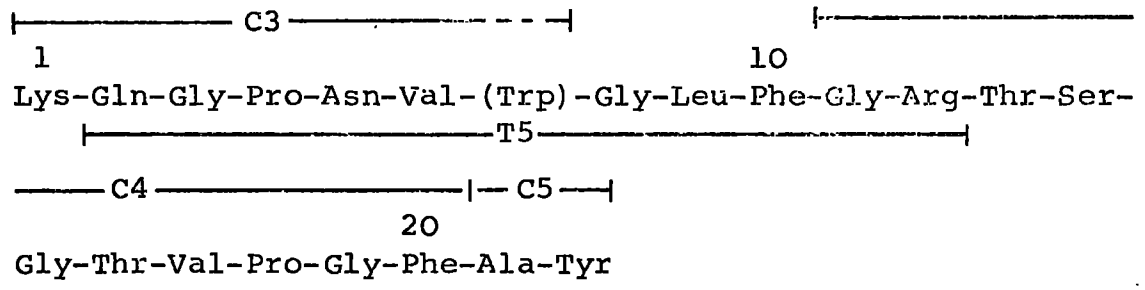
FIGURE 17.

The Amino Acid Sequence of Three Fragments of *Rhodymenia* Cytochrome *c*

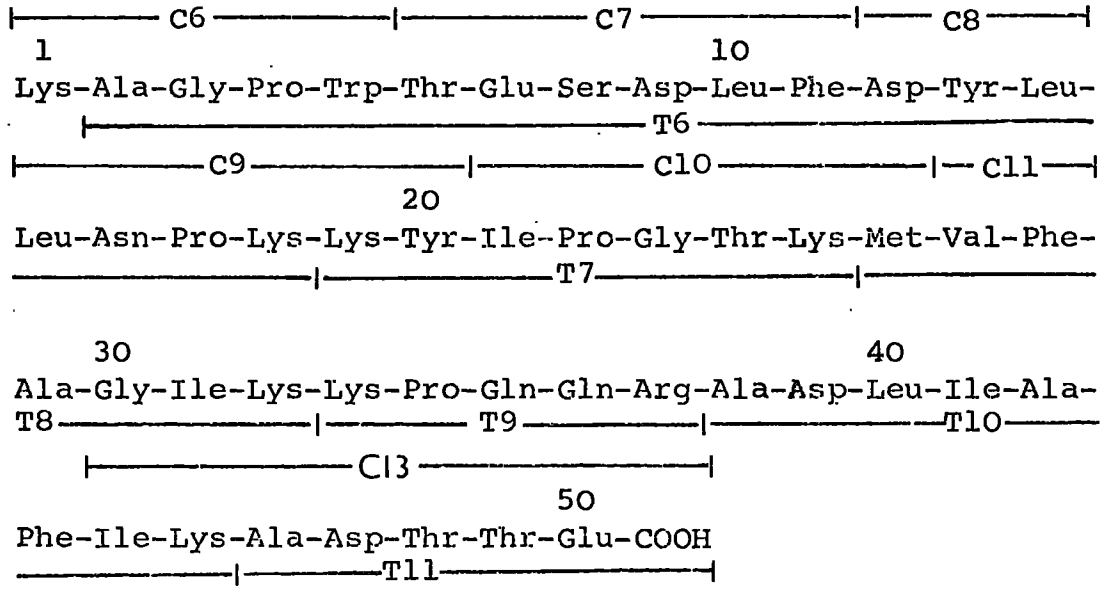
Fragment I



Fragment II



Fragment III



The alignment of these Fragments by homology with other cytochromes *c* is shown in Appendix 8.

22 as threonine.

Separation of peptides on Biogel G-10

A 1 cm x 110 cm column of Biogel G-10 was used to separate the peptides of a chymotryptic digest. The flow rate was 8 ml/h and fractions were tested for peptide material by removing samples for N-terminal amino acid analysis by the dansyl method. These results were used to pool the fractions to give two samples which were then subjected to electrophoresis in the usual way.

Peptide C1 (Ala-Pro-Ala-Ala-Ala-Tyr)

This neutral peptide was recovered in high yields from all chymotryptic digests.

Peptide C2 (Ala-Asx-Leu-Lys-Gly-Asx-Pro-Thr-Lys-Gly-Ala (Lys-Ile-Phe))

This peptide was recovered from the neutral zone and was purified at pH 1.9.

Peptide C3 (Lys-Glx-Gly-Pro-Asx-Val(Trp)

This weakly Ehrlich positive peptide was recovered in low yield from the zone of mobility 0.6-0.75 at pH 6.5, this mobility indicating that positions 2 and 4 of Fragment II are both amides.

Peptide C4 (Gly-Arg-Thr-Ser-Gly-Thr-Val-Pro-Gly-Phe)

This peptide was Sakaguchi positive.

Peptide C5 (Ala-Tyr)

Peptide C6 (Lys-Ala-Gly-Pro-Trp)
positive

This Ehrlich/peptide was purified at pH 6.5. Carboxypeptidase-A digestion yielded tryptophan on short digestion and tryptophan with a trace of proline on prolonged (overnight) digestion. Dansyl-Edman analysis gave the rest of the sequence.

Peptide C7 (Thr-Glx-Ser-Asx-Leu-Phe)

The mobility of this peptide indicated that one of positions 7 and 9 of Fragment III was an amide.

Peptide C8 (Asp-Tyr-Leu)

The mobility of this peptide indicated that position 12 of Fragment III was aspartic acid.

Peptide C9 (Leu-Asn-Pro-Lys-Lys-Tyr)

The mobility of this peptide indicated that position 16 of Fragment III was an amide.

Peptide C10 (Ile-Pro-Gly-Thr-Lys-Met)

This platinic iodide positive peptide was obtained in high yields from all chymotryptic digests.

Peptide C11 (Val-Phe)

Peptide C12

Only tentative evidence of this peptide was obtained. The sequence was determined by the corresponding tryptic peptides and cyanogen bromide Fragment II.

Peptide C13

This peptide was purified at pH 1.9. Dansylation without subsequent hydrolysis after six Edman degradation steps yielded glutamic acid, showing this to be the C-terminal peptide of the protein. The mobility thus placed position 47 of Fragment III as asparagine.

Peptide T1 (Ala-Pro-Ala-Ala-Ala-Tyr-Ala-Asp-Leu-Lys)

This neutral peptide was purified at pH 1.9.

Peptide T2 (Gly-Asx-Pro-Thr-Lys)

The mobility of this peptide at pH 6.5 indicated that position 12 of Fragment I was an amide.

Peptide T3 (Gly-Ala-Lys)

Peptide T4 (Ile-Phe-Lys)

Peptide T5 (Glx-Gly-Pro-Asx-Val-(Trp)-Gly-Leu-Phe-Gly-Arg)

This peptide was Ehrlich positive and a tryptophan residue was tentatively assigned as shown. The mobility of this peptide indicated that both positions 2 and 6 of Fragment II were amides.

Peptide T6 (Ala-Gly-Pro-(Trp))

This Ehrlich positive peptide was digested with α-chymotrypsin. Three peptides were separated at pH 1.9:-

T6C1 (Ala-Gly-Pro-Trp). This was weakly Ehrlich

positive. Carboxypeptidase-A released tryptophan and the rest of the sequence was determined in the usual way.

T6C2 (Thr-Glx-Ser-Asx-Leu-Phe).

T6C3 (Asx-Tyr-Leu-Leu-Asx-Pro-Lys).

Peptide T7(a) and 7(b)

Dansyl-Edman analysis gave the two sequences.

Peptide T8

This platinic iodide positive peptide was purified at pH 6.5.

Peptide T9 (Lys-Pro-Glx-Glx-Arg)

The mobility of this peptide indicated that one of the two glutamyl residues was an amide.

Peptide T10 (Ala-Asx-Leu-Ile-Ala-Phe-Ile-Lys)

The mobility of this peptide indicated that aspartyl residue in position 39 of Fragment III was aspartic acid.

Peptide T10(c) (Ala-Asx-Leu-Ile-Ala-Phe)

This peptide represented the only detectable trace of chymotryptic activity in the tryptic digests.

Peptide T11 (Ala-Asx-Thr-Thr-Glu)

The only remaining amino acid after four Edman degradation steps was found to be glutamic acid, indicating that this was the C-terminal peptide. The mobility thus placed position 47 of Fragment III as asparagine.

Additional Peptides. (a) (Gly-Leu-Ile-Asx-Ala-Tyr)

(b) (Gly-Leu-Ile-Asx)

These two peptides were detected and characterized in several chymotryptic digests. They were neutral at pH 6.5, indicating that the aspartyl residue was asparagine, also shown by chymotryptic specificity, and had the mobility at pH 1.9 of 0.45 and 0.50 respectively. No corresponding overlapping tryptic peptides were detected and it was not possible to assign these peptides to the Rhodymenia sequence.

The Haem Region

Despite the several methods employed to obtain and purify peptides no clear information was obtained about the haem region. Both tryptic and chymotryptic haem residues were neutral at pH 6.5 as determined by their positions on electrophoretograms. Total hydrolysis and subsequent dansylation of chymotryptic haem zones gave the composition (Pro, Lys,His,Ala,Gly,Glu,Thr,Ser,Cys). After dehaeming in the usual way dansyl-Edman analysis gave no definitive results. N-terminus analysis of tryptic haem residues showed alanine to be the N-terminal amino acid of the haem peptide. Results obtained from the sequenator ,indicate that threonine (position 22) would be expected as a tryptic N-terminus.

The sequenator data also gave no information as to the sequence of the haem region.

10. The partial sequence of Porphyra cytochrome c

The results of the amino acid analysis of Porphyra cytochrome c are given in Table 12.

Preliminary chymotryptic and tryptic digests were used to determine approximately 50% of the total sequence.

The results of the dansyl-Edman analysis of chymotryptic fragments are given in Table 13, and those of tryptic fragments in Table 14.

Where possible the sequence was determined by the overlapping chymotryptic and tryptic fragments, otherwise homology with other cytochromes c was used to determine the position of the fragments in the sequence, shown in Fig. 18.

The alignment of the known sequence with other cytochromes c is shown in App. 8. N-terminus determination of the native protein indicated the presence of alanine, the protein therefore not possessing an N-terminal blocking group.

An acidic chymotryptic peptide, mobility -0.79 at pH 6.5, gave a strong alanine N-terminus and on further purification at pH 1.9 mobility 0.34, gave the sequence Ala-Asx-(Pro)Asx- but was recovered in insufficient yield to obtain further data.

TABLE 12.

Amino Acid Analysis of *Porphyra* Cytochrome *c*

Amino Acid	Integral Value
Asp	11
Thr	7
Ser	6
Glu	8
Pro	6
Gly	10
Ala	12
Val	3
Cys	1.5
Met	2.5
Ile	8
Leu	8
Tyr	5
Phe	5
Me_3Lys	-
Lys	12
His	1.5
Arg	3
Trp	2*

*Trp was not determined; the best spectral ratios indicate 2 residues present. The analysis was based on one 20 h. hydrolysis and the integral values obtained using the leucine value as a standard.

TABLE 13.

Chymotryptic Fragments of Porphyra Cytochrome c

Peptide	Mobility at pH 6.5	Mobility at pH 1.9	Dansyl-Edman Results
C1	0	0.73	Gly-Leu-Phe
C2	0.3-0.63	0.85	Gly-Arg-Thr-Ser-Gly/Asx-Ser-Val
C3	0	0.67	Ser-Tyr
C4	0	0.63	Lys-Ala-Gly-Glx-Trp
C5	-2.22		Gly-Glx-Asx-Thr-Leu-Gly-Glx-Tyr
C6	1.8		Leu-Leu-Asx-Pro(Lys)-(Lys)-Tyr
C7	1.06	0.79	Ile-Pro-Gly-Thr-Lys-Met
C8	1.06	0.82	Val-Phe-Pro-Gly-Leu
C9	0	0.50	Ile-Ala-Phe
C10(a)	1.06	0.85	Leu-Lys-Gly-Ala-Thr-Ala
C10(b)	1.49		Lys-Gly-Ala-Thr-Ala

TABLE 14.

Tryptic Fragments of Porphyra Cytochrome c

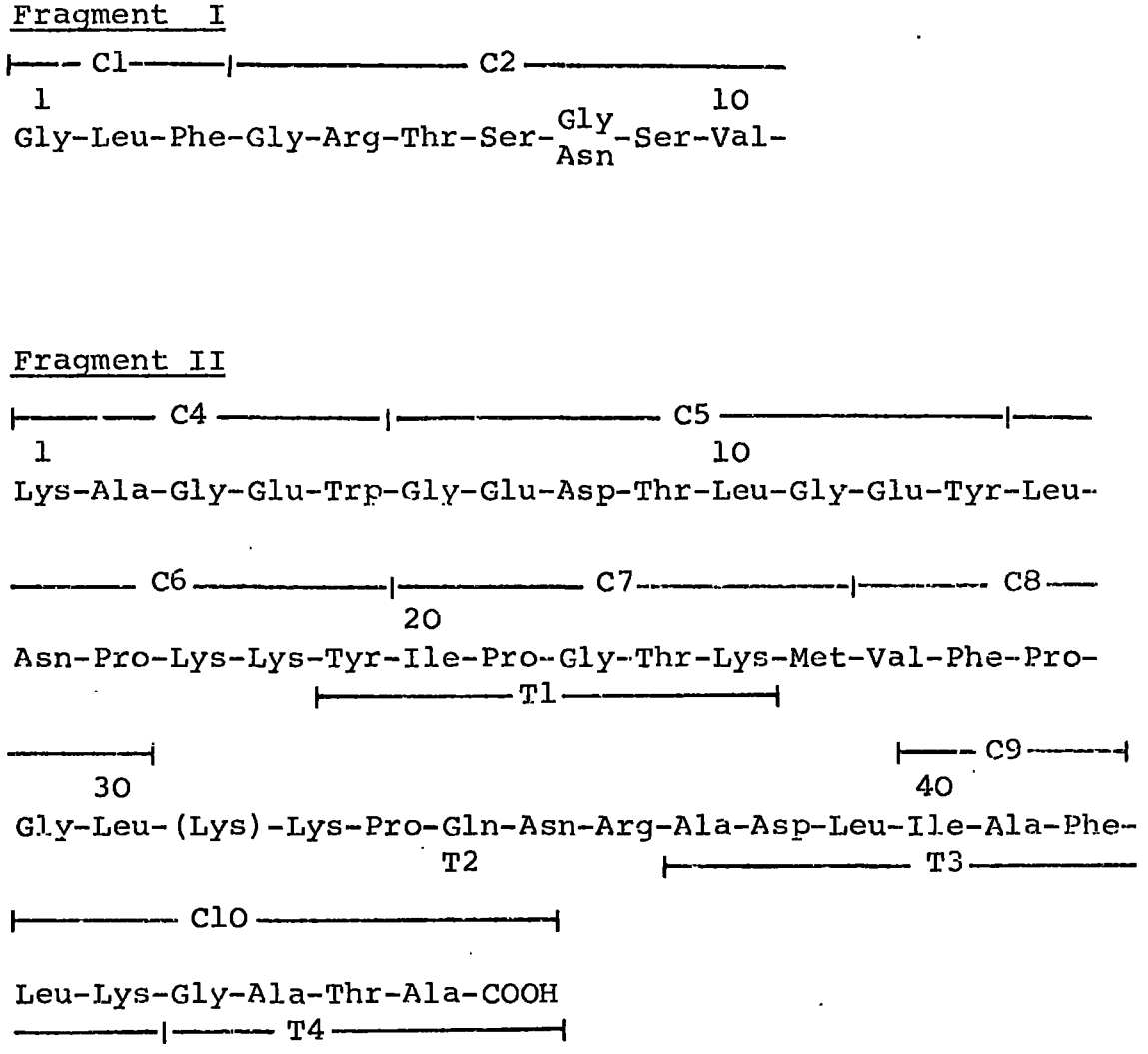
Peptide	Mobility at pH 6.5	Mobility at pH 1.9	Dansyl-Edman Results
T1	1.02		Tyr-Ile-Pro-Gly-Thr-Lys
T2	0.95		Lys-Pro-Glx-Asx-Arg
T3	0	0.45	Ala-Asx-Leu-Ile-Ala-Phe-Leu-Lys
T4	0	0.65	Gly-Ala-Thr-Ala

FIGURE 18.

The partial amino acid sequence of *Porphyra* cytochrome *c*.

FIGURE 18.

The Partial Sequence of Porphyra Cytochrome c



(Peptide C3 (Ser-Tyr) could also be assigned to the sequence by homology with other cytochromes (Dayhoff, 1972))

Haem Peptides

The chymotryptic haem zone was visible on the pH 6.5 electrophoretogram and had a mobility of 0.30. Total hydrolysis revealed cysteic acid and histidine to the present but no sequence information was obtained.

The tryptic haem was neutral at pH 6.5. Elution of the electrophoretogram with 20% (v/v) pyridine gave a red solution which on dansyl analysis gave an N-terminus, glycine. Total hydrolysis composition analysis showed the presence of cysteic acid and histidine but no further sequence information was obtained.

Chymotryptic fragments (Table 13)

Peptide C1 (Gly-Leu-Phe)

This neutral peptide was purified at pH 1.9.

Peptide C2 (Gly-Arg-Thr-Ser-Gly-Ser-Val-(-))

This basic peptide was purified in low yield at pH 1.9. Total hydrolysis analysis indicated that the C-terminus was phenylalanine. At the fourth Edman position no distinction could be made between glycine and aspartyl. The mobility suggested that the latter would be an amide. The peptide was recovered in insufficient yield to obtain the complete sequence.

Peptide C3 (Ser-Tyr)

Peptide C4 (Lys-Ala-Gly-Glx-Trp)

This Ehrlich positive peptide was purified at pH 1.9. Carboxypeptidase-A released tryptophan on short digestion and tryptophan with glutamic acid on more prolonged digestion. The mobility also placed the glutamyl residue as acidic.

Peptide C5 (Gly-Glx-Asx-Thr-Leu-Gly-Glx-Tyr)

The mobility of this acidic peptide indicated that both the glutamyl residues and the aspartyl residue were acidic.

Peptide C6 (Leu-Leu-Asx-Pro-Lys-Lys-Tyr)

The mobility of this peptide indicated that the aspartyl residue was an amide.

Peptide C7 (Ile-Pro-Gly-Thr-Lys-Met)

This platinic iodide positive peptide was recovered in high yield and purified at pH 1.9.

Peptide C8 (Val-Phe-Pro-Gly-Leu)

This peptide gave an unambiguous sequence on dansyl-Edman analysis but no explanation could be given for its observed mobility which indicated the presence of one basic amino acid.

Peptide C9 (Ile-Ala-Phe)

Peptide C10(a) and 10(b)

These peptides were recovered and purified in good yield. Dansylation without subsequent hydrolysis at the -5 Edman

position for peptide 10(a) and the -4 Edman position for peptide 10(b) showed the C-terminus to be alanine, thus placing these peptides at the C-terminus of the protein.

Tryptic Fragments (Table 14)

Peptide T1 (Tyr-Ile-Pro-Gly-Thr-Lys)

This peptide was purified at pH 6.5.

Peptide T2 (Lys-Pro-Glx-Asx-Arg)

The mobility of this peptide indicated that either of the glutamyl or aspartyl residues was an amide.

Peptide T3 (Ala-Asx-Leu-Ile-Ala-Phe-Leu-Lys)

The mobility of this peptide indicated that the aspartyl residue was acidic.

Peptide T4 (Gly-Ala-Thr-Ala)

This neutral peptide was purified at pH 1.9. No ϵ-dansyl-lysine was observed during sequence analysis and dansylation without subsequent hydrolysis after three Edman degradation steps placed this as the C-terminal peptide of the protein with the C-terminal amino acid alanine.

11. Preliminary Sequence investigations of Saprolegnia cytochrome c

The pooled fractions of the P-30 column with the highest purity ratios were lyophilized, taken up in a minimum quantity of water and denatured with ethanol. The

protein was not as readily denatured as the cytochrome c studied from algae, and was left overnight at 2-4°C before centrifuging and washing.

The pellet dissolved readily in water and was transferred to the digest vessel. On the addition of 5% (w/w) chymotrypsin the autotitrator showed only a slight rise in the base line, suggesting little or no digestion to be occurring. After 100 min a further 2.5% (w/w) chymotrypsin was added but no change in the base line was detected. The digest vessel was then removed from the titrator and left overnight at room temperature. Overnight the pH of the digest fell to 7.0 and after readjustment to pH 8 a further 2.5% (w/w) chymotrypsin produced no response on the titrator. An aliquot taken for dansyl analysis revealed the presence of valine, isoleucine, leucine, phenylalanine, tyrosine, lysine, alanine and glycine in high quantity with glutamic acid, aspartic acid, threonine and serine in lesser amounts, as N-terminal amino acids. The presence of trimethyllysine in this respect was also indicated.

The digest was subjected to pH 6.5 electrophoresis. The mobilities and N-terminal amino acids of the cadmium-ninhydrin positive zones are shown in Table 15. The zone at mobility 2.4, was strongly Ehrlich positive. Digestion with carboxypeptidase-A gave only tentative evidence of a

TABLE 15.

Mobilities and N-terminal Amino Acids of *Saprolegnia* Chymotryptic Peptides

Peptide	Mobility at pH 6.5	Mobility at pH 1.9	N-terminal amino acids and dansyl-Edman analysis
C1	-1.60		Glu
C2	-1.00		Asx-Glx-Asx-Pro-
C3	0	0	Lys-Ala-Asx-
C4	0	0.65	Leu-Tyr
C5	0	0.79	Val-Asx-Ala-
C6	1.10	1.13	Lys-Pro-Glx

C-terminal tryptophan. Total hydrolysis analysis showed only lysine and arginine to be present.

The haem zone was neutral and was eluted in the usual way. Dansyl analysis indicated the N-terminus of the chymotryptic haem peptide to be alanine. Semi-quantitative composition analysis showed the presence of proline, lysine, histidine, alanine, glycine, threonine, serine and cysteic acid, with glutamic acid and aspartic acid in lower amounts. N-terminal analysis of the total protein indicated that there was no blocking group and that the N-terminal amino acid was alanine.

Many free amino acids were characterised on electrophoretograms, possibly resulting from the prolonged digestion.

12. Cyanogen bromide cleavage of Mushroom cytochrome c

The material of $E_{280}(ex)/E_{410}(ex)$ ratios of greater than 3.5 was subjected to cleavage by cyanogen bromide.

The reaction mixture was chromatographed on a column of G-50 Sephadex. The eluate was monitored by an Isco model 222 u.v. spectrophotometer, and two peptide fragments, one red and one colourless, were identified.

The non-haem fragment was shown to have the N-terminus, valine, thus indicating a Met-Val sequence characteristic of other cytochromes c (Dayhoff, 1969).

FIGURE 19.

The overall phylogenetic tree relating thirty-eight species. The branch lengths are expressed in 'Accepted Point Mutations' (PAMs) (see Dayhoff, 1972). Figures in brackets refer to PAM values calculated where blanks occur in ancestral node sequences (see text).

