Molecular evolution of cytochrome c from plants

Ramshaw, John A. M.

MOLECULAR EVOLUTION OF CYTOCHROME *c* FROM PLANTS

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

by

John A. M. Ramshaw

February 1972.

Department of Botany.

SUMMARY

Cytochrome *c* has been extracted and purified from four higher plants, mung bean, sunflower, niger and black gram, and from one fungus, *Cephalosporium acremonium*. The complete amino acid sequences of the cytochrome *c* of three angiosperms, mung bean, sunflower and niger, and of one gymnosperm, *Ginkgo biloba*, have been determined. The plant cytochrome *c* sequences determined in this investigation were compared with the other plant cytochrome *c* sequences and the homology of this group with other eukaryotic cytochromes was discussed. The cytochrome *c* sequences were used to examine times of divergence relating various taxonomic groups and to determine a higher plant phylogeny.

The results of both numerical comparisons and of tree constructions suggest that the three eukaryotic kingdoms all diverged at roughly the same time; the numerical comparisons indicate that this occurred approximately 1,500 million years ago. Both numerical matrix methods and the ancestral sequence method were used for constructing phylogenetic trees, but no unique phylogenetic tree relating the plant species could be constructed. The advantages and disadvantages of both construction methods were discussed, and it was concluded that the ancestral sequence method provided the better approach. However, this method did not give a unique solution, but the alternative trees found were all very similar and had many features in common. These features were in agreement with the available biological evidence and most of them were also found in the trees constructed by the numerical matrix methods. A unique tree relating ten plants was discovered which may represent a definite framework, from which a plant phylogeny can be established.

C O N T E N T S

STATEMENT

The determination of the amino acid sequence and part of the purification of mung bean cytochrome *c*, was done in collaboration with Dr. E. W. Thompson.

The ancestral sequence method program was written by Mr. A. A. Young.

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)

In figures of phylogenetic trees the following species abbreviations have been used:

AR	=	Arum
BW	=	Buckwheat
CA	=	Castor
CF	=	Cauliflower
CO	=	Cotton
GI	=	Ginkgo
MA	=	Maize
MB	=	Mung bean
NI	=	Niger
PU	=	Pumpkin
RA	=	Rape
SE	=	Sesame
SU	=	Sunflower
TO	=	Tomato
WH	=	Wheat

INTRODUCTION

The history of systematic botany has been divided by Alston & Turner (1963) into five major historical or developmental periods. During the first period, the megamorphic period, formal group concepts and a descriptive language were developed. Leeuwenhoek's invention of the microscope was responsible for the next developmental period, the micromorphic period, in which the recognition of hitherto unknown features made possible the acquisition of new morphological data. The third period, the evolutionary period, resulted from Darwin's work; most subsequent classification systems were constructed on a phylogenetic as opposed to artificial basis. The rediscovery of Mendel's laws on inheritance led to the first truely experimental approach to systematics during the fourth, cytogenetical period. The most recent period is the biochemical period, in which various chemical constituents of plants have been examined and compared in the solving of taxonomic problems. This has been possible because of the development of rapid and simple analytical techniques, particularly chromatography.

Initially, most work was concerned with micromolecules enabling 'biochemical profiles' to be constructed. More recently, limited use of macromolecules has been made through establishing biochemical pathways and by comparative enzymology.

It became clear when DNA was shown to be the genetic material, and to act by the production of messenger RNA molecules, that the sequence of base pairs in part of the

DNA determined the sequence of amino acids in particular proteins, and that vast amounts of evolutionary history may be hidden within their structures (Crick, 1958). Zuckerkandl & Pauling (1965) recognised that the total chemical content of any species was a document of its own evolutionary history. They divided molecules into three classes depending on the relative amounts of information which they carried; they inferred that the amount of history preserved will be the greater, the greater the complexity of the elements and the smaller the parts of the elements that have to be affected to bring about a significant change. The first class, the semantides, consists of the molecules which carry the information of the genes or a transcript thereof. The second class, the episemantic molecules, consists of molecules synthesised under the control of enzymes, semantides, in the absence of a template. The third class, the asemantic molecules, is those molecules that are not produced by the organisms which do not, except by their presence or absence, express any information that the organism contains. The largest information content lies within the semantides which may be further divided into three levels. The genes are the primary semantides, messenger RNA molecules, secondary semantides, and the resulting polypeptides, the tertiary semantides.