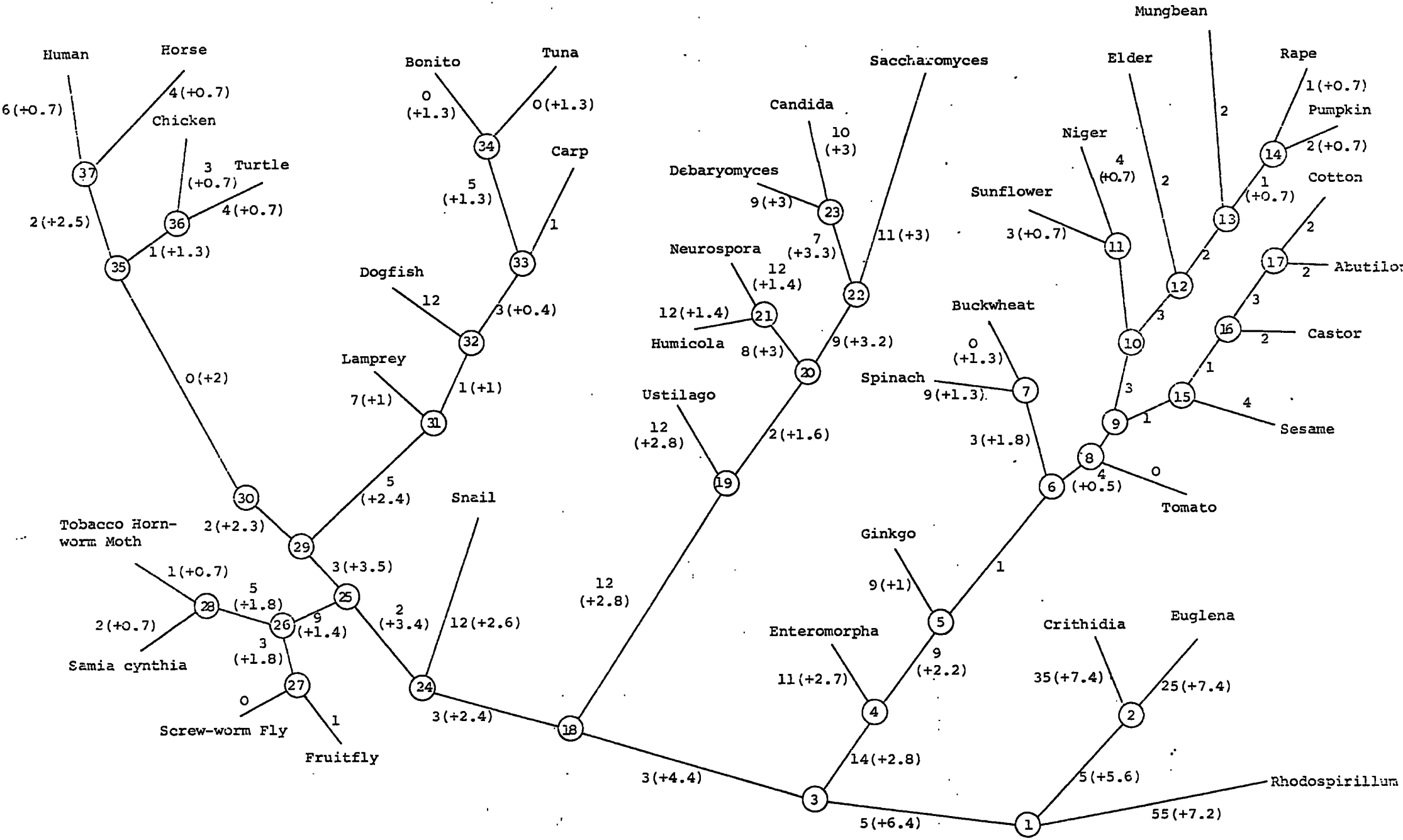

Human
Horse
Chicken
Turtle
Bonito
Tuna
Carp
Dogfish
Lamprey
Tobacco Horn-worm Moth
Samia cynthia
Screw-worm Fly
Fruitfly
Snail
Saccharomyces
Candida
Debaryomyces
Neurospora
Humicola
Ustilago
Enteromorpha
Ginkgo
Spinach
Buckwheat
Tomato
Sesame
Castor
Abutilon
Cotton
Pumpkin
Rape
Mungbean
Elder
Niger
Sunflower
Crithidia
Euglena
Rhodospirillum
6(+0.7)
4(+0.7)
3 (+0.7)
4(+0.7)
2(+2.5)
1(+1.3)
0 (+1.3)
0(+1.3)
5 (+1.3)
1
12
3(+0.4)
7(+1)
1(+1)
0(+2)
5 (+2.4)
2(+2.3)
3(+3.5)
1(+0.7)
5 (+1.8)
2(+0.7)
9 (+1.4)
3 (+1.8)
0
1
2 (+3.4)
12(+2.6)
3(+2.4)
12 (+2.8)
10 (+3)
9(+3)
7 (+3.3)
11(+3)
12 (+1.4)
12(+1.4)
8(+3)
9(+3.2)
12 (+2.8)
2(+1.6)
11(+2.7)
9 (+2.2)
14(+2.8)
3(+4.4)
5(+6.4)
0 (+1.3)
9(+1.3)
3(+1.8)
4 (+0.5)
0
1
4
3
2
1(+0.7)
2(+0.7)
1 (+0.7)
2
2
3
2
2
3
1
4
1
3
2
2
35(+7.4)
25(+7.4)
5(+5.6)
55(+7.2)
37
36
35
34
33
32
31
30
29
28
26
27
25
24
23
22
21
20
19
18
3
4
5
7
6
8
9
10
11
12
13
14
15
16
17
2
1

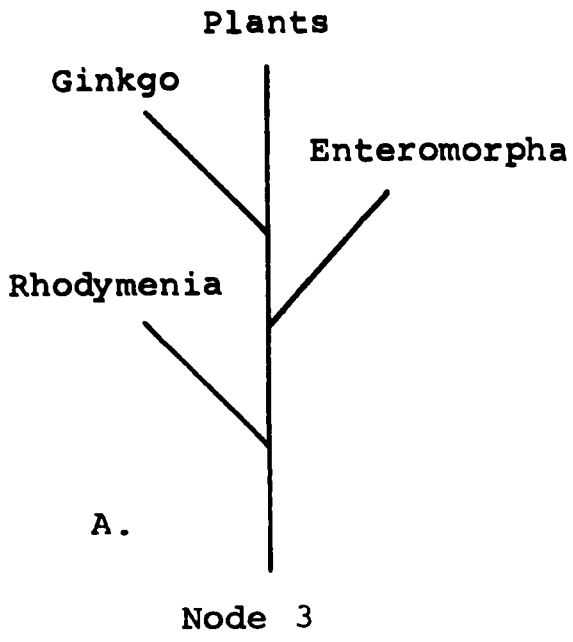
FIGURE 20 (A-D).

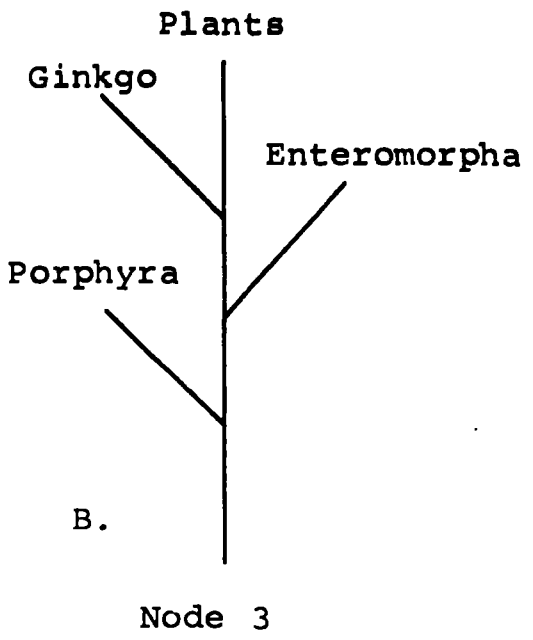
The trees obtained using the available sequence data for the red algae.

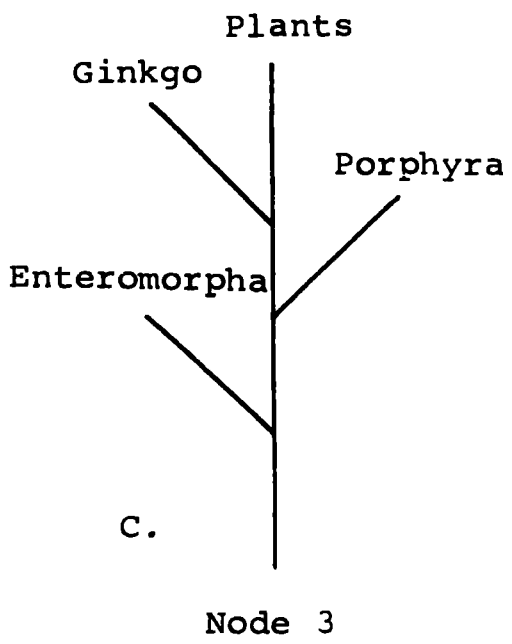
Fig. 20A. The tree obtained when Rhodymenia was added as a single sequence.

Fig. 20B and C. The trees obtained when Porphyra was added as a single sequence.

Fig. 20D. The tree obtained when Porphyra was added to the tree as shown in Fig. 20A.


Plants
Ginkgo
Enteromorpha
Rhodymenia
A.
Node 3


Plants
Ginkgo
Enteromorpha
Porphyra
B.
Node 3


Plants
Ginkgo
Porphyra
Enteromorpha
C.
Node 3

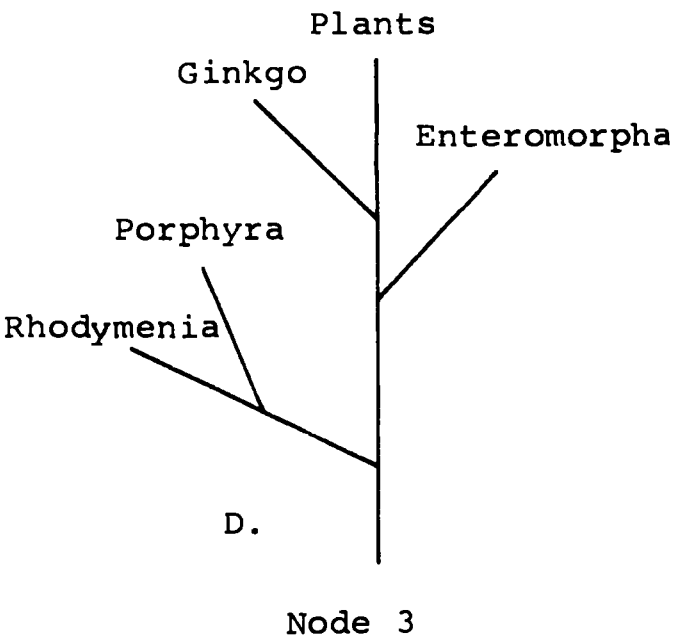
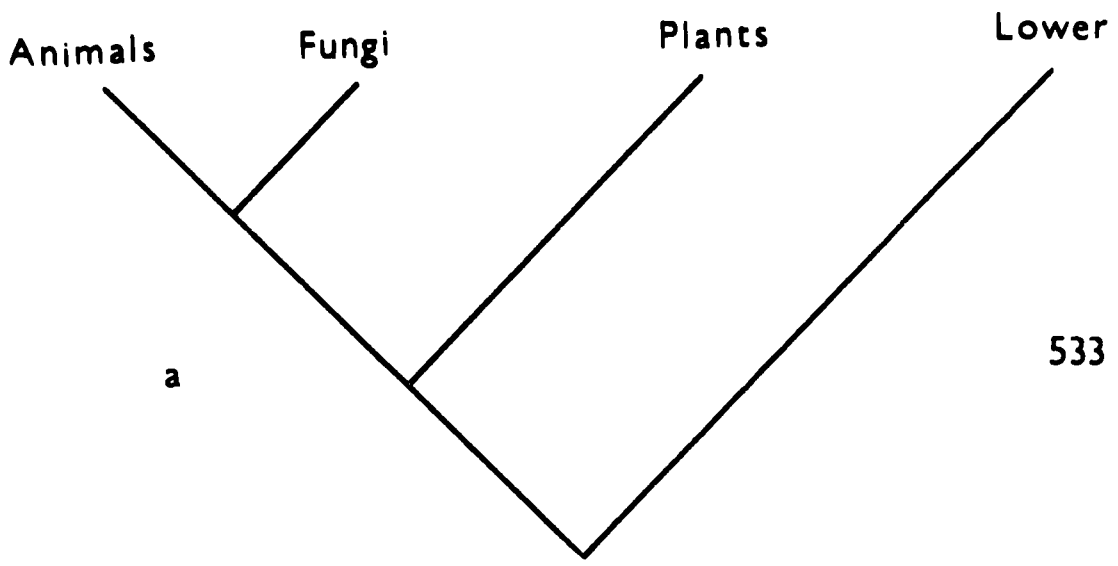

Plants
Ginkgo
Enteromorpha
Porphyra
Rhodymenia
D.
Node 3

FIGURE 21 (a, b, c).

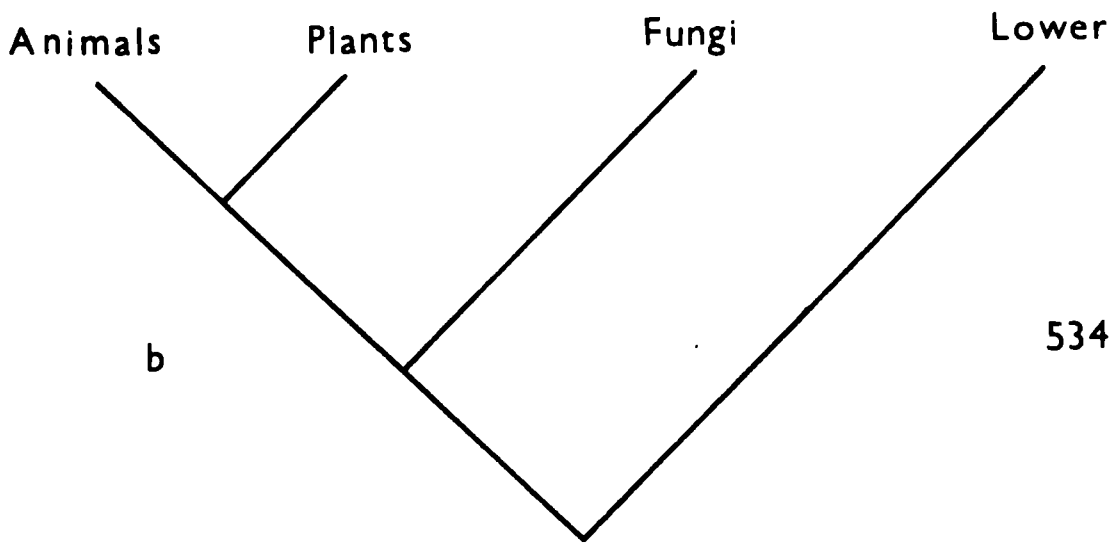
The alternative ordinations of the tree kingdoms when the sequence data is considered as four groups. The Crithidia, Euglena and Rhodospirillum sequences comprise the 'lower' group.

The mutation values given are the scores obtained when the three alternative trees were evaluated using the ancestral sequence method.



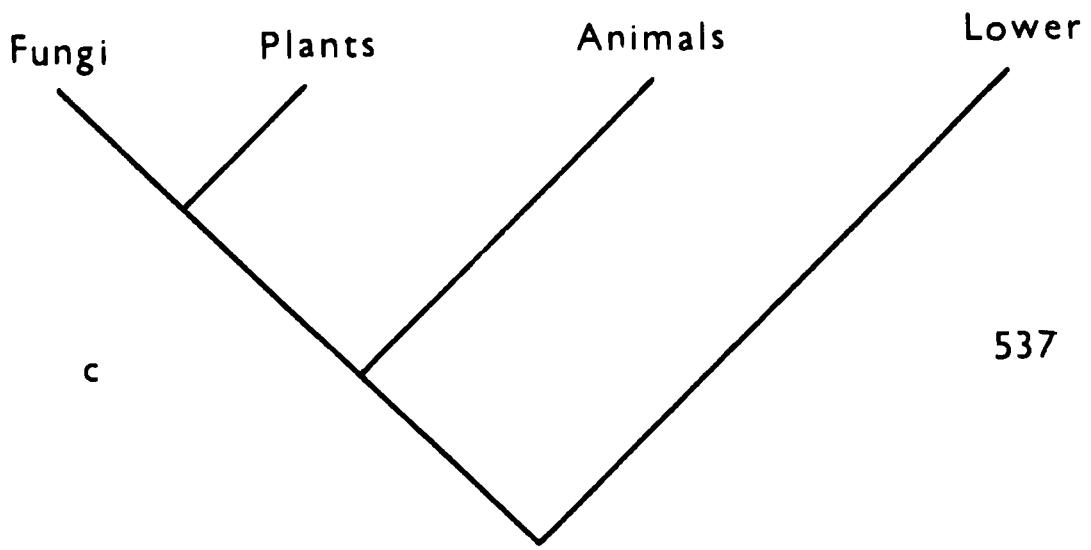
a

533



b

534



c

537

FIGURE 22(a-o).

The alternative ordination of the tree kingdoms when the sequence data is divided into groups according to the Whittaker system (see Appendix 4). The mutation values given are the scores obtained when the fifteen trees were evaluated by the ancestral sequence method.

A = Animals

P = Plants

F = Fungi

U = Protista (Unicells)

M = Monera.

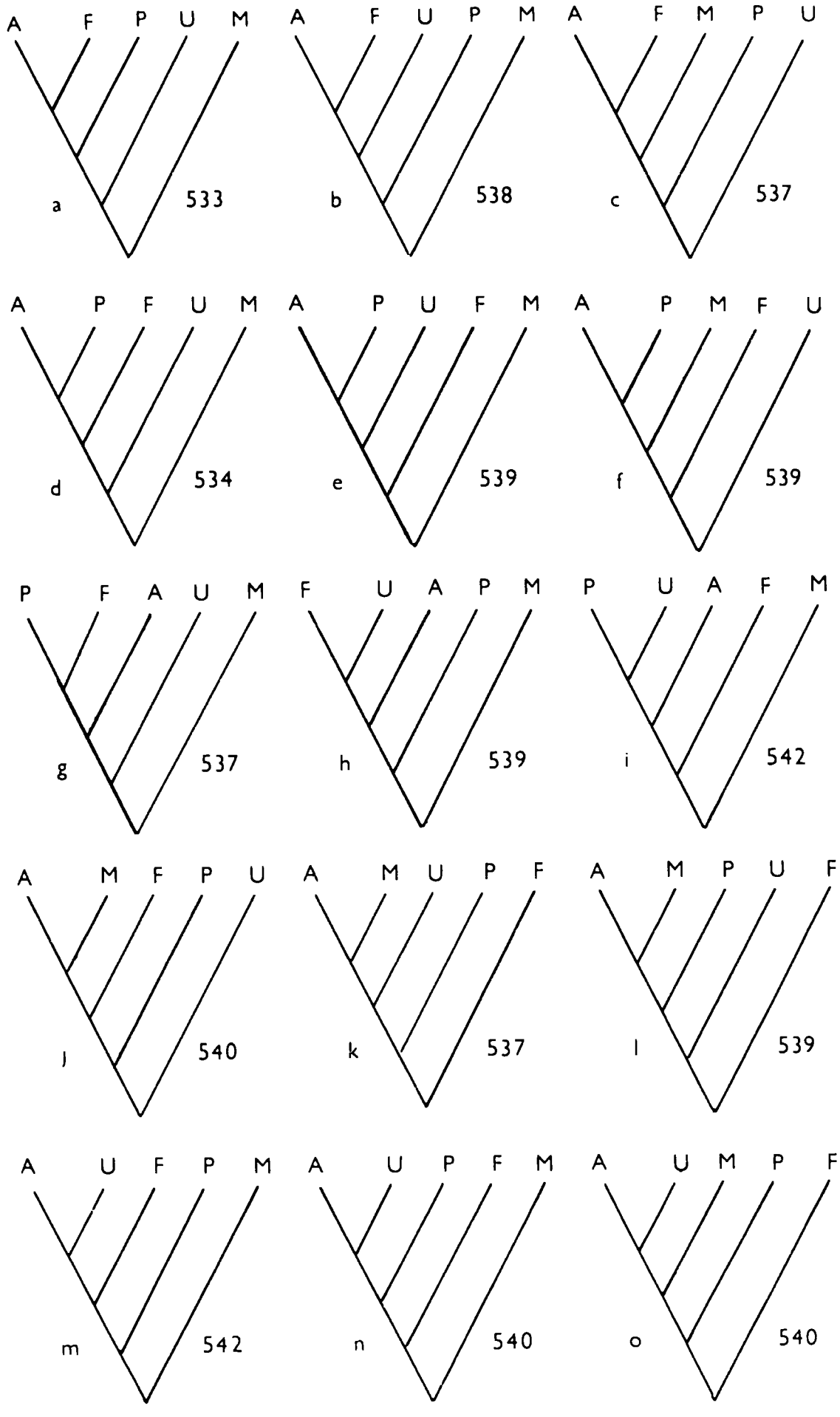

A F P U M
a 533
A F U P M
b 538
A F M P U
c 537
A P F U M
d 534
A P U F M
e 539
A P M F U
f 539
P F A U M
g 537
F U A P M
h 539
P U A F M
i 542
A M F P U
j 540
A M U P F
k 537
A M P U F
l 539
A U F P M
m 542
A U P F M
n 540
A U M P F
o 540

TABLE 16.

Times of divergence of some animal and plant groups

	Groups compared	Amino acid differences between sequences: Average number of variant residues	Amino acid differences between sequences: Times of* divergence ($\times 10^6$ years)	Amino acid differences between sequences: Times of* divergence (corrected) ($\times 10^6$ years)	Minimum mutation distance between sequences: Average of the minimum mutation distance	Minimum mutation distance between sequences: Times of* divergence ($\times 10^6$ years)	Minimum mutation distance between sequences: Times of* divergence (corrected) ($\times 10^6$ years)
(A)	Animal groups						
	Mammals - Birds	9.9 ± 1.7 (36)	280+	280+	13.7 ± 2.4 (36)	280	280
	Vertebrates - invertebrates	26.6 ± 3.1 (84)	750	830	33.6 ± 4.9 (84)	690	740
(B)	Plant groups						
	Angiosperms - Ginkgo	17.8 ± 2.3 (18)	500	520	18.9 ± 2.6 (18)	390	400
	Spermatophyta - Enteromorpha	25.3 ± 2.0 (19)	720	790	31.6 ± 2.4 (19)	650	700
	Spermatophyta - Rhodymenia	46.2 ± 2.1 (19)	1310	1760	-	-	-
(C)	Kingdoms						
	Animals - plants	47.2 ± 2.7(400)	1330	1860	56.4 ± 4.8(400)	1160	1370
	Animals - fungi	47.9 ± 2.6(100)	1350	1910	64.6 ± 3.9(100)	1320	1620
	Plants - fungi	52.3 ± 2.4 (64)	1480	2120	68.3 ± 3.2 (64)	1400	1770
	Protista - plants	56.9 ± 3.4 (30)	1600	2900	75.8 ± 2.5 (30)	1550	1980

The species considered are those given in Dayhoff (1972; 1973) with the addition of Euglena (Pettigrew, 1973). In cases in which identical cytochromes c occur in more than one species a single value was used in calculating average deviations.

For each group of comparisons the average differences or distance is given followed by its standard deviation from the mean and the number of comparisons used.

*Times of divergence were calculated by a linear extrapolation using the respective 'unit evolution periods' from the point of the mammal to bird divergence. The corrected times of divergence were derived after application of the formula of Feller (1950) to the data (see Margoliash and Fitch, 1968).

+Dating obtained direct from the fossil record.

FIGURE 23.

Relations between amino acid differences (A) and minimum mutation distances (B) of cytochrome c from different classes and phyla of organisms to elapsed time since divergence of the lines of descent.

The straight lines were calculated on the basis of a value of 280 million years since the time of divergence of the mammalian and avian lines of descent. The curved lines show the relation corrected for multiple changes according to the formula $r = n \ln(n/n-\lambda)$ (Feller, 1950). The Rhodymenia data was calculated on the assumption that the average number of changes in the undetermined part of the sequence is the same as that in the known fragments.

The major group divergences indicated are:-

A Enteromorpha - Spermatophyta

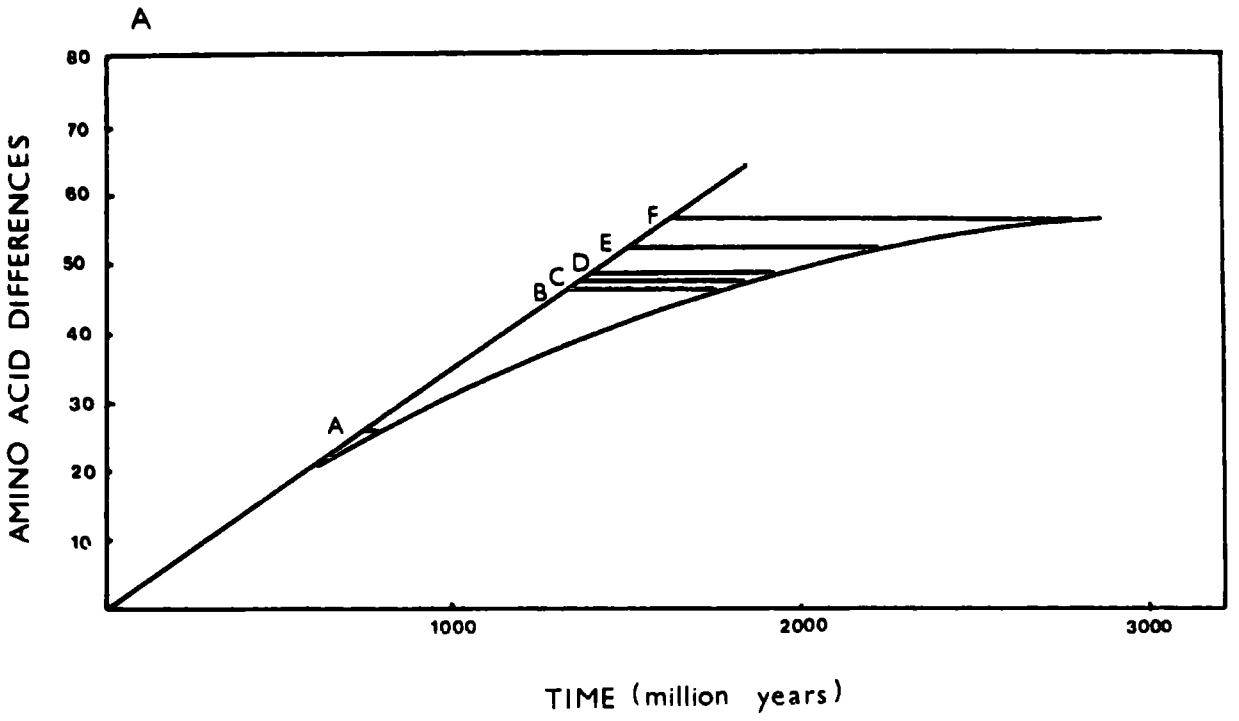
B Rhodymenia - Spermatophyta

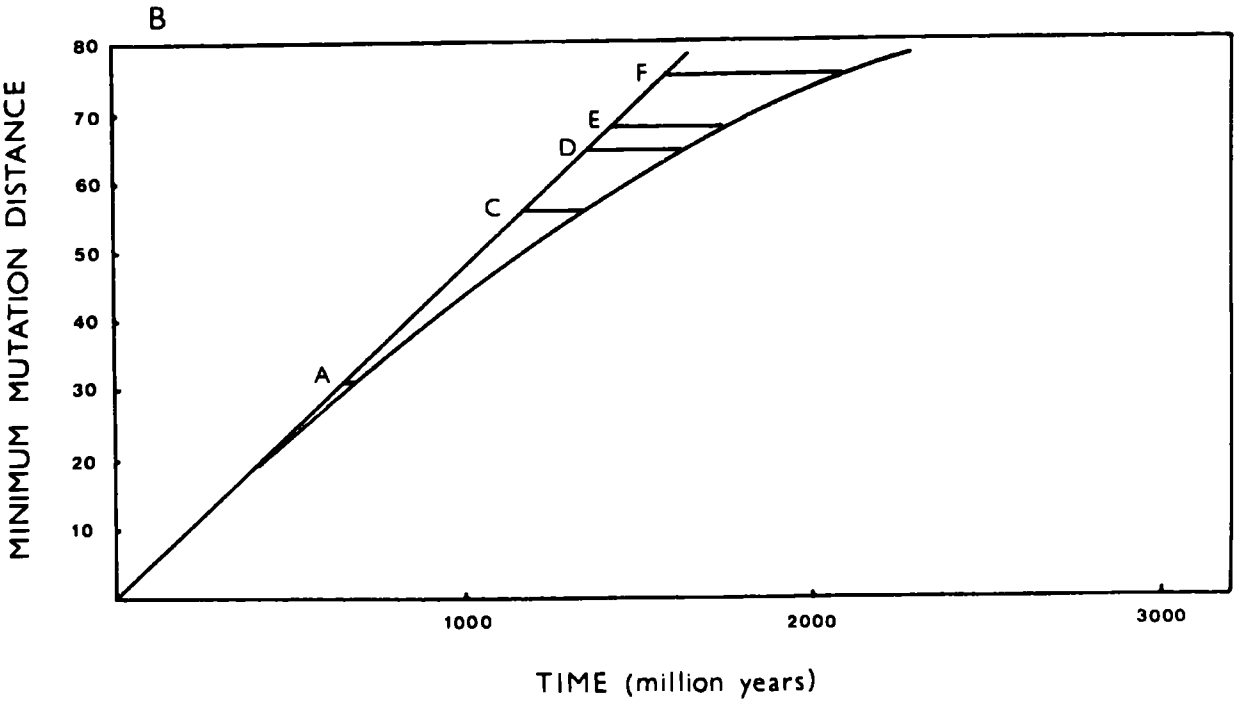
C Animals - Plants

D Animals - Fungi

E Plants - Fungi

F Protista - Spermatophyta


A
AMINO ACID DIFFERENCES
80
70
60
50
40
30
20
10
A
B
C
D
E
F
1000
2000
3000
TIME (million years)


B
MINIMUM MUTATION DISTANCE
80
70
60
50
40
30
20
10
A
C
D
E
F
1000
2000
3000
TIME (million years)

DISCUSSION

The Distribution and Characterization of c-type Algal Cytochromes

Keilin (1925; 1927) showed that cytochrome c occurred not only in members of the animal kingdom, as had been demonstrated previously by McMunn (1884; 1886; 1887), but that it was also present in plants. Since Keilin's early work, the preparation and characterization of cytochrome c from a variety of higher plant sources has been accomplished (see Richardson et al., 1970).

In 1935 Yakushiji described a c-type cytochrome solubilised from a red alga, Porphyra tenera (Yakushiji, 1935) and later Katoh (1959) investigated the physical and chemical properties of several algal cytochromes. Crystallization of algal cytochrome was first reported from Porphyra by Katoh (1960) and Yakushiji et al. (1960). Katoh (1959) found that this cytochrome was localised in the chloroplast, had a high redox potential similar to that of cytochrome f, and, therefore, concluded that it belonged to the photosynthetic system. Nishimura (1959) reported a similar cytochrome from Euglena gracilis, which was only detectable in autotrophically-grown cells.

Sugimura et al. (1968) characterized c-type cytochromes from a variety of red, brown and green algae (see also Kamen et al., 1971). However, these cytochromes have their α band

absorption maxima at 553 nm, are acidic proteins with no appreciable oxidation by mammalian cytochrome oxidase (Yamanaka and Okunuki, 1963), and hence do not participate in cell respiration. Holton and Myers (1967a and b) demonstrated the presence of c-type cytochromes in the blue-green alga, Anacystis nidulans, and characterized two acidic c-types and partially characterized a basic form which was obtained in much lower yields. This latter cytochrome is presumably homologous with the basic eukaryotic cytochromes c. Many other c-type cytochromes of uncertain affinities are also reported from bacteria, fungi and yeasts (see Kamen et al., 1971).

Table 17 gives a summary of the occurrence and properties of some of the presently recognised groups of c-type cytochromes, i.e. those characterized by the possession of the covalently bound haem c prosthetic group (see Keilin, 1966); however, variations in structure are found within the various groups shown in Table 17. Many of the photosynthetic bacterial c_2-type cytochromes have been shown to have more than one haem group molecule (Dus and Kamen, 1963; Kamen et al., 1971). Also, subtle differences in the haem environment as well as a slightly higher content of alpha helix in cytochrome c_2 lead to the observed variations in the α band absorption peaks. In the case of the protozoan Crithidia, which has the α band absorption at 557 nm, sequence determination

TABLE 17

Occurrence and Properties of c-type Cytochromes

Cytochrome	Occurrence	Main absorption bands in nm	Remarks	E'_o pH 7 mV.	Function
c	Widespread, typical mitochondria of eukaryotic organisms	550 521 415	Soluble, rather stable to heat and acids, MW 12-13,000; isoelectric point above 10. Many crystallized. Complete amino acid sequence known for many[1].	+255	Electron carrier in respiratory chain between cytochromes c_1 and aa_3. Not autoxidizable, does not react with CO.
		553 555 558	In Protozoa[2,3,4] sequence homologous with above but haem attachment different[5,6].	+280 +307	
c_1	In mitochondria of animals, possibly also present in plants and some bacteria	553 524 418	Membrane-bound, but solubilizable, thermolabile MW 37,000 per one Fe	+226	Electron carrier in respiratory chain between cytochromes *b* and *c*, not autoxidizable, does not react with CO.
c_2	In photosynthetic non-sulphur purple bacteria	551 521 417	Soluble, stable *Rhodospirillum* sequence known[1]	+310 to +340	Electron carrier in photosynthesis. Not autoxidizable, does not react with CO. Reacts only slowly with cytochrome aa_3.
c_3	In *Desulphovibrio desulphuricans*[7,8,9]	553 525 419	Soluble, thermostable MW 13,000 with 2 haems	-205	Electron carrier in sulphate reduction, autoxidizable.
c_4	In *Azotobacter vinelandii*[10]	551 522 416	Soluble, stable. MW 12,000	+300	Electron carrier in the respiratory chain, but does not react with cytochrome aa_3. Not autoxidizable.
c_5	In *Azotobacter vinelandii*[9]	555 526 420	As for c_4	+320	As for c_4.
c_6 = *f*	Chloroplasts of higher plants[10]	554 525 423	Bound in chloroplast thylakoids, solubilizable, MW 110,000 with 2 haems	+350 to 365	Electron carrier in plant photosynthesis, does not react with cytochrome aa_3.

P.T.O.

TABLE 17 continued.

Cytochrome	Occurrence	Main absorption bands in nm	Remarks	E'_o pH 7 mV.	Function
f-like	Algal chromatophores in green, brown, red and blue-green algae[9,10]	552 520 415 to 556 521 419 frequently 553	More readily extractable than cytochrome *f*. MW 16,500 to 23,000	+340 to +390	Electron carriers in algal photosynthesis, do not react with cytochrome aa_3, but with cytochrome *cd* of *Pseudomonas*.
c_x	Some blue-green algae, e.g. *Anacystis*[11,12]	549	Autoxidizable, reacts with CO. MW 34,000?	-260	? oxidase
c_x	Bacterial cytochromes of *E. coli*, *Pseudomonas*; photosynthetic non-sulphur bacteria, *Mycobacteriaceae*[9]	Variable 551-554	Variable properties. Some (*Chromatium*, *Chlorobium*) contain flavin. *Pseudomonas* sequences[6] known.	from -200 to +350	Variable, respiratory photosynthetic, nitrate and sulphate reductases.
c' and *cc'*	Bacteria ("RHP") in *Chromatium* etc.[9,13]	Ferrous spectrum resembling that of ferromyoglobin	Autoxidizable, reacts with CO; soluble, MW 25,000?		? oxidase

(1) Dayhoff (1972).
(2) Meyer & Cusanovich (1972).
(3) Yamanaka *et al.* (1968).
(4) Hill *et al.* (1971).
(5) Pettigrew (1972).
(6) Pettigrew (1973).
(7) Postgate (1954).
(8) Postgate (1956).
(9) Kamen *et al.* (1971).
(10) Lemberg & Barrett (1973).
(11) Holton & Myers (1967a).
(12) Holton & Myers (1967b).
(13) Bartsch & Kamen (1968).

has shown that only one cysteine residue is involved in the binding of the haem (Pettigrew, 1972).

However, comparatively few of the bacterial c-type cytochromes have been fully characterized and few sequences are at present known (see Table 17). The cytochromes investigated in this study are of the mitochondrial type and were shown to be homologous with other eukaryotic mitochondrial cytochromes c from higher plants (Boulter, 1970) and animals (Margoliash and Smith, 1965; Margoliash and Schejter, 1966; see also Dayhoff, 1972).

Technical Problems in the Purification and Sequencing of Algal Cytochrome c

Initial experiments with Enteromorpha thalli indicated that the major problem would be the low yield of cytochrome c, as compared with animal sources and those obtained from germinating seeds (Richardson et al., 1970). The yields obtained from Enteromorpha were of the same order as the above authors reported for cytochrome extracted from green tissues of higher plants (Richardson et al., 1970). A comparison of the yields obtained from other sources is given in Table 18.

Whereas the material of Rhodymenia and Porphyra represented the single species, Enteromorpha source probably represented a mixture of several closely related species including E. linza, E. flexuosa and E. intestinalis (P. R. Edwards, personal

TABLE 18.

Comparison of Yields and Spectral Properties of Various Plant and Algal Cytochromes c

Species	550 nm(R) / 280 nm(O)	416 nm(R) / 410 nm (O)	410 nm(O) / 280 nm(O)	416 nm(R) / 550 nm(R)	Yield mg cytochrome/ Kg starting material
Wheat germ (1)	1.16	1.20	3.8	4.4	-
Maize (seeds) (2)	1.10	1.25	4.9	5.7	0.55
Ginkgo biloba (seeds) (2)	1.06	1.20	4.4	5.2	0.12
Arum maculatum (shoots) (2)	0.95	1.20	4.0	5.3	0.09
Cauliflower (inflorescences) (2)	1.22	1.22	4.9	5.0	0.20
Spinach (leaves) (3)			4.0	4.7	0.07
" (seeds) (4)	0.9				0.05
Enteromorpha intestinalis	1.00	1.20	4.1	5.2	0.13
Porphyra umbilicalis	0.9	1.22	3.9	5.3	0.1
Rhodymenia palmata	1.0	1.12	4.8	5.3	0.24-0.80
Horse heart (5,6)	1.20	1.20	4.6	4.4	250

(1) Stevens et al. (1967).
(2) Richardson et al. (1971).
(3) Asada and Takahashi (1971).
(4) Brown et al. (1973).
(5) Margoliash and Frohwirt (1959).
(6) Margoliash and Walasek (1967).

communication). Species of this genus are notoriously difficult to identify except by details of their life cycles (Bliding, 1963; Cook and Hoffman, 1971); however, a unique sequence was obtained indicating that predominantly only one species was present, or, alternatively, that the several species of the genus have identical cytochromes c as has been reported for Brassica spp. (Richardson et al., 1971 ; Thompson et al., 1971c).

Because of the large mass of starting material required to compensate for the low yields, the initial problems encountered during the extraction were largely technological. Speed of handling in the initial extraction stages of cytochrome c is reported to mitigate against losses in yield (Margoliash and Walasek, 1967), which often result from the action of degradative enzymes. The use of the Broadbent centrifuge proved particularly useful in this respect for the bulk filtration of crude homogenates, rather than Buchner filters. The Terylene bag material chosen represented the most efficient filtering agent, giving the best rate of separation.

Various methods have been employed by Japanese workers to prepare photosynthetic cytochromes c from algae (see Sugimura et al., 1968) and of these homogenization procedures have led to problems with slimy substances. No problems

were encountered in this respect with either Enteromorpha or Porphyra; homogenization of Rhodymenia thalli gave an oily slurry which filtered satisfactorily in the Broadbent centrifuge. Homogenization of Fucus and Laminaria gave very slimy extracts, which filtered very slowly, and necessitated frequent removal of the filter cake.

It has been shown that many of the polysaccharides extracted from algae have sulphonated side-chains (Drum, 1969; Percival, 1968) and the cationic binding properties of polysaccharides from Rhodophycean and Phaeophycean sources has also been reported (Springer and Abdelmessih, 1969). It is suggested that the zero yields of cytochrome obtained from the species rich in these substances are attributable to these processes. Microscopic investigation of Enteromorpha homogenates revealed a very low percentage (estimated < 5%) of cell breakage, but in spite of this, extraction methods involving mere soaking of thalli in the extraction medium proved ineffectual. It is significant to note that in the case of Nitella where ~100% cell breakage could be guaranteed, no detectable cytochrome was extracted; also, Lonsdale (personal communication) failed to extract RNA from the same material.

Previous investigations of cytochromes c have often involved rather drastic methods for the initial extraction of

the protein. These include treatments with organic solvents (Goddard, 1944; Hagihara et al., 1958, 1959; Wasserman et al., 1963; Fridman et al., 1968) or with hot acid (Goddard, 1944). Inhomogeneities of most preparations have been shown to result from such treatments (Margoliash, 1954 ; Yamanaka et al., 1959). The use of organic solvents can cause denaturation (Kaminsky and Davison, 1969), while extremes of pH or high temperature may lead to denaturation, deamidation or polymer formation (see Margoliash and Schejter, 1966).

The blending of plant material in dilute phosphate buffer was shown to give an efficient extraction of cytochrome c (Morita and Ida, 1968; Ida and Morita, 1969). A similarly mild technique was developed by Richardson et al. (1970) to extract cytochrome c from germinating seedlings. The method used to prepare algal cytochromes was a modification of this technique. Algal thalli were extracted at pH 4.5-4.6 in the presence of ascorbic acid and EDTA to reduce oxidative destruction of the cytochrome by compounds produced in the browning reaction (Wasserman et al., 1963). Initially the homogenate was filtered at this pH but no apparent loss of yield or decrease in purity was observed when the homogenate was adjusted directly to pH 8 prior to filtering. Filtering at the acid pH precipitated many of the storage proteins found in the germinating seed (Richardson

et al., 1970) and this was considered an unnecessary step with algal material, particularly as it was desirable to keep the time that the protein was under acidic conditions to a minimum to avoid modification.

The large volumes and low cytochrome content provided major problems of concentration in the initial stages. Yamanaka et al. (1964a) showed that the ability of cation exchangers to absorb cytochrome c from solution provided an efficient method of concentrating very dilute cytochrome solutions obtained from poor sources of the protein. J. Ramshaw (personal communication) reported that with plant material large quantities of material precipitated in and on the resin, thus necessitating batchwise elution of the column. Batchwise elution was also found to represent the best method for algal sources, as much debris and extraneous material (particularly for the red algae studied) accumulated on the resin. The pH was maintained at 8 during elution since below pH 6 the cytochrome becomes very strongly bound to the resin (Boardman and Partridge, 1953; 1954; 1955).

The batchwise elution of the cation exchangers still gave fairly large volumes of eluate, which necessitated a further concentration step. CM-Sephadex had proved the most useful way of affecting this (Richardson et al., 1970), since impurities which had a higher affinity for the resin than the reduced cytochrome, were left adsorbed to the upper

part of the column after the cytochrome had migrated downwards during the washing procedure, thus giving rise to a significant increase in purity. Also, the flow rates for CM-Sephadex were better compared with those for CM-cellulose. The use of CM-Sephadex still involved dialysis of relatively large volumes of eluates. Even after this concentration step, care was still needed in order to avoid modifying the protein, and for this reason ammonium sulphate was initially avoided since Flatmark (1966) had suggested this led to deamidation. Precipitation with ammonium sulphate has been used as an alternative method for the initial concentration of cytochrome c from spinach leaves (Asada and Takahashi, 1971) and elder flowers (Brown and Boulter, 1973c). Margoliash and Schejter (1966) had suggested that the inhomogeneities attributed to ammonium sulphate fractionation were due to poor pH control during the ammonium sulphate step, or to the drastic extraction techniques. It was found that carefully controlled ammonium sulphate fractionation of algal cytochrome c did not give rise to changed forms of cytochrome c, possibly due to deamidated forms of the protein. The subsequent inclusion of this step gave a useful and substantial purification of the cytochrome preparation as judged by the copious precipitates of protein material obtained during fractionation. Subsequent molecular exclusion chromatography using the method of

Flatmark (1964) provided further purification and removed the remaining ammonium sulphate from the cytochrome residues. At this stage selection of best fractions was used; in all previous steps the total preparation was maintained. Final purification was achieved by gradient elution from CM-cellulose. The salt-free preparation after gel filtration was applied directly to the resin and was oxidised by potassium ferricyanide. Elution was by a linearly increasing cation concentration (Margoliash, 1962; Margoliash and Lustgarten, 1962), rather than the alternative of constant cation concentration and increasing pH (Boardman, 1959; Margoliash, 1962; Margoliash and Lustgarten, 1962). This technique will separate deamidated forms and modified forms which are not separated by gel-filtration. The results show that the use of ammonium sulphate precipitation step did not lead to any appreciable deamidation and the degree of modification generally present was low. The ampholine isoelectric focusing method used by Flatmark and Vesterberg (1966) for beef heart cytochrome c, has been shown to be unsuitable for the preparation of plant cytochromes c (Richardson et al., 1970) and was not investigated for the algal cytochromes c studied.

Dialysis was avoided where possible, especially in the later stages of purification and desalting, as at low salt

concentrations plant cytochrome c shows a tendency to bind to the dialysis tubing (Laycock, unpublished experiments). Wasserman et al. (1963) reported that wheatgerm cytochrome c is unstable at low ionic strength and it was routinely observed that eluates of the desalting column appeared cloudy.

The crystallization of the cytochromes c was not attempted, although the crystallization of wheat cytochrome c (Hagihara et al., 1958) and rice cytochrome c (Morita and Ida, 1968; Ida and Morita, 1969) has been reported and this appears to be easier than crystallization of cytochrome c from animals (Ida and Morita, 1969). The crystallization of acidic c-type cytochromes from algae is widely reported (see Sugimura et al., 1968), and Mitsui and Tsushima (1968) report the crystallization of Euglena c_{552} cytochrome.

With such a variety of sources of cytochrome c available no single method of preparation is ideal. However, it is more convenient to establish a basic method which can be followed inprinciple and modified to suit individual requirements.

The cytochromes c from the algal cytochromes used in this study are spectrally related to other mitochondrial cytochromes c, as shown in Table 18. The spectral ratio of E_{416} nm(reduced)/E_{550} nm(reduced) obtained for higher plants and algae is noticeably higher than that for animal cytochromes.

The Enteromorpha preparation was shown to be pure by polyacrylamide gel electrophoresis (Valentine and Roberts, 1971). The Porphyra preparation separated with two cytochrome bands of approximately equal density on electrophoresis at both pH 4.3 and 8.3 and the isoelectric points for each of these were calculated (Valentine and Roberts, 1971). No evidence of inhomogeneity during the preparation of Porphyra cytochrome c was observed and the presence of the two bands is unexplained. The Porphyra preparation was shown also to contain trace amounts of impurity. The apparent inhomogeneities shown during the preparation of Enteromorpha cytochrome shown in Fig. 5 were considered to be due to streaking of the column, due to overloading and not to modified or deamidated forms of the protein.

The purification methods which have been used for the preparation of fungal cytochromes are basically the same as those used for animal cytochromes (see Margoliash and Schejter, 1966). The major problems are those associated with the initial extraction of the protein. Early work on the fundamental properties of cytochromes involved the use of delft yeast preparations (Hill and Keilin, 1930; Keilin, 1930; Dixon et al., 1931). Extraction was by boiling and sodium chloride treatment to plasmolyse the cells followed by precipitation by sulphur dioxide to extract the protein. It was believed that these drastic procedures

did not lead to modification of the protein, but this belief was founded on the lack of sensitive criteria of purity. However, because fungal cytochromes are generally more difficult to extract, recent procedures have still required drastic treatment, usually by organic solvent autolysis (Sherman et al., 1965; Shirasaka et al., 1968). Alkaline extraction methods have been successfully used with cytochrome c-rich strains of Neurospora mycelium preceded by various freezing techniques (Heller and Smith, 1966a; Scott and Mitchell, 1969). The method used for the preparation of mushroom cytochrome was basically the same as that used to prepare cytochrome from algal thalli. Mushrooms were obtained in the 'cup' stage, as it was considered that at this stage, metabolic activity would be maximal. Freezing of the mushrooms and grinding to a powder in the Hobart blender prior to extraction gave a finer slurry than other methods tried. Various extraction procedures were assessed for the extraction of Saprolegnia cytochrome, involving liquid nitrogen freezing and acetone powdering to disrupt the mycelium, followed by extraction with different buffers of varying pH, with and without the addition of ethyl acetate as used by Sherman et al. (1965) and Sherman et al. (1968). The method chosen was that which gave the best yield of cytochrome (see Table 4).

The initial cytochrome yields of both mushroom and Saprolegnia were higher than those obtained from algal sources, but were considerably lower than those obtained from other fungal sources; 100-400 mg/kg fresh cells for Neurospora (Heller and Smith, 1966a), 400 mg/kg wet cells for Candida, and 100 mg/kg wet cells for Baker's yeast (Shirasaka et al., 1968). However, in these cases the strain of the fungus, the growth media and harvesting were all selected for maximal yield of cytochrome. The high yields obtained allowed direct elution from the amberlite which was used to initially concentrate the extract.

The mushroom preparation gave multiple bands on gel filtration. The major band which contained the majority of cytochrome extracted was certainly not a normal native mitochondrial c-type cytochrome as judged by its lack of reduction by ascorbic acid. Further purification of this band using gradient elution gave evidence of more inhomogeneity. Further purification of the other bands was not attempted, owing to lack of material; however, the last band to be eluted was considered from its spectral characteristics to represent native cytochrome c. It is easy to ascribe these results to denaturation; however, the extraction method used was mild and caused no apparent inhomogeneities in the cytochromes from algae. Also, the Saprolegnia

cytochrome which had been subjected to an ethyl acetate treatment, appeared as an homogeneous preparation, as judged by its gel filtration properties. An alternative explanation of the mushroom results is that the peaks may represent isoenzymes. Saccharomyces has been shown to possess two different cytochromes which exist in the same cell (Slonimski et al., 1965), although only one cytochrome is reported for Debaryomyces (Narita and Sugeno, 1968). The Basidiomycetes appear to be a somewhat neglected group in terms of cytochrome studies and a typical situation has yet to be characterized. Much selection work on strains of yeasts and antibiotic and food yielding fungi has been done during the past ten years, and it is difficult to assess the effect that this may have had on the genetic material. Amino acid compositions or peptide maps would indicate fairly simply if any differences between the various peaks existed but would not prove them to be the same protein, e.g. the cytochromes c from Kloechera and Candida krusei have quite different properties (Yamanaka et al., 1964b), presumably as a result of sequence differences, yet give the same amino acid compositions (Narita et al., 1964).

Apparently equally pure samples of the same cytochrome have been shown to vary in their spectral ratio (Paleus and Neilands, 1950; Paleus, 1960), and this demonstrates

the inadequacies of using this method alone to assess purity. The E_{410}/E_{280} ratio for algal cytochromes as shown in Table 18 was generally similar to that quoted for higher plant cytochromes c (Richardson et al., 1970). The same authors also quote varying $E_{550}(red)/E_{280}(ox)$ ratios for samples of plant cytochrome shown to be pure by gel disc electrophoresis. Differences in amino acid composition cause variations in the absorbance at 280 nm, and this is particularly apparent with changes in tryptophan content and also to a lesser extent, phenylalanine and tyrosine. The absorbance of the soret band is particularly susceptible to modifications in tertiary structure becoming larger as the protein unfolds (Stellwagen, 1968). Thus, no absolute purity criteria based on spectral ratios can be established for all species. The use of iron content as an estimate of purity has also given apparent inconsistencies (Paleus and Neilands, 1950; Bodo, 1955; Paleus, 1960). This method will indicate protein purity but cannot distinguish between native and modified protein. Analytical column chromatography on weak cation exchangers in conjunction with the inability of cytochrome c fractions to bind carbon monoxide, is considered a good criteria for purity (Margoliash, 1962; Margoliash and Schejter, 1966). Similarly, analytical electrophoresis on media such as starch gel or polyacrylamide gel provides a good criterion of purity. The

greater range of cytochrome sources now becoming available is making apparent a greater range of structures and properties within the mitochondrial cytochrome c group, and this means that a single set of purity criteria cannot be devised to cover all of the group. Ideally, any method that does not involve direct iron content should be used (see Keilin and Hartree, 1939; Van Gelder and Slater, 1962).