The determination of DNA sequences is at present a very difficult technical problem; the preparation of homologous sections of DNA from different species presents a further problem to the examination of primary semantides. Similar preparation and sequencing difficulties exist in the study

of messenger RNA. Nucleic acid hybridisation technique provides the only really useful method at present available for obtaining information from these categories of molecule (Kohne, 1968). However, with this method difficulties in both technique and interpretation exist; particular difficulties arise from the fact that in evolution inversions, translocations and repetitions of the DNA occur. Investigations on proteins, tertiary semantides, offer the best approach for the study of macromolecules. Generally they are easily purified and homologous proteins may be readily obtained from different species. The use of serological cross-reactivity, amino acid composition, peptide maps, electrophoretic mobility, catalytic activity, and chromatographic characteristics, have all been used for such comparisons (Bryson & Vogel, 1965; Boulter, Thurman & Turner, 1966; Hawkes, 1968; Nolan & Margoliash, 1968; Vaughan, 1968). These methods are relatively non-specific, and the greatest insights into molecular evolution could be expected to come from primary structure determinations. However, in the translation from DNA to protein, there is a loss of information. Because of the degeneracy of the genetic code differences in base sequences in allelic stretches of DNA may not lead to differences in amino acid sequence (isosemantic heterozygosity). Secondly, many regions of DNA exist which are not expressed in polypeptide products. These losses of information are not however, sufficient to preclude the use of protein sequence determination for studying evolutionary history.

The amino acid sequences are now known for many proteins

of various types from a wide range of species (Dayhoff, 1969); this data has been used to construct phylogenies relating the species examined (Fitch & Margoliash, 1967; Dayhoff, 1969). The tree derived by classical comparative methods and from the fossil record (Romer, 1966), agrees closely with this 'protein phylogeny'. Generally, the proteins examined have been from animal sources so that the phylogenies produced are complementary to the existing data. A major use of protein taxonomy is its application to the plant kingdom, where no continuous fossil record exists comparable to that for the animal kingdom. By the Upper Cretaceous, when flowering plant fossils became abundant, many of the major present-day orders were already represented (Walton, 1953). Numerous conflicting phylogenetic schemes relating plant species have been constructed (Lawrence, 1951), based on morphological data and later incorporating cytological, serological and simple biochemical evidence. The construction of a protein phylogeny provides a potential means of resolving this confusion. The choice of a suitable protein for such an investigation is important. Although sequencing techniques are now well established, the difficulty involved increases rapidly with increasing protein size; ideally, the protein should be small, having a molecular weight between 10,000-20,000. Plants are generally poor sources of enzyme proteins so a protein which is soluble and readily extracted is desirable; thermal stability, pH stability and an easy assay method, for example, by colour due to a prosthetic group, are all advantageous to the purification. The protein must be completely distributed through at least the

range of species to be studied and be performing the identical function throughout. The rate at which different proteins "accept" mutations varies (Dayhoff, 1969), and the rate of change must be suitable to provide a differentiation between the species examined.

For a study of the relationships within the plant kingdom and between the plant kingdom and other kingdoms, cytochrome c provides an excellent choice of protein.

Cytochrome c is distributed throughout all aerobic eukaryotic organisms and consists of a single polypeptide chain of molecular weight about 13,000, with a covalently bound haem group. It is located on the inner mitochondrial membrane (Backmann, Allman & Green, 1966), and functions in electron transport in the terminal oxidation chain. It is relatively stable to both temperature and moderate pH changes, and has a characteristic haemochrome spectrum, a basic isoelectric point of about pH 10 and a redox potential of around +250 mv. It reacts with mammalian cytochrome oxidase preparations while closely related proteins of bacterial and photosynthetic origin do not react (Yamanaka, 1966).