Previous sequence determinations of cytochrome c have often involved the use of large quantities of protein, e.g. Samia Moth, 1.30 g (Chan and Margoliash, 1966); Rabbit, 1.71 g (Needleman and Margoliash, 1966); Dog, 0.82 g (McDowall and Smith, 1965); Wheat, 0.26 g (Stevens et al., 1967). The wheat sequence was determined using only chymotryptic peptides; thus, if the same quantities of algal protein were to be prepared very considerable quantities of starting material would be required. The quantities of cytochrome c available from the preparations undertaken during this study, <30 mg, required a micro-method approach to sequencing different to the direct or subtractive Edman methods previously used in cytochrome c sequence investigations. A suitable method has been used successfully by Boulter and his co-workers for sequence studies on higher plant cytochrome c where only limited amounts of protein are available, e.g. Rape (12 mg) (Richardson et al., 1971),

Abutilon and cotton (12 mg each) (Thompson et al., 1971a), and Sesame and Castor (18 mg each) (Thompson et al., 1970). The method used was that devised by Gray and Hartley (1963a). This involves the separation of peptides, derived from proteolytic enzyme digestion of the protein, by high-voltage paper electrophoresis and sequence analysis by the dansyl-Edman method. In conjunction with using the homology between sequences as little as 0.5 μmol (6 mg) cytochrome was used in the determination of the sequence of Ginkgo biloba cytochrome (Ramshaw et al., 1971) and 0.2 μmol (2.5 mg) for the Nigella sequence (Brown and Boulter, 1973a).

The proteolytic enzymes used in this investigation were chymotrypsin and trypsin. These are the most specific of the enzymes available and have been used successfully in the analysis of other cytochromes (see for example, Chan and Margoliash, 1966; Stevens et al., 1967; Boulter et al., 1970). The two sets of peptides obtained from digestion by each of these usually give sufficient overlapping peptides to logically establish the sequence by inspection.

The algal cytochromes c were readily digested with either of these enzymes and with the specificities generally expected. Chymotrypsin however, occasionally showed wider specificity breaks, e.g. at asparagine residues. This was more common

in large peptides and on more prolonged digestion, indicating that this probably represents secondary cleavage specificity. Trypsin showed very few unexpected cleavages; that giving the Rhodymenia peptide T1(c) being the most obvious. Anomalies in tryptic specificity have been attributed to pseudotrypsin in the enzyme preparations (Keil-Dlouha et al., 1971). No tryptic activity was observed at ε-N-trimethyllysine residues.

Cyanogen bromide cleavage has been used in the determination of several protein sequences (see Kasper, 1970), and this method has been applied previously to cytochrome c (Heller and Smith, 1966b). The method owes its success to its specificity for methionyl residues and to the fact that methionine is relatively rare in proteins. This increases the probability of obtaining large peptide fragments which can provide valuable information important for either the correct positioning of peptides derived from enzymic digests or for the confirmation of previously established sequences. Cleavage of 96% of the susceptible bonds has been achieved using 70% (v/v) formic acid as the solvent (Steers et al., 1965); the method used was that described for β-Galactosidase (Steers et al., 1965).

Cyanogen bromide cleavage of cytochrome c was used as a method of obtaining two fragments of the protein which could then be digested enzymically, thus making subsequent

separation of the peptides easier; also, where only small amounts of protein were available, and the complete sequence could not be determined by overlapping peptide data, this method would give stronger support to the assigning of peptides to the sequence by homology with other cytochromes c (see for example, Brown and Boulter, 1973a). The cyanogen bromide fragments were readily separated by gel filtration but subsequent digestion and paper chromatography gave poor yields of peptides. Direct dansyl-Edman analysis on the smaller, non-haem fragment gave useful sequence data. The use of this method is suggested in cases where only very limited amounts of cytochrome of uncertain purity are available; subsequent sequence analysis of the second fragment by sequenator methods could then give data on some 20% of the sequence which could be used in comparative studies. However, larger quantities of material than were used in this investigation appear to be required to give full sequence information.

High-voltage paper electrophoresis was used to separate peptides as this is preferable to column separation methods when only small quantities of material are available. Paper chromatography proved unsatisfactory probably due to the lack of material and was not routinely used as electrophoresis at pH 6.5 and pH 1.9 gave adequate separation of the peptides.

The dansyl-Edman method for sequencing peptides is an extremely sensitive one and has been used in a micro-form to determine the sequence of as little as 10 pmol of peptide (Bruton and Hartley, 1970). When the method of Woods and Wang (1967) is used to separate the dansyl derivatives by thin-layer chromatography, it also has the advantage of excellent resolution of all normally occurring protein amino acid derivatives, especially leucine and isoleucine. Resolution of unusual amino acids such as ε-N-trimethyllysine, and good separation of the basic amino acids, can also be achieved. The main difficulties of this and other 'end group' techniques are those of identifying amino acids whose derivatives are labile during the acid hydrolysis of the labelled peptide. Both asparagine and glutamine are destroyed to their corresponding acids, and if tryptophan has not been previously destroyed during Edman degradation its dansyl derivative is totally destroyed. Bis-dansyl-histidine is degraded to α-N-dansyl-histidine and after several Edman degradation steps lysine can also be difficult to identify. Dansyl-proline is degraded on prolonged acid hydrolysis and when proline was expected in the sequence the labelled peptide was hydrolysed for a shorter time (6 h). The method is also only qualitative at each step of the Edman degradation, although an indication of the peptide

composition can be obtained from the intensity of the fluorescence of the spots after chromatography of a dansylated acid hydrolysate of the peptide.

Of these difficulties the identification of amides presents the major problem. The method of Offord (1966) using the mobility at pH 6.5 often suffices to assign most amide residues, and this is an added advantage of this separation technique. However, certain residues cannot be placed as amides by this method and, unless sufficient peptide material is available to determine the mobilities of degraded peptides at pH 6.5, amides can only be tentatively assigned by homology with cytochromes where the amide residues have been experimentally determined.

The difficulties encountered with haem peptides, particularly with those of Rhodymenia, lie in the fact that these peptides were routinely recovered in only low yield. It is possible that the haem moiety interfered with nearby susceptible bonds (Margoliash and Schejter, 1966) and that this, coupled with the initial limited amount of material and the performic oxidation of haem peptides was responsible for the lack of information obtained. The technical problems involved in sequence determination and the likely sources of inaccuracy, when using limited amounts of protein, have been discussed by Ambler and Wynn (1973), and these problems were

largely those encountered during this study and which necessitated the analysis, in most cases, of repeated digests.

Sequence Comparisons

The sequences determined so far for higher plants show several characteristics which can be considered as being typical of the group. All consist of a single polypeptide chain of 111-113 residues with a covalently bound haem group. They are longer than any of the other cytochromes c examined, the additional residues being present as an acetylated N-terminal 'tail'. Enteromorpha cytochrome c is thus of the 'higher plant type', whereas the cytochromes c of Rhodymenia and Porphyra do not bear such a striking resemblance to this group. Rhodymenia cytochrome c has the longest N-terminal tail yet found, consisting of 10 amino acids, the longest previously being the 9 residue tail of Crithidia (Pettigrew, 1972). The Enteromorpha sequence differs from higher plant sequences in the possession of only a single ε-N-trimethyllysine residue in position 80. This probably represents a secondary modification of the protein after synthesis is complete, as has been shown for the single residue of this amino acid in Neurospora cytochrome c (Scott and Mitchell, 1969). Enteromorpha differs from higher plants in 12 of the 85 residues that have been shown to be invariant in this group. These are Ser-1, Asp-5, Ala-20,

Ala-43, Thr-47, Ala-51, Phe-54, Thr-64, Ala-65, Ala-91, Lys-94 and Phe-105. However, alanine in position 51 is also reported for Nigella (Brown and Boulter, 1973a) and leek (Brown and Boulter, 1973b). Lysine at position 94 is included in the list since this is ϵ-N-trimethyllysine in the higher plant sequences. The number of 'invariant' residues in all the mitochondrial cytochromes so far determined is now 28; the value of 34 (Boulter and Ramshaw, 1972) having been previously increased to 29 by the publication of the sequences of snail (Brown et al., 1972) and Crithidia (Pettigrew, 1972) cytochromes c. The number of proline residues in the Enteromorpha sequence is similar to that in the higher plant sequences. Proline also occurs more frequently in both the algal and plant sequences than in the animal sequences. The number of basic amino acids in both the plant and algal sequences is lower than that for the animal sequences, and this explains the lower observed pI values for plant cytochromes (Laycock, unpublished experiments; Valentine and Roberts, 1971). The ratio of phenylalanine to tyrosine residues also varies between different taxonomic groups. In the higher plants the ratio is invariably 4 Phe-6 Tyr, as compared with 6 Phe-4 Tyr for Enteromorpha, as is also the case in Neurospora (Heller and Smith, 1966b). In Rhodymenia the ratio is apparently 6 Phe-4 Tyr.

No positive evidence of heterogeneity in the algal sequences studied was obtained. This has not always been the case with the other plant cytochromes c and heterogeneity was observed in Castor and Sesame (Thompson et al., 1970), Pumpkin (Thompson et al., 1971b) and Rape (Richardson et al., 1971).

Heterogeneity may be due to any of four different causes:- (1) The gene specifying the polypeptide may exist as a heterozygote at one or more loci; (2) polymorphism, due to inter-cistronic differences of the gene specifying the protein either within or between populations; (3) Translational ambiguity of the messenger RNA template (Carbon et al., 1966; Rifkin et al., 1966); or (4) as an artefact caused by modification of the protein during extraction, purification or sequencing.

With the exception of the iso-cytochromes -1 and -2 of Baker's yeast (Slonimski et al., 1965), little evidence has been found for different forms of cytochrome c in an individual. Mules and hinnies carry 50% horse and 50% donkey cytochrome c, which differ by a single residue (Walasek and Margoliash, 1969). Matsubara and Smith (1962) report that using a mixed population of 70 individuals leucine was found to replace methionine in position 65 in 10% of human cytochrome c. Attempts to demonstrate poly-

morphism in cytochrome preparation from 12 horses and 18 humans proved unsuccessful (Margoliash, 1969a), and Stewart and Margoliash (1965) showed that cytochrome c preparations from different organs of the same species were identical. The sequencing methods used by Margoliash (1969b) were estimated to be able to detect 5% heterogeneity if it was present as a single change. The dansyl-Edman method requires a greater degree of heterogeneity than 5% for it to be detectable, especially in cases where the minimum of material is being used. The Enteromorpha source certainly represented a mixture of species yet the dansyl-Edman method gave a unique sequence.

Certain specific points about the sequence determination of the individual species studied deserve comment. Enteromorpha cytochrome c was shown to have an acetyl blocking group at the N-terminus, indicating further its homology with higher plant cytochromes. Rhodymenia cytochrome c is interesting in being the first plant cytochrome to be sequenced without a blocking group and also in its possession of the 10 residue N-terminal tail. This may be characteristic of the red algae as Porphyra was also shown to possess an unblocked N-terminal alanine. Despite the investigation and purification of peptides from several digests, several peptides of Rhodymenia were

clearly not detected. In the Porphyra sequence a substitution of glycine for tyrosine in position 73 was indicated. This position is invariably tyrosine in the higher plants and more commonly methionine in the animals. In Lamprey, the insects, snail and Neurospora, position 73 is phenylalanine, and Candida, Saccharomyces and Debaryomyces have serine (see Dayhoff, 1972).

Strong evidence for the presence of two tryptophan residues in the red algae was obtained. For both Rhodymenia and Porphyra cytochrome the best spectral ratios of E_{280}(ox)/ E_{550}(red) were found to be 1.05 and 0.93, although in the case of Rhodymenia the protein was apparently pure as judged by polyacrylamide gel electrophoresis. In Rhodymenia the presence of two tryptophan positive tryptic peptides was indicated, although only Trp-5, Fragment III could be positively identified. The other tryptophan residue was present in peptide T5 and has been tentatively assigned to Fragment II, position 7. The corresponding chymotryptic zones gave no clear evidence for this assignment. Indirect evidence for the occurrence of tryptophan in the same position in the Porphyra sequence is given by peptide C1 (Gly-Leu-Phe). This assignment of tryptophan places it as homologous with position 41 of other cytochrome sequences (see Appendix 8), in which position tryptophan has also been recorded for

Euglena, Bonito, Tuna and Carp cytochromes c.

Two peptides were characterized in Rhodymenia chymotryptic digests, which could not be assigned to the sequence either by corresponding overlapping tryptic peptides or by homology. It is possible that these peptides may fill the 'missing links' in the sequence.

No Є-N-trimethyllysine residues were detected in the Rhodymenia sequence either from amino acid analysis or during sequence analysis. Porphyra cytochrome c also appears not to contain this amino acid. If, as has been shown by Scott and Mitchell (1969), the methylation of lysine occurs after the initial synthesis of the protein, then the absence of Є-N-trimethyllysine in the red algae implies a lack of the methylating enzymes, and may imply a fundamental biochemical difference between this group and the green algae and higher plants. An indication of the specificity of the methylating enzymes of cytochrome c may be obtained by a study of the Enteromorpha sequence. The sequence around the Є-N-trimethyllysine residue in position 80 in Enteromorpha is identical with that of the higher plants, which also have position 80 as this amino acid, i.e. -D-Y-L-L-N-P-J-K-Y-I-P-G- (see Appendix 7 for code). For position 94, in which the lysine is methylated in higher plants, the sequence is M-V-F-P-G-L-J-K-P-Q-D, but in

Enteromorpha a substitution of alanine for proline in position 91 is accompanied by no methylation of the lysine at position 94. Crystallographic studies with animal cytochrome c have shown that both of these lysine residues are on the surface of the molecule (Dickerson, 1971) and are, therefore, presumably susceptible to modification by methylating enzymes in the native protein. The three-dimensional shape of plant cytochrome c has been shown to be very similar to that of the animal protein (Boulter and Ramshaw, 1972). Proline is known to be important in affecting the tertiary structure of cytochrome c (Dickerson et al., 1971) and a proline substitution is liable to permit modifications in the conformation of the protein. No animal cytochrome c sequence is reported to contain ϵ-N-trimethyllysine and this may be due to either, (1) absence of cytochrome c-lysine-methylating enzymes, or (2) conformational changes due to amino acid substitutions affecting enzyme specificity. The latter cause is not indicated, as Neurospora crassa cytochrome c has been shown to contain ϵ-N-trimethyllysine in position 80 and the surrounding sequence is identical with the animal sequences (Delange et al., 1969). It would therefore appear that the enzymes responsible for the methylation of lysine in plants and fungi are not present in animals. The case for the absence of similar enzymes in the red algae is

strengthened by the identical sequences of Rhodymenia, Porphyra and higher plant cytochromes c around Lys-80 with no methylation of this residue in either Rhodymenia or Porphyra. Pettigrew (1972 & 1973) has shown that ϵ-N-trimethyllysine is also present in protozoan cytochromes c, at position 80 in Crithidia and position 94 in Euglena, and the situation appears to be more complicated in these species. Although a Phe-Met sequence in Crithidia is substituted for a Tyr-Val sequence in Euglena in positions 82 and 83 (see Appendix 8), the sequences around position 94 are identical, yet Lys-80 in Euglena and Lys-94 in Crithidia are non-methylated. The presence of ϵ-N-dimethyllysine has been observed in cytochrome c from the thermophilic fungus, Humicola lanuginosa, and this was considered to represent incomplete methylation of the lysine residue (Morgan et al., 1972). However, various ϵ-N-methyllysines have been observed to occur in other proteins, e.g. histones (Hnilica, 1967; Paik and Kim, 1967) and flagellar proteins (Ambler and Rees, 1959), and presumably do not represent incomplete methylation in all cases. Scott and Mitchell (1969) report no finding of mono- or dimethyllysine derivatives during the formation of ϵ-N-trimethyllysine in Neurospora. It is suggested that in Humicola the ϵ-N-dimethyllysine does not represent an intermediate stage in ϵ-N-trimethyllysine synthesis and that methylation of lysine does not occur

sequentially as this would involve a series of enzymes with different charge and presumably steric specificities towards the lysine residues. It thus appears that the occurrence of ϵ-N-trimethyllysine in cytochrome c is more widespread than only the higher plant and fungal cytochromes, as quoted by Lemberg and Barrett (1973). The functional significance of the methylation is, however, as yet unclear, although Scott and Mitchell (1969) have suggested that for Neurospora it may be involved in the binding of the protein to the mitochondrial matrix.

The sequences of the cytochromes from algae showed remarkable similarities, both among themselves and with other cytochromes c from eukaryotic mitochondrial sources. The mitochondrial cytochromes c are chemically and physically all very similar, functioning as electron carriers in the mitochondrial respiratory electron transport chain (see Margoliash and Schejter, 1966; Keilin, 1966). This is amplified by the reports that cytochromes c of other species can successfully replace the cytochrome c of the same species in its intra-mitochondrial reaction with its cytochrome oxidase (Jacobs and Sanadi, 1960; see also Margoliash et al., 1971). This interchangeability is however, not unlimited. Yamanaka and others (Yamanaka and Kamen, 1965; Yamanaka 1966, 1967; Yoshioda, 1966; Yamanaka et al., 1969) have

attempted to make evolutionary deductions from the relative rates of interactions of the cytochromes with mammalian cytochrome oxidase (cytochrome aa_3) on the one hand, and that of *Pseudomonas*, on the other. The results obtained using *c*-type cytochromes from algae are probably due to the use of photosynthetic *c*-types (c_6) and not the homologous mitochondrial *c*-types as described in this study.

As yet little is known of the relationship of the mitochondrial cytochromes *c* with the cytochromes which function in anaerobic energy-yielding reactions and which occur in chemosynthetic bacteria, nor of those which function in the photo-reduction processes of photosynthetic organisms (Kamen *et al.*, 1971). All these contain haem c as the prosthetic group (Dickerson, 1971); however, with the exceptions of those noted in Table 17, sequence determinations have been confined largely to eukaryotic mitochondrial cytochromes *c*. The data of Dus *et al.* (1962) and particularly those of Dus *et al.* (1968) for the cytochromes cc′ and c_2 of *Chromatium* and *Rhodospirillum* respectively showed strong evidence for homology. Also, Dickerson (1971) has found strong evidence for homology of *Pseudomonas* cytochrome c_{551} with horse heart cytochrome *c* if a deletion of 16 residues in a hairpin loop is assumed. A photosynthetic cytochrome *c* from the algae *Monochrysis lutheri* has also been shown to

have some sequence homology with mitochondrial cytochrome c (Laycock, 1972).

Evolutionary Aspects

Similarities between sequences could have arisen, (1) by chance (parallel or coincident change); (2) by convergence of correlated proteins during evolution, certain amino acid sequences being more efficient for a common biological function (analogy); (3) by descent from a common ancestor (homology).

The evidence now available points to there having been a common ancestor for all cytochromes c and thus evidence for a divergent homologous, rather than a convergent analogous evolution having taken place. The major work in the study of the evolution of mitochondrial cytochrome c has been carried out by Margoliash and his co-workers (Margoliash, 1963; 1964; Margoliash and Smith, 1965; Margoliash and Fitch, 1968) on animals, and by Boulter and his colleagues (Boulter, 1972, 1973; Ramshaw et al., 1972) on plants. The whole field of the evolutionary aspects of the primary structure of proteins has been reviewed by Nolan and Margoliash (1968) and more recently by Lemberg and Barrett (1973).

Fitch (1970) and others (see Fitch, 1966a and b;

Cantor and Jukes, 1966), by using a statistical approach, showed it is possible to distinguish between the first process quoted above and the other two possibilities. The method involves the comparison of the frequency of occurrence of the minimum mutation distances between various peptide segments of determined length with those expected from a purely random situation. However, Dickerson (1971) has pointed out that the method of evolutionary homology based on the 'minimum mutation distance' has an inbuilt error when used for detecting otherwise invisible homologies. This is due to the degeneracy of the genetic code. A change in the third purine or pyrimidine base seldom changes the amino acid coded for and alteration of the first base usually converts one amino acid to another of similar chemical or physical properties, the alteration of the second base being the most influential in changing the chemical character of the amino acid residue (see for example, Crick et al., 1961; Crick, 1966).

However, even when a departure from a random situation has been shown for mitochondrial cytochromes c (Margoliash et al., 1969), it is necessary to distinguish between convergence and homology. Fitch and Margoliash (1967b) have presented an approach for use with a set of proteins in which many invariant residues exist. The method is to

statistically determine the expected number of residues that will be found to be invariant, irrespective of the number of sequences determined; two sequences are then held to be ancestrally related if the number of identical residues between the two exceed the statistically determined value. Fitch and Margoliash (1967b) calculated the expected number of invariant residues for eukaryotic cytochromes c to be 27-29; and recent more sophisticated calculations (Fitch and Markowitz, 1970) have given a value of 32 invariant residues. The number of observed invariant residues between the eukaryotic cytochromes c is presently 28 (see Brown et al., 1973) and it is thus concluded that they are all ancestrally related.

In the comparison of sequences a further difficulty arises in the cases where additions or deletions of residues have occurred. Certain methods to show departure from randomness between sequences show the probability of homology or convergence without the need to consider such events (Fitch, 1969, 1970; Gibbs and McIntyre, 1970). For the mitochondrial cytochromes c the numbers and sizes of these are small and may be readily identified. For example, the additional residues in the plant sequences are clearly located at the N-terminus when the two cysteine residues are aligned (see Dayhoff, 1972). Even in the

protozoan Crithidia cytochrome c where only one cysteine is involved in the binding of the haem (Pettigrew, 1972), alignment of the Gln-Cys sequence at the haem attachment point, which is now considered invariant (Lederer et al., 1972), also places the extra residues at the N-terminus.

In attempting to compare the similarity of more distantly related c-type cytochromes particular difficulties arise in this correction, and especially when considering those not belonging to the same group. The sequences of Pseudomonas fluorescens c_{551} (Ambler, 1963) and Rhodospirillum rubrum c_2 (Dus et al., 1968) have been aligned with the sequences of mitochondrial cytochromes c, but the proposed alignments disagree with one another (Cantor and Jukes, 1966; Dus et al., 1968; Needleman and Blair, 1969; Dickerson, 1971). Ambler and his co-workers have aligned various bacterial cytochrome sequences and some similarity is evident within this group (see Ambler et al. in Dayhoff (1972); Ambler, 1971; Ambler and Wynn, 1973).

The only certain way to test the validity of such comparisons is by a study of x-ray crystallographic analysis of the three-dimensional structure of the protein molecules. An experimental approach of this nature offers a solution to the differentiation between convergence and homology. If the structure determination shows that the three-dimensional structure of the two molecules is essentially the same, this

is much more likely to have arisen by homology; however, the minimal sequence constraints required by function would not necessarily require the same tertiary structure over the entire protein molecule.

So far the data obtained for cytochromes c indicate that they are closely related in terms of their tertiary structure. Dickerson et al. (1971) have determined the three-dimensional structures of horse heart and bonito cytochrome c and studies are in progress on the tertiary structures of spinach (Morita et al., 1973) and rice (Morita and Ida, 1972) cytochromes c.

Dickerson (1971) illustrates the usefulness of tertiary structure determinations in postulating the location of deletions in the sequence of Pseudomonas fluorescens c_{551}, when compared with horse heart cytochrome c. Despite only 25% similarity between the sequences the proposed alignment is fairly clear. The plant and algal cytochrome sequences show a clear fit with the structural constraints determined for the horse heart and bonito proteins; many major points of similarity exist between the sequences which have been shown to be basic structural or functional requirements in the horse heart cytochrome (Dickerson, 1971). It is, therefore, concluded that the tertiary structures of the cytochromes of both plants (including the algae) and animals are essentially

identical, indicating a common ancestry for both.

There are many features of the plant and algal cytochrome c sequences which give a great deal of confidence in this conclusion; however, only a complete x-ray investigation, such as that being carried out for rice and spinach, will verify the identities in tertiary structure.

Dickerson et al. (1971) showed the horse heart cytochrome to be a classic example of the "oil drop" model of a protein with a hydrophobic interior and polar exterior. The polypeptide chain is wrapped around the heme group in two halves, thus forming a hydrophobic crevice in which the heme group is located. Two hydrophobic "channels" filled with hydrophobic side-chains lead right and left from the heme to the surface of the molecule, where there are specifically grouped charged residues. The horse heart, bonito, plant and algal proteins show large stretches of identical residues and where differences are found these are generally of a chemically conservative nature. The essential residues required by the heme group are the two cysteine residues which are covalently bound to the porphyrin ring. The heme is apparently the principal folding influence on the molecule and the configuration is maintained such that histidine-26 and methionine-88 residues provide the fifth and sixth ligands for the iron atom. Generally, the

sequences of plant and algal cytochromes are identical to the horse heart sequence in those regions and certain isolated amino acids that are held to be important in maintaining the tertiary structure. The number of invariant residues in all mitochondrial cytochromes c now sequences is 28 and these are the positions that have been cited as being important for the maintenance of the three-dimensional structure.

Despite the striking similarities several notable differences in the properties of plant and animal cytochromes still exist.

The stability of plant cytochromes during preparation is less than that for animal cytochromes; they are more readily denatured by ethanol and digested by proteolytic enzymes, although the greater similarity of Saprolegnia cytochrome c to the animal protein in these respects was apparent in this study. The ratio of the absorbances at 416 nm(red)/550 nm(red) are higher and this possibly suggests a more open tertiary structure (Stellwagen, 1968), which may also account for the other observations.