Cytochrome c was first observed in 1884 by MacMunn (1884, 1886, 1887), as part of a four-banded adsorption spectrum, which he observed to be widely distributed among vertebrates and invertebrates. He named the material either myohaematin or histohaematin, depending on the tissue type in which it was found, and noted that the spectrum appeared on chemical reduction and disappeared on oxidation; from this he concluded that these compounds had a respiratory function. The same four-banded spectrum was rediscovered in 1925 by Keilin, who

confirmed MacMunn's observations. He confirmed the wide distribution of the cellular pigment, "the cytochrome", and showed its presence in fungi and higher plants, as well as in animals (Keilin, 1925). Its importance in respiration was indicated by the demonstration of oxidation-reduction *in vivo*. He interpreted the spectral bands at 604, 566 and 550 nm, as being the α-bands of three haemochrome components, a, b and c, of the cytochrome and the band at 520 nm to be the combined ß-bands of these components. The possible reasons behind the total dormancy of the field between the two discoveries have been reviewed (Keilin, 1966).

Of the cytochrome components only cytochrome *c* could be easily extracted and purified (Keilin, 1930). Early purification procedures used its high thermal stability (Theorell, 1936), its stability to acid pH (Keilin & Hartree, 1937), and its high basic isoelectric point (Theorell & Åkesson, 1939), but none gave a completely pure or native product. The use of cation exchange resins for purifying cytochrome *c* (Paleus & Neilands, 1950; Nielands, 1952; Boardman & Partridge, 1953, 1954, 1955; Margoliash, 1954a, b) enabled preparations of a purity compatible with crystallization to be readily prepared (Bodo, 1955; Kuby *et al.*, 1956; Okunuki, 1960).

This availability of pure preparations has enabled many physical and chemical properties of the protein from animal, fungal and plant sources, to be examined (see Margoliash & Schejter, 1966), and despite the wide range of sources these properties are found to be essentially similar throughout. The complete amino acid sequence of horse heart

cytochrome c was determined in 1961 (Margoliash, Smith, Kreil & Tuppy, 1961), and the three-dimensional structure of horse ferricytochrome c to 2.8 Å resolution in 1971 (Dickerson et al., 1971). Cytochrome c from many other sources has also now been sequenced (see Dayhoff, 1969); however, the majority of these are of animal origin and at the beginning of the current investigation only one sequence, that of wheat, was of plant origin (Stevens, Glazer & Smith, 1967). These sequences have been used to construct a phylogeny relating the species examined (Fitch & Margoliash, 1967; Dayhoff, 1969), which, with minor exceptions, agrees with that constructed from other data. The sequence data indicated that cytochrome c accepts mutations very slowly and to be ideally suited for determining relationships between major taxonomic groupings.

The aim of the present investigation was to purify and sequence various plant and one fungal cytochrome c, in order to use the data as part of an investigation into the phylogeny of plants. Phylogenetic trees were to be constructed by computer programmes, written in collaboration with Mr. A. A. Young of the Computing Laboratory, Durham, using the strategies already described by Dayhoff & Eck (1966), Fitch & Margoliash (1967), and Lance & Williams (1966). Plant data accumulated by other workers in the Botany Department of Durham University, have been used, in addition to my own, in the construction of phylogenetic trees.

MATERIALS AND METHODS

I. GENERAL

1. Biological Materials

Seeds were obtained from the following suppliers:

Phaseolus aureus L. = *Vigna radiata* (mung bean) and

Phaseolus mungo L. var. *radiatus* = *Vigna mungo* Linn. Heppe (black gram)

from Hussein and Co., Newcastle upon Tyne.

Helianthus annuus L. (sunflower)

from Tyneside Seed Stores, Gateshead.

Guizotia abyssinica Cass. (niger)

from H. Garnham and Sons, Sacriston, Co. Durham.

Seeds were soaked for 24 h in running tap water prior to being germinated in trays under an intermittent fine spray of water at 20-25 °C for 3-7 days in the dark. The amount of water used was adjusted to prevent flooding of the trays. The seeds were occasionally turned by hand during the first 2-3 days of germination.