Compared to other cytochromes the plant and algal cytochromes are longer and Rhodymenia exhibits the longest cytochrome c yet studied. Thus, it would appear that the cytochrome gene has shortened during evolution (Smith, 1968). In animals and fungi there is strong evidence that

cytochrome c is synthesized on 80S cytoplasmic ribosomes (Scott and Mitchell, 1969; Boulter, 1970), and it has been demonstrated that the mechanism of initiation of protein synthesis on these particles involves $tRNA_F^{Met}$ (Smith and Marcker, 1970). This would imply that the N-terminal amino acid of cytochrome c is initially methionine and that subsequently a hydrolytic enzyme is responsible for the cleavage of residues from the N-terminus. The specificity of the postulated enzyme may thus be responsible for the variations in sequence length at the N-terminus. The presence of the acetyl group in many cytochromes may be linked with this process, and it would appear that its presence is not for removing the otherwise free amino group from the heme environment (Dickerson et al., 1971), as had been previously suggested (Margoliash and Smith, 1965).

Despite the large amount of evidence, it is still necessary to assume homology between plant and algal and other eukaryotic cytochromes, since it is still impossible to rigorously rule out the possibility of convergence having occurred. Assuming, however, that homology has been sufficiently demonstrated, it is possible to examine the molecular evolution of the cytochrome gene. As the cytochrome c molecule is an expression of a small part of the genome of the species from which it was obtained, then it follows

that relationships between the sequences could relate directly to the relationships between the species themselves. It is thus possible to derive the relations between the various kingdoms and also between the individual members of each kingdom from the relationships which exist between their cytochromes.

The first sequenced cytochromes c were mainly from animal species (Dayhoff, 1969). This data showed that when the cytochromes c of members of a taxonomic class are compared with each other, the number of amino acid differences between them is less than that found when cytochromes c of members of different taxonomic classes are compared. The number of differences becomes greater the more distantly related the organisms. Within a class the number of differences between its members lies within a range. This contrasts strongly with the more or less constant difference found in each interclass comparison, irrespective of which members from each of the classes are compared. This constancy of difference in interclass comparisons is interpreted to imply that elapsed time is an important parameter in determining the number of mutations which accumulate in the structural gene for cytochrome c along any line of evolutionary descent (Zuckerkandl and Pauling, 1962; Nolan and Margoliash, 1968).

It was suggested that over long periods of evolutionary

history, perhaps a minimum of 200 million years for cytochrome c, other factors relating to the rate of fixation of amino acid differences have either cancelled or averaged themselves out, leaving elapsed time as the variable most clearly related to differences between primary structures. For such an empirically derived relation however, it is unnecessary to assume any of the mechanisms through which the differences may have arisen (Margoliash and Schejter, 1966). Accepting the relation enables a direct correlation to be made between time and the rate of change of cytochrome c (Margoliash, 1963).

The times of divergence of many of the major classes of animals can be established from the fossil record (see for example, Colbert, 1969), and from these it is possible to derive the time taken, on average, for a single difference to become established between the sequences of different lines of descent. This time is called the 'unit evolutionary period' and refers to a given protein (Nolan and Margoliash, 1968).

However, before a set of sequences can be used to derive a unit evolutionary period, it is necessary to demonstrate that the proteins belong to an homologous set.

Fitch and Margoliash (1970) draw a distinction between "paralogous" and "orthologous" genes, when establishing gene

or species phylogenies. Both classes of genes are homologous, but an important difference exists between the two. Parologous genes are those which, although originally identical, may, after duplication, have subsequently diverged to the extent that they code for proteins with different functions. Orthologous genes, on the other hand, are those which have remained identical in function throughout. When the plant and algal proteins are included along with the animal ones and the calculations described by Fitch and Margoliash (1967b) applied, they can be shown *in toto* to constitute an homologous set. Furthermore, apart from iso-cytochromes c_1 and c_2 of Baker's Yeast, there is no evidence for gene duplication having occurred with the gene responsible for specifying mitochondrial cytochrome *c*. Therefore, these cytochromes *c* are not only homologous but orthologous in the sense discussed by Fitch and Margoliash (1970). This means that all cytochrome *c* data used in this study that satisfied the requirements of Fitch and Margoliash (1970), can be used legitimately to calculate unit evolutionary periods.

The parologous condition may exist between the mitochondrial cytochrome *c* series and certain of the cytochromes *f* (c_6) in cases where the molecular weight is known to be a multiple of that of cytochrome *c* and several heme groups

per molecule are apparent; both however, share the homologous function of electron transfer.

In this investigation 'unit evolutionary periods' have been calculated using both amino acid differences and minimum mutation distances. However, for Rhodymenia cytochrome c the data was calculated only on the amino acid differences assuming that the amino acid substitutions in the unknown fragments occur at the same frequency as those in the known sequence of the protein.

Since the times of divergence of the major groups of the plant kingdom cannot be established from fossil evidence, 'unit evolutionary periods' have been calculated using fossil datings from the animal kingdom. The time value which has been used to derive the results given in Table16, is the point in time of the divergence of the mammalian and avian lines of descent, which is believed to have occurred 280 million years ago. When the available sequences of members, one from each of these groups, were compared in all possible combinations the average amino acid difference found was 9.9 and the average minimum mutation distance was 13.7. These values gave 28.3 and 20.5 million years respectively for the 'unit evolutionary periods' for cytochrome c. These data have been extended linearly to give divergence times for other taxonomic groups in the

plant and animal kingdoms (see Table 16). However, linear extrapolation will always give an underestimate of the time values since multiple mutations at a single codon, parallel and back mutations occur. It is possible to correct partially for these events, however, by using the statistical method of Feller (1950) and these corrected values are also given in Table 16 (see also Fig. 23).

The use of minimum mutation distance partly corrects for the probability of multiple mutations at a single codon (Margoliash et al., 1963), but fails to take account of back mutations or multiplicities in pathways which exist. Fitch and Margoliash (1969) estimate that of the mutations which have occurred in animal cytochromes 20% are parallel and 1% are back mutations. A statistical correction has been applied (Margoliash and Smith, 1965) but assumes that all the variable sites are equally liable to effective amino acid substitutions, and this certainly is not the case for cytochrome c, in which certain positions are considerably more variable than expected if a random situation existed (Fitch and Margoliash, 1967b). However, these corrections probably give a more accurate idea of the times of divergences.

The justification in applying the 'unit evolutionary period' derived from considerations of animal data to calculate divergence lines for the plant kingdom, is largely

theoretical. Using the 'unit evolutionary period' calculated from the bird-mammal fossil dating, the values obtained by extrapolation for the times of divergence of any two animal lines are approximately the same as those based directly on the fossil record (Young, 1962; Colbert, 1969).

This correlation provides good support for an approximately constant rate of evolution of cytochrome c (linearity) over at least the last 750 million years, during which period the divergence and evolution of the higher plant kingdom from its ancestral stock occurred. Since the algal sequences are homologous with those of higher plants, animals, and fungi, it seems not unreasonable, in the absence of plant fossil datings, to apply the 'unit evolutionary period' calculated from the bird-mammal fossil dating. However, Cronquist (1968) has pointed out fundamental differences in the evolution of higher plants and animals, particularly that evolutionary pressure on vertebrates has been towards adaption to a niche, whereas in the Angiosperms single families usually occupy a variety of different ecological niches. Despite the great degree of convergent morphological evolution apparent in the algae, their present-day distribution would seem to indicate that their evolution has been niche-orientated to a higher

degree than that of the Angiosperms. The importance of these considerations to the present considerations cannot be assessed, particularly since the mechanism of selection and fixation of mutations at the biochemical level is uncertain.

Assuming the application of the animal data to the plant kingdom is justified, the time of divergence of the three eukaryotic kingdoms can be estimated at about 1500 x 10^6 years ago. Considerations of cytochrome and tRNA data have led to the calculation of the divergence of prokaryotic and eukaryotic lines as being 2.6 times more remote than the divergence of the eukaryotic kingdoms (McLaughlin and Dayhoff, 1970). A similar value was obtained using only tRNA data (Jukes, 1969).

The time of divergence of the Protistan branch from the main eukaryotic line is estimated at 2900 x 10^6 years ago. This estimate is 1000 x 10^6 years more recent than the estimate for the divergence of the eukaryotic/prokaryotic lines of descent. Fossil records for this period are unlikely to amplify our knowledge of these events and such calculations are largely speculative. Fossil red algae are well known from the Cretaceous and Triassic periods, and a few questionable forms from the Ordovician have been recorded (see Scagel et al., 1965). The estimate of

1760 x 10^6 years ago places the divergence of the red algae long before this in the middle Pre-Cambrian. The point of divergence of the higher plants from the green algae is placed at 700-800 x 10^6 years ago, and this represents a much earlier estimate than most authors who place the origin of the flowering plants in the Jurassic. However, Takhtajan (1969) has suggested an earlier origin.

The 'unit evolutionary periods' derived for cytochrome c show that it is accepting mutations more slowly than almost all other proteins so far examined (see Dayhoff, 1972). This would suggest that severe structural constraints exist for the cytochrome c molecule, and the reasons for this are evident from the studies on tertiary structure (Dickerson et al., 1971) and the minimum of three interacting functions with oxidase, reductase and membrane (see also McLaughlin and Dayhoff, 1973).

From the evolutionary standpoint, the potentially most useful computation, based on quantitative comparisons of the amino acid sequences of sets of homologous proteins, is that which leads to the construction of phylogenetic trees. The problems involved in the construction of such trees is the enormous number of alternative trees which are possible and ideally have to be evaluated. Even with a computer the assessment of all trees is an almost impossible task;

sixteen species for example, give rise to about 10^{14} alternatives. Therefore, methods must be used which drastically reduce the number of trees to be evaluated, while still finding the optimum solution.

A simple reduction can be made if the available sequence data is split into kingdoms on the basis of biological information. The tree for each kingdom can then be constructed separately and then combined to give a final tree. Even so, further reductions beyond the kingdom level are necessary, e.g. the sequences may need to be grouped into classes within the kingdoms.

There are basically two types of approach to this problem that have been described; these are the 'numerical matrix' method (see for example, Fitch and Margoliash, 1967a; Gibbs and McIntyre, 1970), and the 'ancestral sequence' method (Dayhoff and Eck, 1966).

The method used in this investigation was the 'ancestral sequence' method. The basic tree used was that as computed in Dayhoff (1972), and this was updated to include the higher plant tree of Boulter *et al.* (1972), the protozoan sequences (Pettigrew, 1972 & 1973), and the fungi, *Humicola lanuginosa* (Morgan *et al.*, 1972) and *Ustilago sphaerogena* (Bitar *et al.*, 1972). Other fungal sequences were corrected on the data of Lederer (1972) and Lederer *et al.* (1972).

The following account relies heavily on the simplified explanation of tree construction given in Boulter et al. (1972).

The tree consists of branches, the junctions of which are called nodes. Each node has three branches which either lead to an adjacent node or to a determined sequence. In addition, by using the rules described below, it is possible to reconstruct the most probable ancestral sequences of each node, so that the tree contains not only the determined sequences of extant species, but also the computed ancestral sequences. The point of earliest time on the tree cannot be established from sequence data and must be established by biological considerations.

The computing strategy employed was based on that described by Dayhoff (1972) and the program used in calculating the tree consisted of three main procedures. The initial construction starts with any three sequences; only one tree exists relating these. The next sequence is added to the tree in all possible positions and the best resultant topology is selected and used for further tree construction. Additional sequences are added successively to the best tree obtained at the previous step until a final tree is obtained. Alternatively, different species can be added singly to the same tree to check the effect of the position of one species on another. The second computer procedure is the evaluation

of the trees. It starts by inferring the 'ancestral sequence' at each of the nodes which relate the sequences to the tree. Each position along the sequences is considered in turn. For each node three lists are made which consist of the amino acids found in this position along the three branches attached to the node. Thus, at each nodal position all the sequences in the tree are considered. This is repeated in turn for all the positions in the sequence. If, for a given node and at a given position in the sequence, only one amino acid is found which occurs on more branch-lists than any other, then it is selected as the nodal amino acid; otherwise, at this stage the nodal position is left blank. When all the nodes have been assigned an amino acid or a blank for every position in the sequence, the situation at each blank position is re-assessed. If a blank position has at least two of its three adjacent neighbours (either node or sequence) the same, then this amino acid is assigned to the position. The process is repeated until no more additions occur at any of the blank positions. Finally, the nodal sequences are checked so that if the amino acid at the node is not the same as at least two of its adjacent neighbours, it is changed to a blank. This process gives a definite assignment of the ancestral amino acid whenever one choice is clearly preferable, but will leave blanks when

reasonable doubt exists.

When the ancestral sequences at all the nodes have been determined, the tree is evaluated as follows: the numbers of amino acid changes along every branch of the tree are counted by comparing each sequence with each adjacent sequence, position by position. All the branches are then totalled to give the evaluation for the whole tree. This calculation is, however, complicated by the existence of any blanks in the ancestral sequences. The blank means that two or more equally probable amino acids exist for that position. In such cases it does not matter for the overall evaluation which amino acid is chosen to fill the blank, as the number of total changes on the tree will be the same if any of the alternatives is used, unless two parallel mutations exist at adjacent nodes. Because a minimum route is assumed for evolution, in this case the same amino acid is chosen for both blank nodal positions. When the branch lengths were evaluated for the final tree (Fig.19) the minimum number of mutations counted around a blank or series of blanks was divided equally among all the independent branches and the values thus assigned shown in brackets.

The third procedure evaluates the positions of branches of the tree in alternative positions. This is necessary because a sequence, once fixed to the tree, does not change

its relative position and therefore, the fusion strategy of the initial procedure effectively limits the amount of evaluation required. Because no account is made of the remaining sequences still to be added when each new sequence is fixed, this means that, in retrospect, a wrong decision may be made at any step and the final tree may only be a close approximation to the "best" tree. This procedure then, evaluates the positioning of the various branches in alternative arrangements to see if a better tree exists.

The 'ancestral sequence' method involves two assumptions: first, that evolution has taken place by the minimum number of amino acid substitutions, and secondly, that the final tree accepted is that which has the minimum number of amino acid substitutions of all possible trees. With regard to the first of these assumptions, it is evident that back- and parallel-mutations have occurred during the evolution of cytochrome c. Boulter et al. (1972) detected 27% parallel and 6% back mutations during the evolution of the higher plants, and Fitch and Margoliash (1969) estimated 20% and 1% respectively, for the same processes in the evolution of animal cytochrome c. The problem lies in the fact that there is no certain way of knowing to what extent undetected parallel and back mutations occur, especially when the number of sequences in a given taxonomic unit is small.

However, the likelihood of relating two sequences which are similar as a consequence of parallel and/or back mutations, rather than because of common ancestry, is limited, since in each comparison all positions of the sequence are considered.

With regard to the second assumption, because of the large number of possible trees, clearly it cannot be certain that the selected tree has the fewest possible amino acid substitutions. However, computer procedure three is specifically designed to reduce the number of comparisons which have to be made in order to obtain, with reasonable certainty, the absolute minimum tree.

The tree shown in Fig. 19 was the best overall tree obtained by these methods. The species used in the computation of the tree fell clearly into four groups:- plants, animals, fungi and 'lower' organisms. The latter group comprised only three species, Euglena, Crithidia and Rhodospirillum. The three possible arrangements of these four groups were evaluated (Fig. 21) and the chosen minimum topology is that shown in Fig. 19. However, the degree of difference between the scores for these topologies indicates that no single arrangement is clearly the best or absolute minimum tree.

Dividing the known sequences into the kingdoms as posulated by Whittaker (1969) gives the same minimum topology.

There are fifteen possible combinations of trees involving five groups, and these were evaluated as shown in Fig. 22. The minimum tree is in agreement with that computed by McLaughlin and Dayhoff (1973), which does not include the Euglena and Enteromorpha data.

The small degree of difference between the trees shown in Figs. 21a and 21b and those shown in Figs. 22a and 22d, and would seem to indicate that the major groups, particularly the fungi, plants and animals, diverged from each other within a relatively short space of time. However, the branch lengths joining the major nodes of the tree as calculated by McLaughlin and Dayhoff (1973), indicate a much remoter ancestry of these groups if the 'unit evolutionary period' is applied to them. These high values are due to the assignment of values to blank residues in the ancestral nodes; the number of blanks increase in number when not many sequences are available to establish kingdom ancestral nodes. Where a blank exists the minimum number of changes that must have occurred is averaged out along the branches concerned. This method probably does not give an accurate assessment of blank areas and where several blanks exist in one area of the tree may considerably over-estimate node to node distances. The variation in distances between those calculated during this investigation and those quoted by

McLaughlin and Dayhoff (1973), are presumably due to differences in the 'weighting' of blank residues. The blanks in the ancestral nodal sequences are often due to the deletion or addition of more than one amino acid in the sequences used to compute the nodes. These are considered as multiple mutational events by the ancestral sequence method. However, there is no certain method of determining if these are due to multiple events or single events affecting a larger region of the cytochrome c gene. Alternatively, as many blanks exist at the N-terminus of the protein, the deletion of residues in this region may be due to a difference in the specificity of the hydrolytic enzyme responsible for cleaving the methionine N-terminal amino acid after the initial synthesis of the cytochrome c polypeptide chain. These two alternatives may be responsible for the unexpectedly high values for internode distances between the major kingdoms. For this reason the values assigned to branch lengths involving the evaluation of blanks in ancestral sequences in this thesis, are expressed in brackets after the actual real value calculated from definitely assigned amino acids (see Fig. 19). The occurrence of blanks arises largely due to the lack of sequence information available to compute ancestral node sequences. The addition of the Protozoan species to the

list of known sequences has reduced the number of blanks that have to be considered in ancestral nodes, and it is predicted that as more sequence information becomes available the number will be further reduced. This is particularly necessary if an accurate evaluation of the more remote nodes is to be made.

The branch lengths in Fig. 19 are expressed in 'accepted point mutations' or PAMs. When pairs of sequences are compared as in the ancestral sequence method, this approach probably gives a more accurate estimate of evolutionary distance than do matrix methods. PAMs were calculated by the method of Dayhoff (1972). For cytochrome c elapsed time can be calculated from PAMs on the basis of 3 PAMs : 100 million years (Dayhoff, 1972). However, at the present time with insufficient sequence data for the more remote nodes of the tree the matrix methods allow a greater number of comparisons to be made and probably give a more accurate estimate of evolutionary distance.

Having established the minimum tree as shown in Fig. 19 several trials were run to determine the position of the red algae. The red algal sequences were added to the tree shown in Fig.19 as single species. The trees obtained (Fig.20) show that two equally acceptable positions for Porphyra were obtained (Figs. 20b and c). This is almost

certainly due to the lack of sequence information presently available. Rhodymenia showed only one acceptable topology (Fig. 20a) and when Porphyra was added to this tree, only one arrangement was found to give a minimum score (Fig. 20d). Because incomplete sequence data was available the computing strategy employed to compare the known sequence fragments with the homologous sections of other sequences, was as follows. For every unknown position in the Porphyra and Rhodymenia sequences an amino acid, which has never occurred in those positions, was substituted. The amino acid chosen throughout for convenience for ∈-N-trimethyllysine (J). In this way the computer does not recognise this amino acid as being one of the possible alternatives in these positions and, therefore, makes no comparison with the corresponding positions in other sequences. Because only partial sequences for the red algae were used, they have not been included in Fig. 19 as the branch lengths leading to them cannot be accurately calculated. However, it is possible to predict that the branch to the red algae occurs approximately half-way between nodes 40 and 44, and that the branch length to Rhodymenia from this point is 20 PAMs (see Dayhoff, 1972).

The tree in Fig. 19 is drawn without regard for evolutionary time. The branch lengths and overall topology

remain the same whatever time considerations are given. The fungal and animal kingdoms adjoin the same node which is then indirectly linked to the plant, protistan and bacterial kingdoms (see Fig. 22a). The position of certain individual groups on the tree is interesting. The green algae, represented by the Enteromorpha sequence, are placed on the direct line of descent of the Spermatophyta, as is in accordance with general biological opinion. However, the Euglenophyta, as represented by the one sequence, are given a more remote ancestry from the green algal/higher plant line and are grouped on a separate branch with the Protozoan, Crithidia. This raises two points. The first is that the close relationship of the Euglenophyta with the Chlorophyta is questioned. This is postulated by, among others, Christensen (1962, 1966); Fott (1965); Scagel et al. (1965); and Nichols (1970). The second is that two lines of divergence on the tree involving photosynthesis are indicated.

The relationship of the euglenoids with the Chlorophyta has been questioned by Honigberg et al. (1964), and these authors place the euglenoids as a class within the Protozoa. The various modes of nutrition of the euglenoid flagellates have been reviewed by Leedale (1967). These include obligate photo-autotrophy, saprotrophy and phagotrophy. With such a variety of nutritional types involving methods

other than photosynthesis, the linking of the euglenoids with the green algal/higher plant line on the basis of chlorophyl b occurring in some forms, would appear to be largely conjectural. However, at this stage there is insufficient protein sequence data to infer definite phylogenetic relationships between the euglenoids and other Protozoa, although it would appear that such a relationship may exist.

With the placing of Euglena and Crithidia on the same branch the postulated ancestor of the eukaryotic kingdoms proposed by McLaughlin and Dayhoff (1973), must now be revised. This is held as being non-photosynthetic, photosynthesis having evolved along the green plant line and also along the prokaryotic bacterial line. The present work suggests that this hypothetical ancestor may well have been photosynthetic and that certain lines of descent within the Protista lost this faculty to a greater or lesser extent, providing the range of nutritional types presently found within this group (see for example, Leedale, 1967). The lines of descent leading to the animals and fungi can thus be held to have lost the ability to photosynthesise completely.

The position of Euglena on the tree can also be considered with respect to the best topology found for the

red algae. The chosen topology (see Fig. 20d) places the red algae as a group on the line of descent of the green algae and Spermatophyta. Previous phylogenetic speculation has indicated a more divergent relationship between the red algae and the green algal line than is apparent from the cytochrome c data (see for example, Scagel et al., 1965; Nichols, 1970; Christensen, 1964, in Appendices 2, 3 and 6 respectively). The evolutionary relationship between the red algae and the Basidiomycetes has been suggested on the basis of similarities of life cycles (Gwynne-Vaughan and Barnes, 1937), but a direct relationship is not supported by the present work.

The addition of the sequences of the fungi Humicola (Morgan et al., 1972) and Ustilago (Bitar et al., 1972) has made groupings within the fungi apparent. The minimum configuration found after several trials in this investigation, as shown in Fig. 19, does not agree with that of McLaughlin and Dayhoff (1973). These authors found a minimum score when Ustilago joined the main fungal branch between the Humicola/Neurospora branch and the Candida/Debaryomyces/Saccharomyces branch; this and the difference in several of the branch lengths may be attributable to the modifications in the method reported by the above authors. Within the fungi however, definite groupings are becoming apparent, although these do not correspond to the

current taxonomy of the species involved. The Basidiomycete, Ustilago, appears on its own branch, but members of the Ascomycetes and Deuteromycetes appear on both of the other main branches. The classification of the Deuteromycetes is however, extremely artificial and the class largely represents a convenient receptacle for grouping the imperfect stages of fungi from other classes together. As more sophisticated studies are made, it seems likely that the term 'fungi imperfecti' will become obsolete. However, Scagel et al. (1965) argue that as the fungi are classified on the basis of their sexual processes, then the term will remain for those types which have lost the ability to reproduce sexually. It is suggested that work on other aspects of the biology of the group, such as that involving protein sequence data, may assist in clarifying its relationships.

The most remote branch on the tree is that of the single bacterial representative of the Moneran kingdom, Rhodospirillum. This is in accordance with current biological thought regarding the prokaryotic line of descent. No blue-green algal cytochrome sequence is presently available, and the addition of such information to the tree should prove of interest.

Other investigations involving similar treatment of

sequence data from other proteins are currently in progress and a similar method has been applied to transfer RNA sequence data (see for example, Dayhoff, 1972). The results of these investigations may need to be considered together with the cytochrome c data, in order that a greater portion of the genome can be compared in making phylogenetic hypotheses.

In view of the problems associated with yields of cytochrome c from the algae and the apparent usefulness of protein sequence data in determining evolutionary relationships, it is probable that other lines of research involving other proteins may be more applicable to this group. In particular, proteins associated with the photosynthetic process may give more information as the yield of such proteins is generally higher than that of respiratory proteins from green tissues (see for example, for algae, Sugimura et al., 1968). Of particular value may be the studies currently being made on higher plants using the copper protein plastocyanin (Ramshaw et al., 1973). It is possible that work of this nature could be extended to include algal species.

Whilst at the present time it cannot be confidently asserted that the ancestral sequence method for examining the primary structure of proteins does give a true indication

of the phylogenetic relationships of organisms, the results to date suggest that further investigations on these lines are urgently needed.

BIBLIOGRAPHY

ALLSOPP, A. (1969). New Phytol. 68, 591.

AMBLER, R. P. (1963). Biochem. J. 89, 341.

AMBLER, R. P. (1971). FEBS Letters 18, 351.

AMBLER, R. P. and REES, M.W. (1959). Nature 184, 56.

AMBLER, R. P. and WYNN, M. (1973). Biochem. J. 131, 485.

ASADA, K. and TAKAHASHI, M. (1971). Plant & Cell Physiol. 12, 361.

BARTNICKI-GARCIA, S. (1970). In: Phytochemical Phylogeny (Ed. J. B. Harborne). Academic Press.

BARTSCH, R. G. and Kamen, M. D. (1968). J. Biol. Chem. 235, 825.

BECKMAN (1972). Instruction Manual for Model 890C Sequencer. Spinco Div. of Beckman Instruments Inc., Palo Alto, California 94304, U.S.A.