Cephalosporium acremonium (Strain 3227E) was obtained from The Glaxo Co. Ltd., Barnard Castle, Co. Durham.

Mycelia was taken from an industrial fermentation tank after 60-70 log hours growth. It was allowed to stand for 1 h and the oil from the medium, which collected on the surface, removed by suction. The mycelia was then harvested by filtration through heavy twill cloth. An acetone dried powder of cells was prepared by twice resuspending, filtering and washing the mycelia with cold acetone. After air drying, the powder was stored at -10 °C.

2. Chemicals and Reagents

With the exceptions listed below, chemicals were obtained from British Drug Houses 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)
Papain, E.C.3.4.4.10 (" ")

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

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

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

Ginkgo Biloba L. (maidenhair tree) cytochrome c

was a gift from Dr. M. Richardson.

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. & R. Balston (Modified Cellulose) Ltd., Maidstone, Kent. (Whatman).

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

(Rohm and Haas Co., Philadelphia, U.S.A.) was obtained through B.D.H. Chemicals Ltd.

Alanyl-alanine
Alanyl-serine
Acetyl alanine
Arginine (free base)
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.

Ampholine Carrier Ampholytes, 8% (w/w) aqueous solution, pH 5-8 and pH 7-10

were obtained from L.K.B. Producter AB, Stockholm, Sweden.

Arginylarginine

was obtained from Cyclo Chemical Corporation, Los Angeles, California, U.S.A.

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.

α-N-acetyl-L-lysine

was obtained from Calbiochem, Los Angeles, California, U.S.A.

All chemicals were used as supplied, except for phenyl-

iso-thiocyanate, which was vacuum distilled once before use.

The preparation of ϵ-N-trimethyllysine was as described by Delange, Glazer & Smith (1969). Part of the product was purified by paper chromatography using the solvent system of Stewart (1963). The R_F values and electrophoretic mobility of the preparation agreed with those quoted by Delange et al. (1969), and it gave a single ninhydrin positive peak on the autoanalyser, eluted 43 min before lysine.

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 Street, 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_3Citrate.$2H_2O$, 58.8 g/litre) 3.6 ml.

b. Chromatographic location reagents

(i) Cadmium-ninhydrin reagent (Heilmann, Barrolier & Watzke, 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) Starch-Iodide reagent (Pan & Dutcher, 1956)

Solution A.	Sodium Hypochlorite solution (2% (w/v) available chlorine).
Solution B.	95% Ethanol.
Solution C.	A mixture of equal volumes of freshly prepared 1% soluble starch and 1% potassium iodide solutions.

The electrophoretograph was sprayed with solution A, and when dry, it was sprayed with solution B. Finally, when solution B had evaporated, it was sprayed with solution C. A positive reaction was shown as a blue colour against a colourless background.

(iii) 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.

(iv) 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 acetone. The paper was dipped and air dried. A positive reaction, given by reducing sulphur compounds, showed as bleached areas on a pink background.

(v) 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 arginyl-arginine 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 deionized 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 3.80 (Hamilton, 1962)

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

Adjusted to pH 3.80 with HCl, made up to 8 litres with deionized 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, 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		1,200.0 ml
Water	made to	12,000.0 ml

Prepared at least 1 day before use; and stored in 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	12,000.0 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 & Wang, 1967).

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

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

Solvent 3A - Ethyl acetate-methanol-acetic acid (20:1:1, by vol.); (Crowshaw, Jessup & Ramwell, 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.).

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.

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 adsorption at 550 nm with a low dispersion, direct vision hand spectroscope (R. & 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 & Frohwirt, 1959).

2. Criteria of Purity of Cytochrome c

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

3. Dialysis

Unless stated otherwise, 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 seedlings

Etiolated seedlings were grown and extracted in separate batches of 20-40 kg fresh weight. Each batch was blended in approximately 2 kg quantities in a 5 litre Waring blender for 3 min with 1-2 litres of prechilled distilled water (2-4 °C), 30 g ascorbic acid and 1 g E.D.T.A. The pH of the total homogenate was adjusted to pH 4.6 ± 0.1 with 2N-HCl.