BENNETT, A. and BOGORAD, L. (1971). Biochemistry, N.Y. 10, 3625.

BITAR, K. G., VINOGRADOV, S. N., NOLAN, WEISS, L. J. and MARGOLIASH, E. (1972). Biochem. J. 129, 561.

BLIDING, C. (1963). Opera Bot. 8, 1.

BLOMBACK, B., BLOMBACK, M., EDMAN, P. and HESSEL, B. (1966). Biochim. Biophys. Acta 115, 371.

BOARDMAN, N. K. (1959). J. Chromatog. 2, 388.

BOARDMAN, N. K. and PARTRIDGE, S. M. (1953). Nature 171, 208.

BOARDMAN, N. K. and PARTRIDGE, S. M. (1954). J. Polymer Sci. 12, 281.

BOARDMAN, N. K. and PARTRIDGE, S. M. (1955). Biochem. J. 59, 543.

BOCK, R. M. and LING, N-S. (1954). Analyt. Chem. 26, 1543.

BODO, G. (1955). Nature 176, 829.

BONEY, A. D. (1966). A Biology of Marine Algae, Hutchinson.

BOUCK, G. B. (1969). J. Cell Biol. 40, 446.

BOULTER, D. (1970). Ann. Rev. Plant Physiol. 21, 91.

BOULTER, D. (1972). Sci. Prog. Oxf. 60, 217.

BOULTER, D. (1973). Pure & Applied Chem. 34, 539.

BOULTER, D., LAYCOCK, M. V., RAMSHAW, J. A. M. and THOMPSON, E. W. (1970). Taxon 19, 561.

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

BOULTER, D. and RAMSHAW, J. A. M. (1972). Phytochem. 11, 553.

BROWN, R. H. and BOULTER, D. (1973a). Biochem. J. 133, 251.

BROWN, R. H. and BOULTER, D. (1973b). Biochem. J. 131, 247.

BROWN, R. H. and BOULTER, D. (1973c). Biochem. J. 137, 93.

BROWN, R. H., RICHARDSON, M., BOULTER, D., RAMSHAW, J. A. M. and JEFFERIES, R. P. S. (1972). Biochem. J. 128, 971.

BROWN, R. H., RICHARDSON, M., SCOGIN, R. and BOULTER, D. (1973). Biochem. J. 131, 253.

BRUTON, C. J. and HARTLEY, B. S. (1970). J. Molec. Biol. 52, 165.

CANTOR, C. R. and JUKES, T. H. (1966). Proc. natn. Acad. Sci. U.S.A. 56, 177.

CARBON, J., BERG, P. and YANOFSKY, C. (1966). Cold Spring Harb. Symp. quant. Biol. 31, 417.

CHADEFAUD, M. (1960). In: Traite de Botanique Systematique, Vol. 1 (Ed. M. Chadefaud and L. Embarger). Paris.

CHAN, S.K. and MARGOLIASH, E. (1966). J. Biol. Chem. 241, 335.

CHAPMAN, V. J. (1962). The Algae. Macmillan & Co. Ltd.

CHESTER, K. S. (1937). Quart. Rev. Biol. 12, pp. 19, 165, 294.

CHRISTENSEN, T. (1962). Ref. cited in Christensen (1964) 'The Gross Classification of Algae'. In: Algae and Man (Ed. D. F. Jackson), Plenum Press, New York.

CHRISTENSEN, T. (1964). In: Algae and Man (Ed. D. F. Jackson), Plenum Press, New York, p.59.

CHRISTENSEN, T. (1966). Ref. cited by Cook, P. W. and Hoffman, L. R. (1971) (q.v.).

COHN, F. (1867). Arch. microskop. Anak. 3, 1.

COLBERT, E. H. (1969). The Evolution of the Vertebrates, 2nd Edition. New York Press.

COOK, P. W. and HOFFMAN, L. R. (1971). In: Selected Papers in Phycology (Eds. Rosowski, J. R. and Parker, B. C.). Dept. of Botany, Univ. of Nebraska, U.S.A.

COPELAND, H. F. (1938). Quart. Rev. Biol. 13, 383.

COPELAND, H. F. (1947). Amer. Naturalist 81, 340.

COPELAND, H. F. (1956). Classification of Lower Organisms. Pacific Books.

CRICK, F. H. C., BARNETT, L., BRENNER, S. and WATTS-TOBIN, R. J. (1961). Nature 192, 1227.

CRICK, F. H. C. (1966). Scientific American 215, 217.

CRONQUIST, A. (1968). The Evolution and Classification of the Flowering Plants. Nelson.

CROWSHAW, J. K., JESSUP, S. and RAMWELL, P. W. (1967). Biochem. J. 103, 79.

DAVIS, P. H. and HEYWOOD, V. H. (1963). Principles of Angiosperm Taxonomy. Oliver & Boyd, Edinburgh and London.

DAYHOFF, M. O. and ECK, R. V. (1966). Atlas of Protein Sequence and Structure, vol. 2. National Biomedical Research Foundation, Silver Spring, Maryland, U.S.A.

DAYHOFF, M. O. (1969). Atlas of Protein Sequence and Structure, vol. 4. National Biomedical Research Foundation, Silver Spring, Maryland, U.S.A.

DAYHOFF, M. O. (1972). Atlas of Protein Sequence and Structure, vol. 5. National Biomedical Research Foundation, Silver Spring, Maryland, U.S.A.

DAYHOFF, M. O. (1973). Atlas of Protein Sequence and Structure, vol. 5 Supplement. National Biomedical Research Foundation, Silver Spring, Maryland, U.S.A.

DELANGE, R. J., GLAZER, A. N. and SMITH, E. L. (1969). J. Biol. Chem. 244, 1385.

DICKERSON, R. E. (1971). J. Mol. Biol. 57, 1.

DICKERSON, R. E., TAKANO, T., EISENBERG, D., KALLAI, O. B., SAMPSON, L., COOPER, A. and MARGOLIASH, E. (1971). J. Biol. Chem. 246, 1511.

DIXON, M. (1953). Biochem. J. 54, 457.

DIXON, M., HILL, R. and KEILIN, D. (1931). Proc. Roy. Soc. B109, 29.

DIXON, P. S. (1970). Ann. N.Y. Acad. Sci. 175, 617.

DRUM, R. W. (1969). J. Phycol. 5, 21.

DUS, K., BARTSCH, R. G. and KAMEN, M. D. (1962). J. Biol. Chem. 237, 3083.

DUS, K., SLETTEN, K. and KAMEN, M.D. (1968). J. Biol. Chem. 243, 5507.

DUS, K. and KAMEN, M. D. (1963). Biochem. 338, 364.

EASLEY, C. W. (1965). Biochem. Biophys. Acta 107, 386.

EDMAN, P. (1956). Acta Chem. Scand. 10, 761.

ELMORE, D. T. and TOSELAND, P. A. (1956). J. Chem. Soc. 1956, 188.

FELLER, W. (1950). Probability Theory and Its Applications, 1st edn., vol. 1. Wiley, New York, p.71.

FITCH, W. M. (1966a). J. Molec. Biol. 16, 9.

FITCH, W. M. (1966b). J. Molec. Biol. 16, 17.

FITCH, W. M. (1969). Biochem. Genet. 3, 99.

FITCH, W. M. (1970). J. Molec. Biol. 49, 1.

FITCH, W. M. and MARGOLIASH, E. (1967a). Science 155, 279.

FITCH, W. M. and MARGOLIASH, E. (1967b). Biochem. Genet. 1, 65.

FITCH, W. M. and MARGOLIASH, E. (1969). Brookhaven Symp. Biol. 21, 217.

FITCH, W. M. and MARGOLIASH, E. (1970). Evolutionary Biology 4, 67.

FITCH, W. M. and MARKOWITZ, E. (1970). Biochem. Genet. 4, 579.

FLATMARK, T. (1964). Acta Chem. Scand. 18, 1517.

FLATMARK, T. (1966). Acta Chem. Scand. 20, 1487.

FLATMARK, T. and VESTERBERG, O. (1966). Acta Chem. Scand. 20, 1497.

FOTT, B. (1959). Algenkunde. Jena:Gustav Fischer (Publishers).

FOTT, B. (1965). Preslia 37, 117.

FREDERICK, J. F. (1968). Phytochem. 7, 931.

FRIDMAN, C., LIS, H., SHARON, N. and KATCHALSKI, E. (1968). Arch. Biochem. Biophys. 126, 299.

FRITSCH, F. E. (1945). The Structure and Reproduction of the Algae, vol. II. Cambridge University Press.

GIBBS, S. P. (1970). Anns. N.Y. Acad. Sci. 175, 454.

GIBBS, A. J. and MCINTYRE, G. A. (1970). Eur. J. Biochem. 16, 1.

GLAZER, A. N., COHEN-BAZISE, G. and STANIER, R. Y. (1971). Proc. Natl. Acad. Sci. U.S.A. 68, 3005.

GODDARD, D. R. (1944). Am. J. Botany 31, 270.

GRAY, W. R. (1967). In: Methods in Enzymology (Ed. C. H. W. Hirs), vol. 11. Academic Press, N.Y. and London, p.469.

GRAY, W. R. and HARTLEY, B. S. (1963a). Biochem. J. 89, 379.

GRAY, W. R. and HARTLEY, B. S. (1963b). Biochem. J. 89, 59P.

GRAY, W. R. and SMITH, J. F. (1970). Anal. Biochem. 33, 36.

GWYNNE-VAUGHAN, H. C. I. and BARNES, B. F. (1937). The Structure and Development of the Fungi, 2nd edn. Cambridge.

HAGIHARA, B., TAGAWA, K., MORIKAWA, I., SHIN, M. and OKUNUKI, K. (1958). Nature 181, 1590.

HAGIHARA, B., TAGAWA, K., MORIKAWA, I., SHIN, M. and OKUNUKI, K. (1959). J. Biochem. (Tokyo) 46, 321.

HAMILTON, P. B. (1962). Anns. N.Y. Acad. Sci. 102, 55.

HEILMANN, J., BARROLIER, J. and WATZKE, E. (1957). Hoppe-Seyler's Z. Physiol. Chem. 309, 219.

HELLER, J. and SMITH, E. L. (1966a). J. Biol. Chem. 241, 3158.

HELLER, J. and SMITH, E. L. (1966b). J. Biol. Chem. 241, 3165.

HILL, G. C., GUTTERIDGE, W. E. and MATHEWSON, N. W. (1971). Biochim. Biophys. Acta 243, 225.

HILL, R. and KEILIN, D. (1930). Proc. Roy. Soc. B. 107, 286.

HNILICA, L. S. (1967). Progr. Nucl. Acid Res. Mol. Biol. 7, 25.

HOLTON, R. and MYERS, J. (1967a). Biochim. Biophys. Acta 131, 362.

HOLTON, R. and MYERS, J. (1967b). Biochim. Biophys. Acta 131, 375.

HONIGBERG, B. M., BALAMUTH, W., BOVEE, E. C., CORLISS, J. O., GOJDICS, M., HALL, R. P., KUDO, R. R., LEVINE, N. D., LOEBLICH, A. R., WEISER, J. and WENRICH, D. H. (1964). J. Protozool. 11, 7.

IDA, S. and MORITA, Y. (1969). Agr. Biol. Chem. (Japan) 33, 10.

ILSE, D. and EDMAN, P. (1963). Australian J. Chem. 16, 411.

INGLIS, A. S., NICHOLLS, P. W. and ROXBURGH, C. M. (1971). Aust. J. biol. Sci. 24, 1247.

JACOBS, E. E. and SANADI, D. R. (1960). J. Biol. Chem. 235, 531.

JEPPSSON, J-O. and SJÖQUIST, J. (1967). Analyt. Biochem. 18, 264.

JUKES, T. H. (1969). Space Life Sci. 1, 469.

KAMEN, M. D., DUS, K. M., FLATMARK, T. and DE KLERK, H. (1971). In: Electron Coupled Energy Transfer in Biological Systems (Ed. King, Tsoo, E.), chapter 7. Dekker, N.Y.

KAMINSKY, L. S. and DAVISON, A. J. (1969). FEBS Letters 3, 338.

KASPER, C. B. (1970). In: Protein Sequence Determination (Ed. Needleman, S. B.). Chapman & Hall Ltd., London, Berlin, Heidelberg, New York.

KATOH, S. (1959). J. Biochem. (Tokyo) 46, 629.

KATOH, S. (1960). Nature 186, 138.

KEIL-DLOUHA, V., ZYLBER, N., IMHOFF, J. M., TONG, N. T. and KEIL, B. (1971). FEBS Letters 16, 291.

KEILIN, D. (1925). Proc. Roy. Soc. B. 98, 312.

KEILIN, D. (1927). C. R. Soc. Biol. Paris 96, 39.

KEILIN, D. (1930). Proc. Roy. Soc. B. 106, 458.

KEILIN, D. (1966). The History of Cell Respiration and Cytochrome. Cambridge University Press.

KEILIN, D. and HARTREE, E.F. (1939). Proc. Roy. Soc. B. 127, 167.

KELLY, J. and AMBLER, R. P. (1973). Biochem. Soc. Trans. 1, 164.

KLEIN, R. M. (1970). Ann. N.Y. Acad. Sci. 175, 623.

KLEIN, R. M. and CRONQUIST, A. (1967). Quart. Rev. Biol. 42, 105.

KOHNE, D. E. (1968). Taxonomic Applications of DNA Hybridization Techniques. In: Chemotaxonomy and Serotaxonomy (Hawkes, J. G., ed.), p. 117. Academic Press, London & N.Y.

KYLIN, H. (1943). Kgl. Fysiograf. Sällskap. Lund, Fösh 13, 169.

LAYCOCK, M. V. (1972). Can. J. Biochem. 50, 1311.

LEDERER, F., SIMON, A. M. and VERDIERE, J. (1972). Biochem. Biophys. Res. Commun. 47, 55.

LEDERER, F. (1972). Europ. J. Biochem. 31, 144.

LEEDALE, G. F. (1967). The Euglenoid Flagellates. Prentice Hall Inc., Eaglewood Cliffs, N.J.

LEMBERG, R. and BARRETT, J. (1973). Cytochromes. Academic Press, London and New York.

MACMUNN, C. A. (1884). J. Physiol. 5, xxiv.

MACMUNN, C. A. (1886). Phil. Trans. 177, 267.

MACMUNN, C. A. (1887). J. Physiol. 8, 51.

MARGOLIASH, E. (1954). Biochem. J. 56, 529.

MARGOLIASH, E. (1962). Brookhaven Symp. Biol. 15, 266.

MARGOLIASH, E. (1963). Proc. U.S. Natl. Acad. Sci. 50, 672.

MARGOLIASH, E. (1964). Can. J. Biochem. 42, 745.

MARGOLIASH, E. (1969a). Unpublished experiments cited by Margoliash, Fitch and Dickerson (1969).

MARGOLIASH, E. (1969b). Discussion to paper to Smithies, O. Brookhaven Symp. Biol. 21, 258.

MARGOLIASH, E. and FITCH, W. M. (1968). Ann. N.Y. Acad. Sci. 151, 359.

MARGOLIASH, E. and FROHWIRT, N. (1959). Biochem. J. 71, 570.

MARGOLIASH, E. and LUSTGARTEN, J. (1962). J. Biol. Chem. 237, 3397.

MARGOLIASH, E. and SCHEJTER, A. (1966). Adv. Prot. Chem. 21, 113.

MARGOLIASH, E. and SMITH, E. L. (1965). In: Evolving Genes and Proteins (Eds. Bryson, V. and Vogel, H. J.), p.221. Academic Press, New York.

MARGOLIASH, E. and WALASEK, O. F. (1967). In: Methods in Enzymology, vol. X (Eds. Estabrook, R. W. and Pullman, M.E). Academic Press, New York and London. p.339.

MARGOLIASH, E., SMITH, E. L., KREIL, G. and TUPPY, H. (1961). Nature 192, 1125.

MARGOLIASH, E., KIMMEL, J. R., HILL, R. L. and SCHMIDT, W. T. (1962). J. Biol. Chem. 237, 2148.

MARGOLIASH, E., NEEDLEMAN, S. B. and STEWART, J.W. (1963). Acta Chem. Scand, 17, S250.

MARGOLIASH, E., FITCH, W. M. and DICKERSON, R. E. (1969). Brookhaven Symp. Biol. 21, 259.

MARGOLIASH, E., BARLOW, G. H. and BYERS, V. (1971). Nature 228, 723.

MATSUBARA, H. and SMITH, E. L. (1962). J. Biol. Chem. 237, PC3575.

MCDOWALL, M. A. and SMITH, E. L. (1965). J. Biol. Chem. 240, 4635.

MCGUIRE, R. F. (1969). Phycol. 5, 220.

MCLAUGHLIN, P. J. and DAYHOFF, M. O. (1970). Science 168, 1469.

MCLAUGHLIN, P. J. and DAYHOFF, M. O. (1973). J. Molec. Evol. 2, 99.

MEYER, T. E. and CUSANOVICH, M. A. (1972). Biochim. Biophys. Acta 267, 383.

MITSUI, A. and TSUSHIMA, K. (1968). 'Osaka', pp. 459-466. Ref. quoted in Lemberg and Barrett (1973) (q.v.).

MOORE, S. and STEIN, W. H. (1963). In: Methods in Enzymology, vol. 6 (Eds. Colowick, S. P. and Kaplan, N. O.), p. 819. Academic Press, N.Y.

MORITA, Y. and IDA, S. (1968). Agricult. and Biol. Chem. (Japan) 32, 441.

MORITA, Y. and IDA, S. (1972). J. Mol. Biol. 71, 807.

MORITA, Y., YAGI, F., IDA, S., ASADA, K. and TAKAHASHI, M. (1973). FEBS Letters 31, 186.

MORGAN, W. T., HENSLEY, C. P. and RIEHM, J. P. (1972). J. Biol. Chem. 247, 6555.

NARITA, K. and SUGENO, K. (1968). In: Structure and Function of Cytochromes (Eds. Okunuki, K., Kamen, D. and Sekuzu, I.), p.304. Tokyo University Press.

NARITA, K., MURAKAMI, H. and TITANI, K. (1964). J. Biochem. (Tokyo) 56, 216.

NEEDLEMAN, S. B. and BLAIR, T. T. (1969). Proc. U.S. Natl. Acad. Sci. 63, 1227.

NEEDLEMAN, S. B. and MARGOLIASH, E. (1966). J. Biol. Chem. 241, 853.

NICHOLS, B. W. (1970). In: Phytochemical Phylogeny (Ed. Harborne, J. B.). Academic Press, London and N.Y.

NISHIMURA, M. (1959). J. Biochem. 46, 219.

NOLAN, C. and MARGOLIASH, E. (1966). J. Biol. Chem. 241, 1049.

NOLAN, C. and MARGOLIASH, E. (1968). Ann. Rev. Biochem. 38, 727.

OFFORD, R. E. (1966). Nature 211, 591.

PAIK, W.K. & KIM, S. (1967). Biochem. Biophys. Res. Comm. 27, 479.

PALEUS, S. (1960). Acta Chem. Scand. 14, 1743.

PALEUS, S. and NEILANDS, J. B. (1950). Acta Chem. Scand. 4, 1024.

PAPENFUSS, G. F. (1955). In: A Century of Progress in the Natural Sciences (Eds. Miller, R. C., Kessel, E. L. and Papenfuss, G. F.), p.115. San Francisco.

PARKER, B. C. (1970). Ann. N.Y. Acad. Sci. 175, 417.

PASCHER, A. (1931). Ref. cited in Prescott (1964) (q.v.).

PERCIVAL, E. (1968). Oceanogr. Mar. Biol. Ann. Rev. 6, 137.

PETTIGREW, G. W. (1972). FEBS Letters 22, 64.

PETTIGREW, G. W. (1973). Nature 241, 531.

PICKETT-HEAPS, J. D. and MARCHANT, H. J. (1972). Cytobios. 6, 255.

PISANO, J. J., BRONZERT, T. J. and BREWER, H. B. (1972). Analyt. Biochem. 45, 43.

POSTGATE, J. (1954). Biochem. J. 56, xi.

POSTGATE, J. (1956). J. Gen. Microbiol. 15, 186.

POTTS, J. T., BERGER, A., COOKE, J. and ANFINSEN, C. B. (1962). J. Biol. Chem. 237, 1851.

PRESCOTT, G. W. (1964). In: Algae and Man (Ed. Jackson, D. F.). Plenum Press, New York.

RAMSHAW, J. A. M., RICHARDSON, M. and BOULTER, D. (1971). Eur. J. Biochem. 23, 475.

RAMSHAW, J. A. M., RICHARDSON, D. L., MEATYARD, B. T., BROWN, R. H., RICHARDSON, M., THOMPSON, E. W. and BOULTER, D. (1972). New Phytol. 71, 773.

RAMSHAW, J. A. M., BROWN, R. H., SCAWEN, M. D. and BOULTER, D. (1973). Biochem. et Biophys. Acta 303, 269.

RHODES, R. G. and STOFAN, P. E. (1967). J. Phycol. 3, 87.

RICHARDSON, M., LAYCOCK, M. V., RAMSHAW, J. A. M., THOMPSON, E. W. and BOULTER, D. (1970). Phytochem. 9, 2271.

RICHARDSON, M., RAMSHAW, J. A. M. and BOULTER, D. (1971). Biochim. Biophys. Acta 251, 331.

RIFKIN, D. B., HIRSH, D. I., RIFKIN, M. R. and KONIGSBERG, W. (1966). Cold Spring Harbour Symp. 31, 715.

ROUND, F. E. (1965). The Biology of the Algae. Edward Arnold Ltd.

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 & Sons Ltd., London and Glasgow.

SCHMER, G. and KRIEL, G. (1969). Analyt. Biochem. 29, 186.

SCOTT, W. A. and MITCHELL, H. K. (1969). Biochemistry 8, 4282.

SHERMAN, F., TABER, H. and CAMPBELL, W. (1965). J. Molec. Biol. 13, 21.

SHERMAN, F., STEWART, J. W., PARKER, J. H., INHABER, E., SHIPMAN, N. A., PUTTERMAN, E. J., GARDINSKY, R. L. and MARCOLIASH, E. (1968). J. Biol. Chem. 243, 5446.

SHIRASAKA, M., NAKAYAMA, N., ENDO, A., HANEISHI, T. and OKAZAKI, H. (1968). J. Biochem. 63, 417.

SKUJA, H. (1938). Ref. cited in Fritsch (1945) (q.v.).

SLINIMSKI, P. P., ACHER, R., PERE, G., SELS, A. and SOMLO, M. (1965). In: Mechanismes de Regulation des Activites Cellulaires chez les Microorganismes, p.435. Du Centre National de la Recherche Scientifique, Paris.

SMITH, I. (1958). Chromatographic and Electrophoretic Techniques, vol. 1, p.7. Heineman, London.

SMITH, E. L. (1968). Harvey Lectures 62, 231.

SMITH, E. L. and MARCKER, K. (1970). Nature 226, 607.

SMITH, G. M. (1950). Freshwater Algae of the United States, 2nd edn. McGraw-Hill Book Co., N.Y.

SOKAL, R. R. and SNEATH, P. H. A. (1963). Principles of Numerical Taxonomy. Freeman & Co., San Francisco.

SPRINGER, R. and ABDELMESSIH, N. (1969). Botanica Mar. 12, 200.

STANIER, R. Y. (1974). In: Evolution in the Microbial World, (Eds. Carlile, M. J. and Skehel, J. J.). Cambridge University Press.

STEERS, E. Jr., CRAVEN, G. R., ANFINSEN, C. B. and BETHUNE, J. L. (1965). J. Biol. Chem. 240, 2478.

STELLWAGEN, E. (1968). Biochemistry 7, 2893.

STEVENS, F. C., GLAZER, A. N. and SMITH, E. L. (1967). J. Biol. Chem. 242, 2764.

STEWART, J. W. and MARGOLIASH, E. (1965). Can. J. Biochem. 43, 1187.

SUGENO, K., MATSUBARA, H. (1968). Biochem. Biophys. Res. Commun. 32, 951.

SUGIMURA, Y., TODA, F., MURATA, T. and YAKUSHIJI, E. (1968). In: Structure and Function of Cytochromes (Eds. Okunuki, K., Kamen, M. D. and Sekuzu, I.), p. 452. University of Tokyo Press & Univ. Park Press, Baltimore.

TAKHTAJAN, A. (1969). Flowering Plants - Origin and Dispersal. Oliver and Boyd Ltd., Edinburgh.

TECHNICON (1963). Instructions for Assembly and Operation of Amino Acid Analyzer. Chertsey, Surrey, Technicon Instruments Ltd.

TECHNICON (1968). 6th Colloquium in Amino Acid Analysis. Monograph 3, p.162. Geneva: Technicon International Division S.A.

THOMPSON, E. W., RICHARDSON, M. and BOULTER, D. (1970). Biochem. J. 121, 439.

THOMPSON, E. W., NOTTON, B. A., RICHARDSON, M., and BOULTER, D. (1971a). Biochem. J. 124, 787.

THOMPSON, E. W., RICHARDSON, M. and BOULTER, D. (1971b). Biochem. J. 124, 779.

THOMPSON, E. W., RICHARDSON, M. and BOULTER, D. (1971c). Biochem. J. 124, 783.

TILDEN, J. E. (1935). The Algae and Their Relations. Univ. of Minnesota Press.

TURRILL, W. B. (1938). Biol. Rev. 13, 342.

VALENTINE, J. and ROBERTS, S. J. (1971). Unpublished experiments.

VAN DEN HOEK, C. (1964). In: Algae and Man (Ed. Jackson, J. F.), p.31. Plenum Press.

VAN GELDER, B. F. and SLATER, E. C. (1962). Biochim. Biophys. Acta 58, 593.

VAN OYE, P. (1961). Hydrobiol. 18, 293.

VAUGHAN, J. G. (1968a). Science Progress (Oxford) 56, 205.

VAUGHAN, J. E. (1968b). In: Chemotaxonomy and Serotaxonomy (Ed. Hawkes, J. G.), p. 93. Systematics Association Special Volume No. 2.

WALASEK, O. F. and MARGOLIASH, E. (1969). Unpublished experiments cited by Margoliash et al. (1969) (q.v.).

WASSERMAN, A. R., GAVER, J. C. and BURRIS, R. H. (1963). Phytochem. 2, 7.

WHITTAKER, R. H. (1969). Science 163, 150.

WOODS, K. R. and WANG, K. T. (1967). Biochim. Biophys. Acta 133, 369.

YAKUSHIJI, E. (1935). Acta Phytochimica VIII, p.325.

YAKUSHIJI, E., SUGIMURA, A., SEKUZU, I., MORIKAWA, I. and OKUNUKI, K. (1960). Nature 185, 105.

YAMANAKA, T. (1966). Ann. Report Biol. Works Fac. Sci. Osaka University 14, 47.

YAMANAKA, T. (1967). Nature 213, 1183.

YAMANAKA, T. and KAMEN, M. D. (1965). Biochim. Biophys. Acta 96, 328.

YAMANAKA, T. and OKUNUKI, K. (1963). Biochim. Biophys. Acta 67, 379.

YAMANAKA, T., MIZUSHIMA, H., NOZAKI, M., HORIO, T. and OKUNUKI, K. (1959). J. Biochem. 46, 121.

YAMANAKA, T., MIZUSHIMA, H., KATANO, H. and OKUNUKI, K. (1964a). Biochim. Biophys. Acta 81, 223.

YAMANAKA, T., MIZUSHIMA, H., KATANO, H. and OKUNUKI, K. (1964b). Biochim. Biophys. Acta 85, 11.

YAMANAKA, T., NAGATA, Y. and OKUNUKI, K. (1968). J. Biochem. 63, 753.

YAMANAKA, T., TAKENAMI, S. and OKUNUKI, K. (1969). Biochim. Biophys. Acta 180, 193.

YOSHIDA, T. (1966). Ann. Reports Biol. Works Fac. Sci. Osaka Univ. 14, 1.

YOUNG, J. Z. (1962). The Life of the Vertebrates, 2nd edn. Oxford University Press.

ZUCKERKANDL, E. and PAULING, L. (1962). In: Horizons in Biochemistry (Eds. Kasha, M. and Pullman, B.), p.189. Academic Press, New York.

ZUCKERKANDL, E. and PAULING, L. (1965). J. Theoret. Biol. 8, 357.

A P P E N D I C E S

1. The major characteristics of groups of Algae (after Scagel *et al.*, 1965).

APPENDIX I.

MAJOR CHARACTERISTICS OF ALGAL CLASSES

DIVISION AND CLASS	PIGMENTS	STORAGE PRODUCTS	FLAGELLATION	CELL WALL
CYANOPHYTA Cyanophyceae	chlorophyll a β-carotene flavacin lutein myxoxanthin myxoxanthophyll zeaxanthin allophycocyanin c-phycocyanin c-phycoerythrin	cyanophyte starch (=amylopectin; α, 1-6, 1-4 linkage) cyanophycin and other proteins no sterols reported	none	cellulose pectin
PYRROPHYTA	chlorophyll a, c diadinoxanthin dinoxanthin	starch (=amylose + amylopectin; α, 1-4, 1-6 linkage) fats, oils		cellulose pectin
Dinophyceae	β-carotene peridinin		2 lateral (or apical) insertion, 1 around girdle (flattened or ?tinsel), 1 trailing (whiplash)	
Cryptophyceae	α⁻, ε-carotene ?zeaxanthin x-c-phycocyanin* x-phycoerythrin+		2 lateral insertion, ?tinsel	
CHRYSOPHYTA	chlorophyll a, c+ β-carotene diadinoxanthin diatoxanthin+ dinoxanthin+ fucoxanthin	chrysolaminarin (=leucosin; β, 1-3, 1-6 linkage) fucosterol and others+		cellulose pectin

MAJOR CHARACTERISTICS OF ALGAL CLASSES (Continued)

DIVISION AND CLASS	PIGMENTS	STORAGE PRODUCTS	FLAGELLATION	CELL WALL
Chrysophyceae	α-carotene lutein		1, 2 apical (or sub-apical) insertion; 1 tinsel (or ?whiplash) 2 equal, whiplash 2 unequal, subequal, longer tinsel	May be absent
Bacillariophyceae	α^-, ϵ-carotene		reproductive cell (male) only, 1 tinsel ?2	silicon
PHAEOPHYTA				
Phaeophyceae	chlorophyll a, c α^-, β-carotene flavoxanthin fucoxanthin lutein violaxanthin	laminarin (β, 1-3, 1-6 linkage) mannitol fucosterol and others	reproductive cells only 2 lateral insertion, forward tinsel, backward whiplash 1 forward tinsel	cellulose pectin alginic acid (algin) fucoidin
RHODOPHYTA				
Rhodophyceae	chlorophyll a, ?d α^-, β-carotene lutein ?taraxanthin zeaxanthin allophycocyanin b-, c-, r-phycoerythrin b-, c-, r-phycocyanin	floridean starch (=amylopectin; α, 1-4, 1-6 linkage, + α, 1-3) floridoside mannoglycerate cholesterol‡ fucosterol, sitosterol	None	cellulose pectin various mucilages (agar, carrageenin xylan) mannan (some with calcium or magnesium carbonate)
XANTHOPHYTA	chlorophyll a β-carotene xanthophylls		2, apical insertion	
Xanthophyceae	chlorophyll ?e	chrysolaminarin (= leucosin; β, 1-3, 1-6 linkage sitosterol	unequal; long tinsel; short, whiplash	cellulose pectin

MAJOR CHARACTERISTICS OF ALGAL CLASSES (Continued)

DIVISION AND CLASS	PIGMENTS	STORAGE PRODUCTS	FLAGELLATION	CELL WALL
Chloromonadophyceae		oil	equal or unequal; forward, ?tinsel; trailing, ?whiplash	generally lacking
EUGLENOPHYTA				
Euglenophyceae	chlorophyll *a*, *b* β-carotene antheraxanthin astaxanthin neoxanthin	paramylon (= paramylum; β, 1-3 linkage) ergosterol	1 (2 or 3) apical insertion, tinsel, single row of mastigonemes	generally lacking ? pectin
CHLOROPHYTA	chlorophyll *a*, *b* β-carotene	starch (= amylose + amylopectin; α, 1-4, 1-6 linkage)	2, apical insertion, equal, whiplash	cellulose pectin (occasionally with calcium carbonate)
Chlorophyceae	α-carotene astaxanthin leutein neoxanthin siphonein§ siphonoxanthin§ violaxanthin zeaxanthin	ergosterol fucosterol, sitosterol, and others	also 4, 8, or a ring of flagella	mannan mucilages (xylan) chitin
Charophyceae	γ-carotene lycopene		male reproductive cell only	

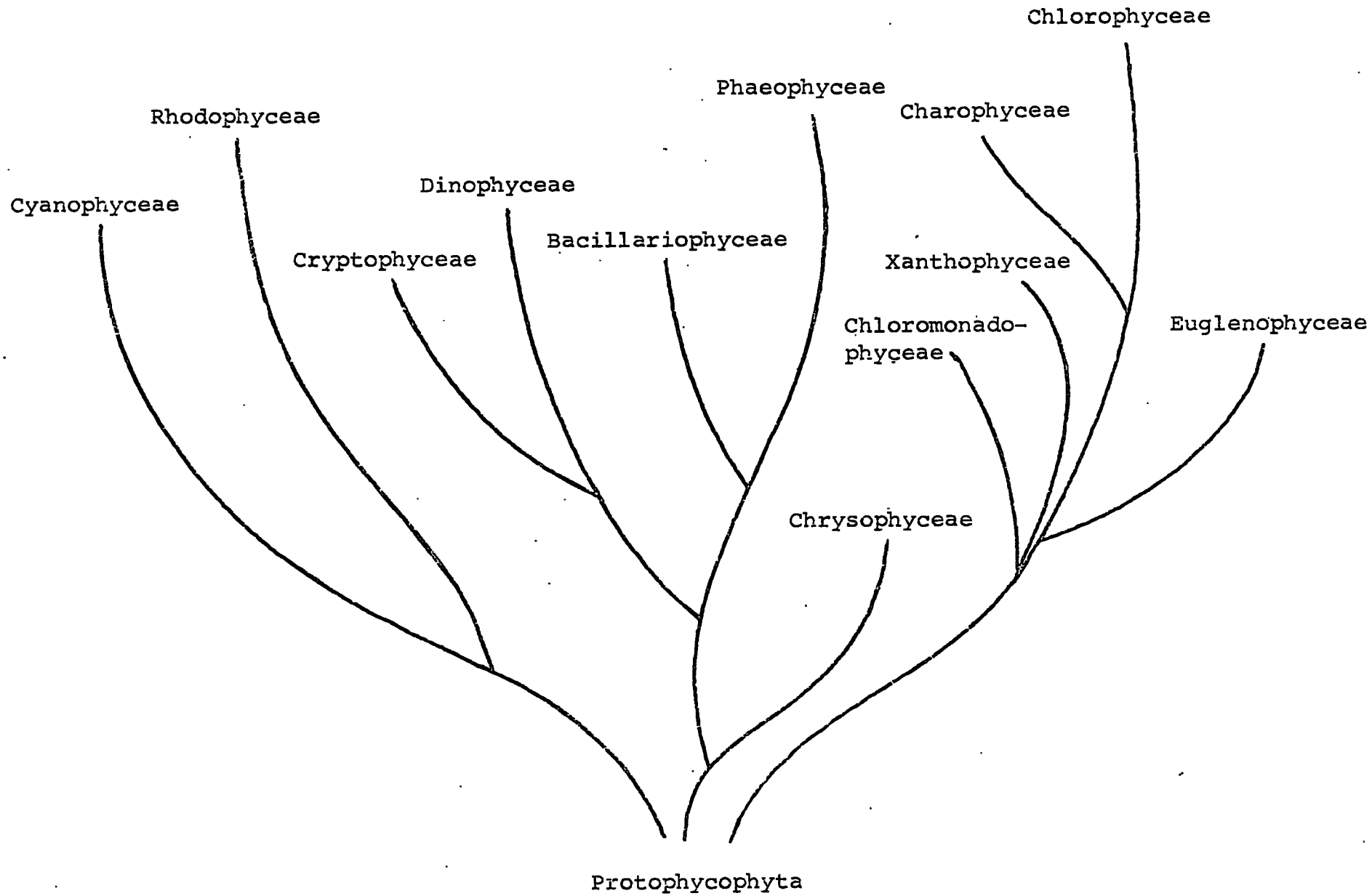
*Phycobilins in Cryptophyceae studied are for the most part different from those occurring in the Rhodophyta and Cyanophyta.

+Too few algae examined to make accurate generalization.

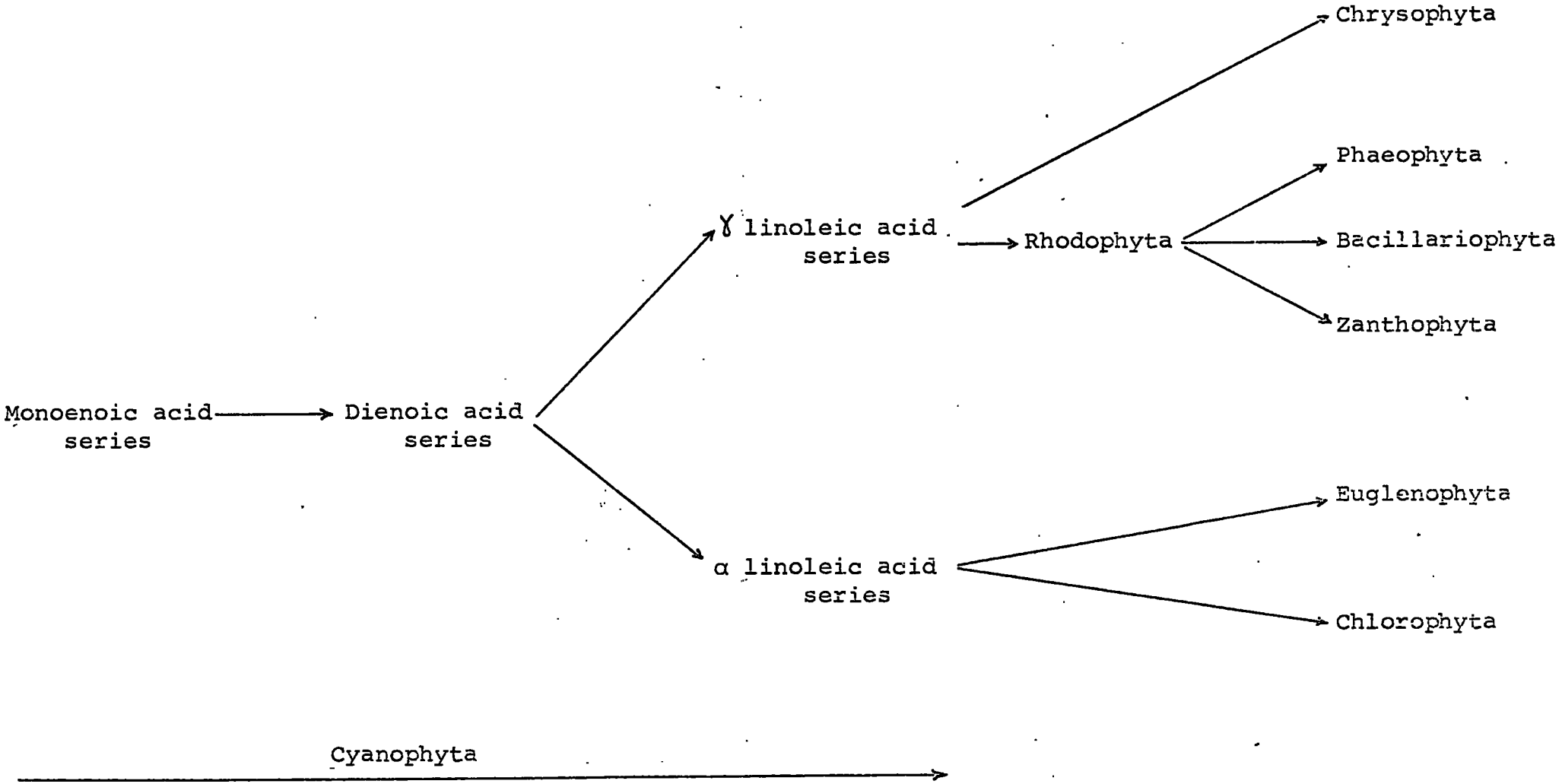
‡Almost all Japanese forms contain cholesterol, but the British forms contain sitosterol or fucosterol instead.

§In siphonous forms only.

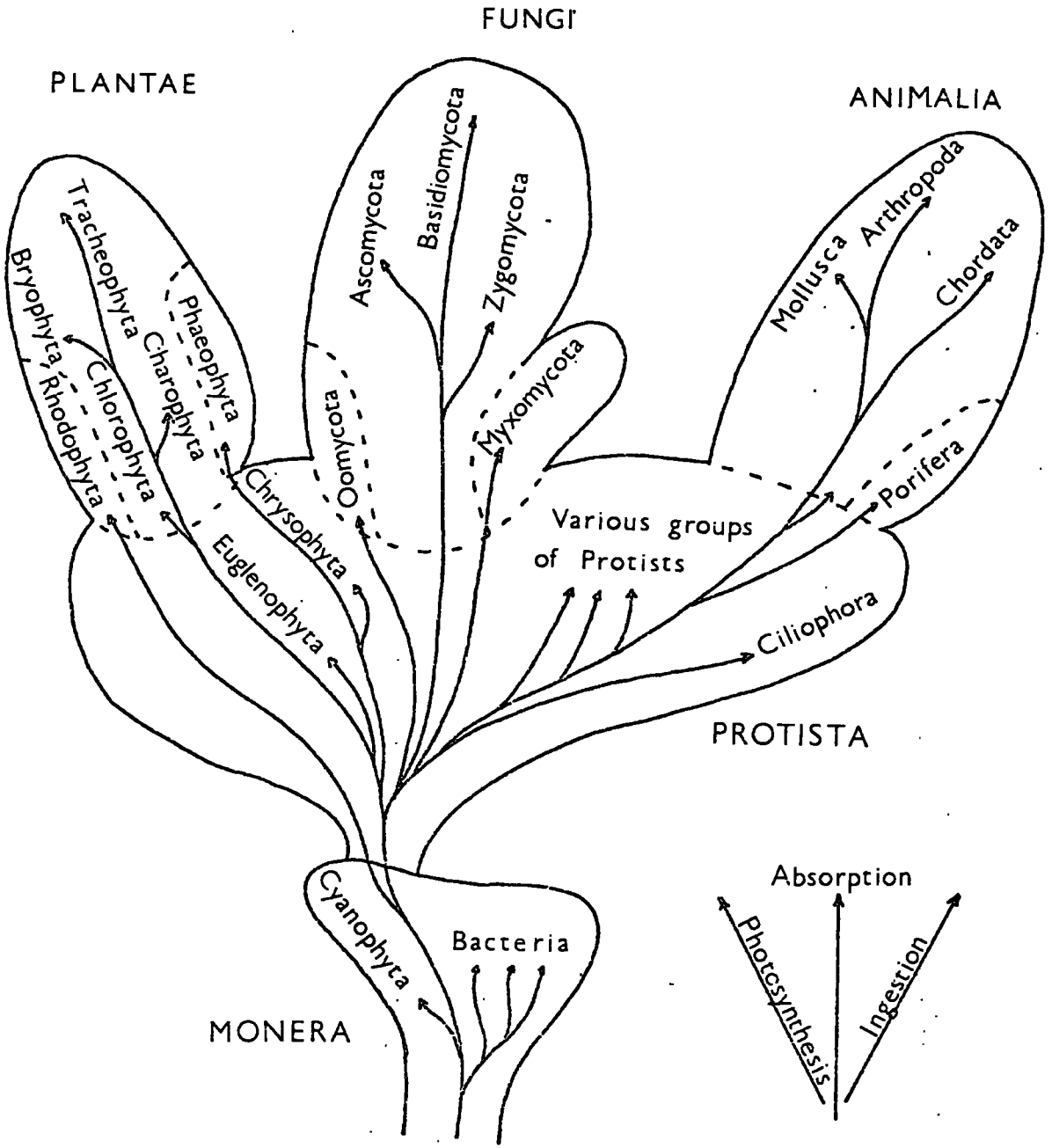
2. The possible interrelationships and phylogenetic arrangement of algal classes (Scagel et al., 1965).


Chlorophyceae
Phaeophyceae
Charophyceae
Rhodophyceae
Dinophyceae
Cyanophyceae
Bacillariophyceae
Cryptophyceae
Xanthophyceae
Chloromonado-
phyceae
Euglenophyceae
Chrysophyceae
Protophycophyta

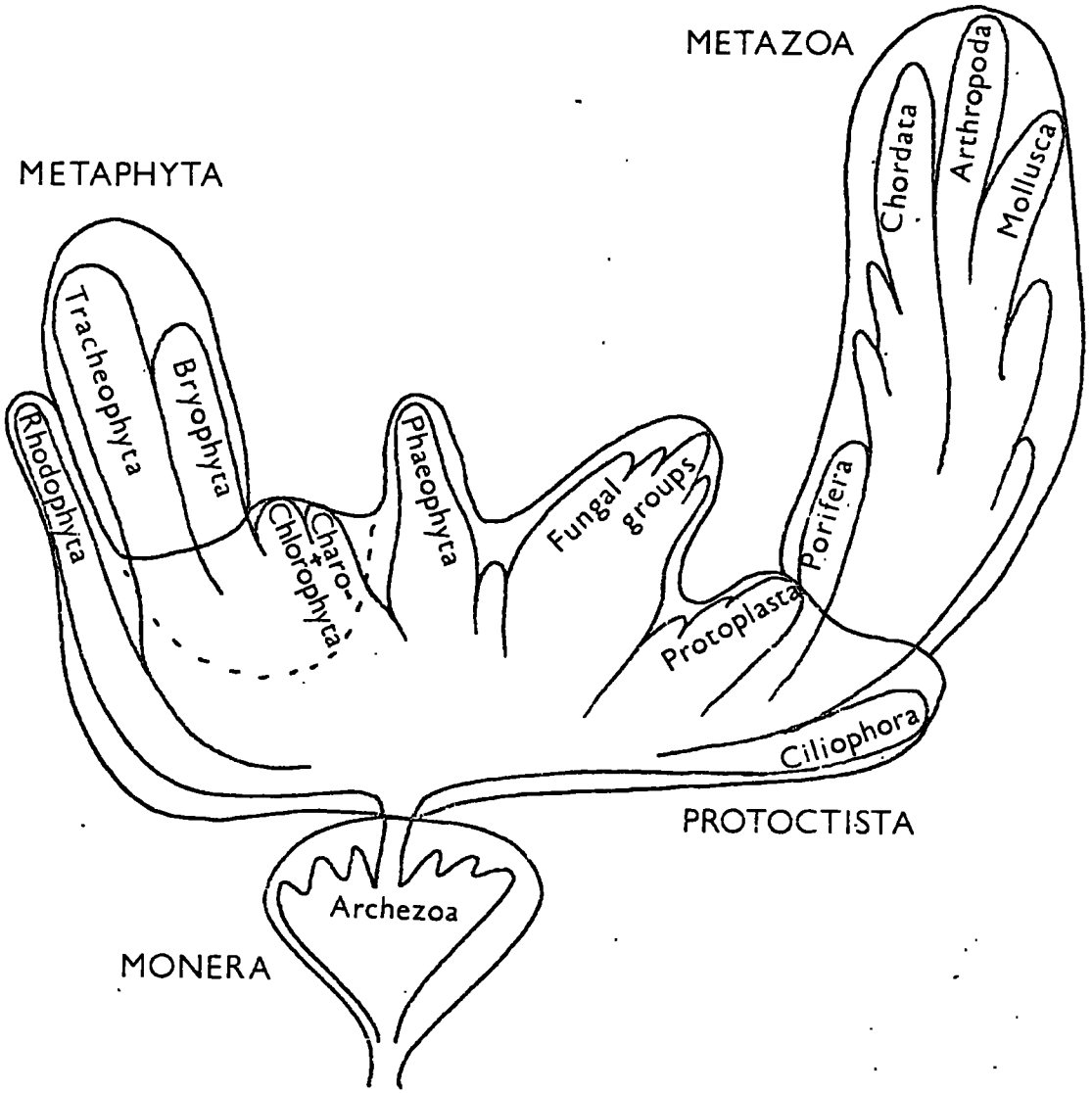
3. A phylogenetic scheme for algal groups based on certain aspects of lipid metabolism (Nichols, 1970).


Chrysophyta
Phaeophyta
γ linoleic acid series
Rhodophyta
Bacillariophyta
Zanthophyta
Monoenoic acid series
Dienoic acid series
Euglenophyta
α linoleic acid series
Chlorophyta
Cyanophyta

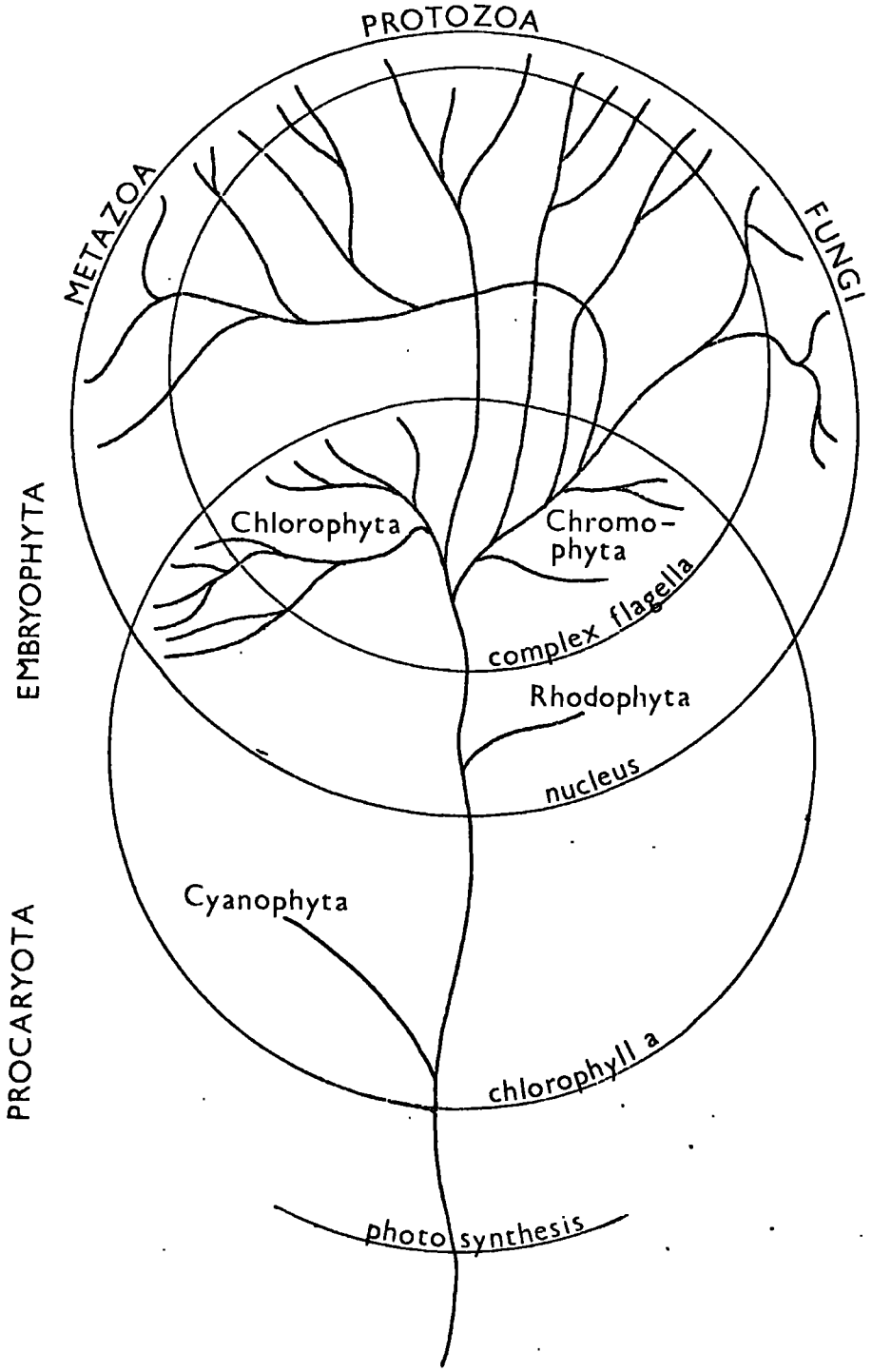
4. The Whittaker five-Kingdom system based on three levels of organization - prokaryotic, eukaryotic unicellular and eukaryotic multicellular and multinucleate. On each level there is divergence in relation to three principle modes of nutrition - photosynthetic, absorptive and ingestive. For clarity only selected phyla have been included (see Whittaker, 1969).


FUNGI
PLANTAE
ANIMALIA
Tracheophyta
Bryophyta
Rhodophyta
Chlorophyta
Phaeophyta
Charophyta
Ascomycota
Basidiomycota
Zygomycota
Oomycota
Myxomycota
Chrysophyta
Euglenophyta
Mollusca
Arthropoda
Chordata
Porifera
Various groups of Protists
Ciliophora
PROTISTA
Cyanophyta
Bacteria
MONERA
Absorption
Photosynthesis
Ingestion

5. The Copeland four-Kingdom system, with the relationships of selected phyla to kingdoms and levels of organisation. Alternative treatments of the Chlorophyta and Charophyta are indicated. These are included in the Metaphyta by Copeland, but in the Protista by other authors (see Copeland, 1956).


METAZOA
Chordata
Arthropoda
Mollusca
METAPHYTA
Tracheophyta
Bryophyta
Rhodophyta
Charo-
Chlorophyta
Phaeophyta
Fungal
groups
Porifera
Protoplasta
Ciliophora
PROTOCTISTA
Archezoa
MONERA

6. The generalized phylogenetic scheme of Christensen (see Christensen, 1962; 1964; 1966). Relationships are drawn with regard to cell type, pigmentation and type of flagellum.


PROTOZOA
METAZOA
FUNGI
EMBRYOPHYTA
Chlorophyta
Chromo-
phyta
complex flagella
Rhodophyta
nucleus
PROCARYOTA
Cyanophyta
chlorophyll a
photosynthesis

7. The single letter code for the nomenclature of amino acids, as given in the tentative rules of the IUPAC - IUB Commission on Biochemical Nomenclature.

J represents ϵ-N-trimethyllysine.

APPENDIX 7.

Table of Abbreviations and Symbols for Amino Acids

Amino Acid.	Abbreviations.	Code Letter.
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asp	D
Aspartic acid	Asn	N
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
ϵ-N-trimethyl-lysine	Me_3Lys	J
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

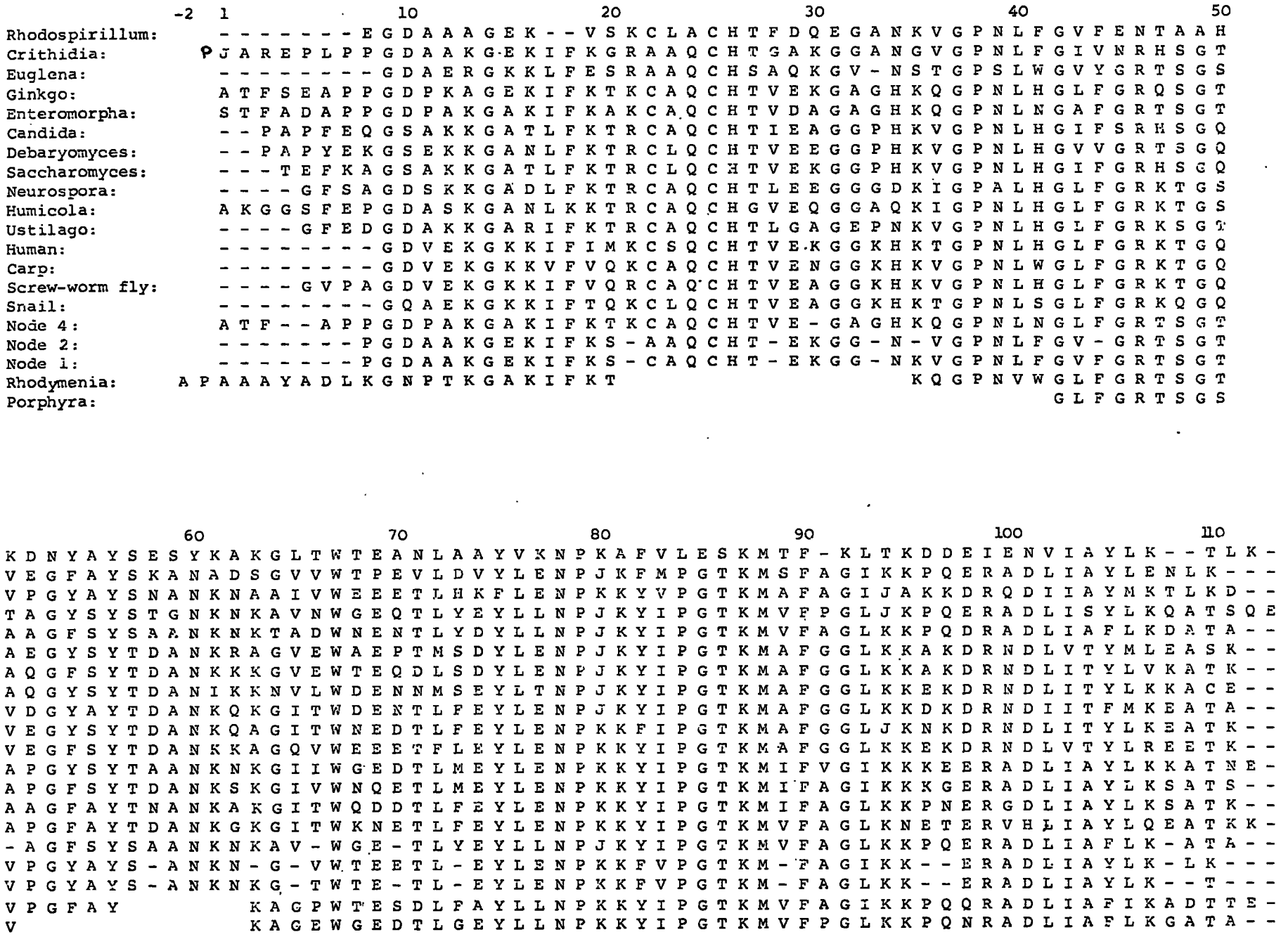
8. Sequence alignment of cytochromes *c*. The sequences are aligned relative to the cysteinyl residues binding the haem group.

Sequence information is as cited in Dayhoff (1972; 1973), with the addition of *Euglena* (Pettigrew, 1973).

The sequences of certain ancestral nodes computed during the compilation of the tree shown in Fig. 19, are also given.

The sequence of *Rhodospirillum rubrum* cytochrome c_2 has been aligned by assigning deletions or insertions so as to obtain maximum homology with other cytochromes *c*. Insertions were not considered in constructing phylogenetic trees.

(See Appendix 7 for single-letter amino acid code).


-2 1 10 20 30 40 50
Rhodospirillum: - - - - - - - E G D A A A G E K - - V S K C L A C H T F D Q E G A N K V G P N L F G V F E N T A A H
Crithidia: P J A R E P L P P G D A A K G E K I F K G R A A Q C H T G A K G G A N G V G P N L F G I V N R H S G T
Euglena: - - - - - - - - - G D A E R G K K L F E S R A A Q C H S A Q K G V - N S T G P S L W G V Y G R T S G S
Ginkgo: A T F S E A P P G D P K A G E K I F K T K C A Q C H T V E K G A G H K Q G P N L H G L F G R Q S G T
Enteromorpha: S T F A D A P P G D P A K G A K I F K A K C A Q C H T V D A G A G H K Q G P N L N G A F G R T S G T
Candida: - - P A P F E Q G S A K K G A T L F K T R C A Q C H T I E A G G P H K V G P N L H G I F S R H S G Q
Debaryomyces: - - P A P Y E K G S E K K G A N L F K T R C L Q C H T V E E G G P H K V G P N L H G V V G R T S G Q
Saccharomyces: - - - T E F K A G S A K K G A T L F K T R C L Q C H T V E K G G P H K V G P N L H G I F G R H S G Q
Neurospora: - - - - G F S A G D S K K G A D L F K T R C A Q C H T L E E G G G D K I G P A L H G L F G R K T G S
Humicola: A K G G S F E P G D A S K G A N L K K T R C A Q C H G V E Q G G A Q K I G P N L H G L F G R K T G S
Ustilago: - - - - G F E D G D A K K G A R I F K T R C A Q C H T L G A G E P N K V G P N L H G L F G R K S G S
Human: - - - - - - - - - G D V E K G K K I F I M K C S Q C H T V E K G G K H K T G P N L H G L F G R K T G Q
Carp: - - - - - - - - - G D V E K G K K V F V Q K C A Q C H T V E N G G K H K V G P N L W G L F G R K T G Q
Screw-worm fly: - - - - G V P A G D V E K G K K I F V Q R C A Q C H T V E A G G K H K V G P N L H G L F G R K T G Q
Snail: - - - - - - - - - G Q A E K G K K I F T Q K C L Q C H T V E A G G K H K T G P N L S G L F G R K Q G Q
Node 4: A T F - - A P P G D P A K G A K I F K T K C A Q C H T V E - G A G H K Q G P N L N G L F G R T S G T
Node 2: - - - - - - - - P G D A A K G E K I F K S - A A Q C H T - E K G G - N - V G P N L F G V - G R T S G T
Node 1: - - - - - - - - P G D A A K G E K I F K S - C A Q C H T - E K G G - N K V G P N L F G V F G R T S G T
Rhodymenia: A P A A A Y A D L K G N P T K G A K I F K T K Q G P N V W G L F G R T S G T
Porphyra: G L F G R T S G S
60 70 80 90 100 110
K D N Y A Y S E S Y K A K G L T W T E A N L A A Y V K N P K A F V L E S K M T F - K L T K D D E I E N V I A Y L K - - T L K -
V E G F A Y S K A N A D S G V V W T P E V L D V Y L E N P J K F M P G T K M S F A G I K K P Q E R A D L I A Y L E N L K - - -
V P G Y A Y S N A N K N A A I V W E E E T L H K F L E N P K K Y V P G T K M A F A G I J A K K D R Q D I I A Y M K T L K D - -
T A G Y S Y S T G N K N K A V N W G E Q T L Y E Y L L N P J K Y I P G T K M V F P G L J K P Q E R A D L I S Y L K Q A T S Q E
A A G F S Y S A A N K N K T A D W N E N T L Y D Y L L N P J K Y I P G T K M V F A G L K K P Q D R A D L I A F L K D A T A - -
A E G Y S Y T D A N K R A G V E W A E P T M S D Y L E N P J K Y I P G T K M A F G G L K K A K D R N D L V T Y M L E A S K - -
A Q G F S Y T D A N K K K G V E W T E Q D L S D Y L E N P J K Y I P G T K M A F G G L K K A K D R N D L I T Y L V K A T K - -
A Q G Y S Y T D A N I K K N V L W D E N N M S E Y L T N P J K Y I P G T K M A F G G L K K E K D R N D L I T Y L K K A C E - -
V D G Y A Y T D A N K Q K G I T W D E N T L F E Y L E N P J K Y I P G T K M A F G G L K K D K D R N D I I T F M K E A T A - -
V E G Y S Y T D A N K Q A G I T W N E D T L F E Y L E N P K K F I P G T K M A F G G L J K N K D R N D L I T Y L K E A T K - -
V E G F S Y T D A N K K A G Q V W E E E T F L E Y L E N P K K Y I P G T K M A F G G L K K E K D R N D L V T Y L R E E T K - -
A P G Y S Y T A A N K N K G I I W G E D T L M E Y L E N P K K Y I P G T K M I F V G I K K K E E R A D L I A Y L K K A T N E -
A P G F S Y T D A N K S K G I V W N Q E T L M E Y L E N P K K Y I P G T K M I F A G I K K K G E R A D L I A Y L K S A T S - -
A A G F A Y T N A N K A K G I T W Q D D T L F E Y L E N P K K Y I P G T K M I F A G L K K P N E R G D L I A Y L K S A T K - -
A P G F A Y T D A N K G K G I T W K N E T L F E Y L E N P K K Y I P G T K M V F A G L K N E T E R V H L I A Y L Q E A T K K -
- A G F S Y S A A N K N K A V - W G E - T L Y E Y L L N P J K Y I P G T K M V F A G L K K P Q E R A D L I A F L K - A T A - -
V P G Y A Y S - A N K N - G - V W T E E T L - E Y L E N P K K F V P G T K M - F A G I K K - - E R A D L I A Y L K - L K - - -
V P G Y A Y S - A N K N K G - T W T E - T L - E Y L E N P K K F V P G T K M - F A G L K K - - E R A D L I A Y L K - - T - - -
V P G F A Y K A G P W T E S D L F A Y L L N P K K Y I P G T K M V F A G I K K P Q Q R A D L I A F I K A D T T E -
V K A G E W G E D T L G E Y L L N P K K Y I P G T K M V F P G L K K P Q N R A D L I A F L K G A T A - -

9. Matrix of amino acid differences.

	Rhodospirillum:	Crithidia:	Euglena:	Pumpkin:	Rape:	Mungbean:	Elder:	Niger:	Sunflower:	Cotton:	Abutilon:	Castor:	Sesame:	Tomato:	Buckwheat:	Spinach:	Ginkgo:	Enteromorpha:	Candida:	Debaryomyces:	Saccharomyces:	Neurospora:	Humicola:	Ustilago:	Horse:	Man:	Turtle:	Chicken:	Bullfrog:	Bonito:	Tuna:	Carp:	Dogfish:	Lamprey:	Fruitfly:	Screw-worm fly:	Tobacco Moth	Samia:	Snail:	Rhodymenia:	Porphyra:
Rhodospirillum:	0																																								
Crithidia:	71	0																																							
Euglena:	70	57	0																																						
Pumpkin:	72	54	58	0																																					
Rape:	71	54	59	4	0																																				
Mungbean:	73	54	59	6	5	0																																			
Elder:	72	55	60	8	7	7	0																																		
Niger:	69	52	59	13	11	13	11	0																																	
Sunflower:	73	54	58	14	14	12	11	8	0																																
Cotton:	71	53	59	8	7	7	7	10	12	0																															
Abutilon:	72	53	58	8	9	9	8	10	12	3	0																														
Castor:	73	54	60	8	8	10	8	11	12	5	6	0																													
Sesame:	72	53	60	7	8	7	11	13	12	7	7	7	0																												
Tomato:	72	53	61	9	7	9	8	10	10	6	5	5	5	0																											
Buckwheat:	71	53	61	15	13	13	41	13	15	12	12	12	12	11	0																										
Spinach:	73	57	62	17	16	16	17	20	18	16	18	18	18	19	11	0																									
Ginkgo:	72	56	64	20	18	20	18	18	20	19	18	16	19	14	14	22	0																								
Enteromorpha:	72	52	62	26	24	27	25	26	24	25	27	25	26	25	27	34	31	0																							
Candida:	78	57	61	51	50	52	53	52	52	51	50	51	52	51	54	54	55	51	0																						
Debaryomyces:	72	56	62	50	48	51	50	49	50	51	50	49	51	50	51	51	51	46	24	0																					
Saccharomyces:	73	59	62	49	48	50	50	46	48	48	47	47	49	48	48	49	49	50	27	26	0																				
Neurospora:	70	63	53	50	49	51	54	50	50	51	50	50	51	50	51	53	52	49	38	39	36	0																			
Humicola:	71	58	56	48	48	49	51	50	49	48	47	48	49	48	47	48	49	51	37	39	41	29	0																		
Ustilago:	73	54	54	50	50	50	52	51	51	50	50	52	52	52	52	52	53	50	31	36	37	34	33	0																	
Horse:	62	56	47	49	49	51	52	48	48	51	49	48	47	48	46	50	47	46	51	45	46	42	44	44	0																
Man:	64	59	49	45	46	48	48	45	44	47	45	44	42	44	42	45	44	46	51	46	45	44	44	47	12	0															
Turtle:	62	54	48	48	47	49	50	46	46	48	47	45	45	45	42	47	45	43	52	47	49	45	45	45	11	15	0														
Chicken:	62	57	50	50	49	51	51	48	48	50	49	48	48	47	42	48	45	45	50	45	46	43	42	44	11	13	8	0													
Bullfrog:	63	56	49	52	50	52	53	49	49	51	51	48	49	48	45	50	47	47	51	47	47	45	46	46	14	18	10	11	0												
Bonito:	62	56	49	51	50	52	53	48	48	51	50	48	48	48	48	50	49	45	46	46	45	42	42	42	18	21	17	17	15	0											
Tuna:	63	56	50	52	51	53	54	50	50	53	52	50	50	50	48	50	50	47	47	47	47	44	42	43	19	21	18	17	15	2	0										
Carp:	64	54	48	50	50	52	53	49	49	52	50	49	48	49	48	50	49	45	50	45	47	44	44	43	14	19	14	16	14	8	8	0									
Dogfish:	64	59	51	53	52	53	54	50	50	53	52	51	51	51	50	53	53	48	52	46	48	45	47	47	17	24	19	19	20	20	20	15	0								
Lamprey:	66	57	51	54	54	54	54	51	51	54	52	53	52	53	51	54	53	50	55	48	49	48	48	47	16	20	19	18	21	19	19	12	17	0							
Fruitfly:	65	54	54	52	50	52	53	48	48	52	52	49	50	49	46	49	49	43	47	42	45	37	40	42	24	29	24	25	22	26	25	23	26	29	0						
Screw-worm fly:	64	53	55	50	48	50	51	46	46	50	50	47	48	47	44	47	47	43	47	43	45	37	40	40	22	27	24	23	22	25	24	22	25	28	2	0					
Tobacco Moth:	64	54	55	46	46	48	48	45	46	47	45	45	45	45	42	44	46	43	46	43	45	42	42	42	28	31	29	28	30	31	30	27	31	33	14	12	0				
Samia:	65	54	53	47	47	47	49	45	46	45	43	47	45	45	41	45	48	45	47	44	47	43	43	42	29	31	28	28	29	33	32	28	32	32	15	14	5	0			
Snail:	63	60	54	53	53	55	56	53	54	54	52	54	52	52	51	54	54	50	51	47	49	46	47	44	21	29	25	23	28	31	30	26	26	26	29	27	29	29	0		
Rhodymenia:	71	60	70	51	49	45	44	48	44	45	45	44	48	44	48	47	47	51	45	71	60	66	60	61	57	57	60	57	60	61	61	58	64	58	58	62	60	61	60	0	
Porphyra:	80	51	53	28	25	28	30	30	28	25	25	25	27	27	25	27	27	23	51	44	36	36	38	44	40	36	40	38	42	42	40	40	46	47	36	36	36	38	44	28	0

10. Matrix of minimum mutation distances.

	Rhodospirillum:	Crithidia:	Euglena:	Pumpkin:	Rape:	Mungbean:	Elder:	Niger:	Sunflower:	Cotton:	Abutilon:	Castor:	Sesame:	Tomato:	Buckwheat:	Spinach:	Ginkgo:	Enteromorpha:	Candida:	Debaryomyces:	Saccharomyces:	Neurospora:	Humicola:	Ustilago:	Donkey:	Monkey:	Turtle:	Chicken:	Bullfrog:	Bonito:	Tuna:	Carp:	Dogfish:	Lamprey:	Fruitfly:	Screw-worm moth:	Tobacco moth:	Samia:	Snail:
Rhodospirillum:	0																																						
Crithidia:	91	0																																					
Euglena:	103	67	0																																				
Pumpkin:	92	78	73	0																																			
Rape:	91	77	74	5	0																																		
Mungbean:	93	74	70	6	6	0																																	
Elder:	91	79	76	10	9	8	0																																
Niger:	86	74	72	15	11	14	14	0																															
Sunflower:	90	73	72	16	16	13	13	9	0																														
Cotton:	91	77	75	10	8	9	10	12	15	0																													
Abutilon:	92	78	73	9	11	10	10	12	14	4	0																												
Castor:	93	80	76	10	9	12	12	14	16	6	8	0																											
Sesame:	92	78	75	8	10	8	13	14	14	9	8	8	0																										
Tomato:	91	78	77	9	8	9	9	11	11	8	6	7	6	0																									
Buckwheat	89	77	75	17	17	16	19	15	18	16	15	15	14	14	0																								
Spinach:	89	78	79	23	24	24	25	26	25	22	23	24	24	24	25	0																							
Ginkgo:	90	79	76	18	18	19	17	19	20	20	18	17	20	14	16	26	0																						
Enteromorpha:	94	74	78	33	29	33	32	31	29	32	33	34	33	30	35	30	36	0																					
Candida:	100	79	84	70	66	70	71	70	71	68	69	68	70	68	71	72	71	71	0																				
Debaryomyces:	90	77	81	70	67	72	70	70	71	71	70	69	70	71	72	72	71	67	29	0																			
Saccharomyces:	95	87	81	60	62	64	63	61	60	61	60	59	61	62	63	66	64	68	27	36	0																		
Neurospora:	92	88	70	60	59	61	64	59	59	61	60	61	62	59	61	66	61	61	52	55	50	0																	
Humicola:	89	79	70	69	68	69	71	68	66	68	67	68	68	66	64	63	66	67	46	49	52	31	0																
Ustilago:	97	77	70	64	63	65	68	65	65	66	67	67	67	67	67	67	66	65	38	46	49	44	40	0															
Donkey:	84	74	62	55	57	59	60	55	57	60	57	57	54	57	53	61	54	51	58	55	62	58	51	57	0														
Monkey:	85	75	66	50	52	53	53	50	49	51	48	48	44	48	46	52	47	59	64	55	54	55	50	57	16	0													
Turtle:	82	70	70	54	53	55	58	51	54	54	53	52	51	51	47	55	52	57	69	62	66	59	54	59	15	19	0												
Chicken:	83	76	72	57	55	58	59	53	55	57	55	54	53	52	47	55	53	58	68	61	62	54	52	59	15	18	8	0											
Bullfrog:	82	75	70	63	62	64	65	57	59	61	60	58	58	58	54	63	58	64	69	65	63	61	59	62	21	26	16	17	0										
Bonito:	81	75	69	65	63	66	69	59	59	64	64	61	61	62	62	64	63	61	66	66	65	63	57	62	27	30	26	25	20	0									
Tuna:	83	74	70	69	66	70	72	62	62	68	68	64	64	65	63	65	65	64	69	69	69	66	58	63	28	31	27	26	21	3	0								
Carp:	84	70	66	61	61	64	66	59	61	64	62	61	59	62	60	63	60	61	70	62	66	65	60	62	20	26	21	24	20	13	13	0							
Dogfish:	89	77	65	60	61	62	63	59	60	63	60	61	59	61	60	66	65	65	70	63	66	64	60	65	22	34	26	27	30	28	28	21	0						
Lamprey:	85	74	67	61	61	62	62	59	60	62	60	61	59	62	60	64	60	63	69	57	62	61	58	63	20	26	28	26	31	27	27	17	23	0					
Fruitfly:	83	72	74	64	63	63	66	57	59	64	63	61	61	61	55	61	59	61	65	62	66	53	53	59	27	37	31	30	29	36	35	31	35	37	0				
Screw-worm fly:	82	71	74	62	61	61	64	54	57	62	61	59	59	59	53	59	57	59	65	63	66	53	53	57	25	35	31	28	29	34	32	30	34	36	2	0			
Tobacco Moth:	83	75	75	57	58	57	60	55	57	59	55	57	54	55	52	55	54	58	63	60	62	57	54	58	30	38	38	36	39	42	41	37	38	39	15	13	0		
Samia:	84	75	72	57	58	54	60	53	52	54	51	57	52	53	48	54	54	59	62	61	62	55	55	58	32	37	36	34	38	44	43	38	40	38	19	16	6	0	
Snail:	87	81	77	62	63	65	65	63	66	66	63	66	63	63	61	65	66	64	64	57	65	62	59	60	32	42	39	38	47	49	46	40	40	39	35	32	33	34	0