The homogenate (50-100 l) was then filtered at 2-4 °C through Whatman No. 6 paper on ten 27 cm Buchner funnels, using a mild vacuum (ca 400 mmHg). Dry filter cakes were each washed with I l of distilled water.

Alternatively, filtration was achieved through Terylene bags (Type 1481, Samuel Hill Ltd., Rochdale, Lancs.), in a 21" perforated basket centrifuge (Type 86, Thos. Broadbent and Sons Ltd., Huddersfield).

The clear filtrate was adjusted to pH 8.0 ± 0.2 by 1 M-tris and allowed to stand for 2-5 h at 2-4 °C; any resulting precipitate was removed by filtration under vacuum through Whatman No. 6 paper on the Buchner funnels.

5. Extraction of cytochrome c from acetone dried Cephalosporium

Batches of 250 g acetone dried cells were resuspended in I l of 0.5 M-NaCl, stirred vigorously for 1 h, and then the slurry filtered through heavy twill cloth. The filter cake was resuspended in 500 ml 0.5 M-NaCl, stirred vigorously for

a further 1 h and then filtered. The filtrates at 2-4 °C were pooled, diluted by 4 volumes of deionized water and adjusted to pH 8.0 ± 0.2 with 1 M-NaOH.

6. Ion Exchange Chromatography on Amberlite CG-50 resin.

a. Preparation of resin

(i) Sodium form

The resin was prepared in the sodium form by the method of Margoliash & Walasek (1967), with final equilibration in 0.1 M sodium phosphate buffer, pH 8.0.

(ii) Ammonium form

Used 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 by 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 by 10-15 vol. distilled water on a sintered funnel, resuspended 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 by 5-10 vol distilled water on a sintered funnel and resuspended as a slurry ready for use.

b. Adsorption

Filtrate at pH 8.0 was passed through columns (6.5 cm x 25 cm) of Amberlite CG50, at a flow rate of 250-400 ml/h. Approximately 15-20 l of filtrate was passed through each individual column. The adsorption was carried out at 2-4 °C.

c. Elution

(i) Batchwise (NH_4^+ form resin used).

The resin was removed from the column and washed with distilled water by stirring and decantation, until the supernatant was colourless. It was then stirred with 500 ml 0.5 M-NaCl and 2 M-NaOH added to maintain the pH at 8.0. After stirring at constant pH for 1-2 h at room temperature, the suspended resin was poured back into the columns and the supernatant collected. The resin was washed through with 0.5 M-NaCl (ca 1 litre) until the washings exhibited no further adsorption at 550 nm when reduced with a trace of ascorbic acid; adsorption was determined with a direct vision, low-dispersion, hand spectroscope (R. and J. Beck Ltd., London).

(ii) Column Elution (Na^+ form resin used).

The cytochrome c was eluted from the resin by 0.5 M-NaCl in 0.1 M sodium phosphate buffer, pH 8.0, containing a trace of ascorbic acid and collected as a single fraction.

7. 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 c 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.

8. 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 re-suspended in a minimum vol of distilled 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.

9. 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.

10. 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 (D.C.L. 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 under a linear gradient from 10 mM- to 300 mM-sodium phosphate, pH 7.2, formed using a device as described by Bock & Ling (1954).

11. Isoelectric Fractionation by Ampholine Electrofocussing

Isoelectric fractionation was carried out using an L.K.B. 8102 model Electrofocussing apparatus.

The anode solution contained 0.1 ml conc. H_2SO_4 in 40 ml water, and the cathode solution contained 0.1 g diaminoethane, 0.2 g morpholine, and 48 g sucrose in 56 ml water. The stock ampholyte solution contained 100 μl ethanolamine, 100 μl diethanolamine, 100 μl of diaminoethane, 400 mg arginine (free-base), 10 ml of 8% (w/v) ampholine solution, pH 7-10, 1 ml of 8% (w/v) ampholine solution, pH 5-8, in water to a volume of 50 ml. For the preparation of the stabilizing sucrose gradient two solutions were prepared. The less-dense solution contained 12.5 ml stock ampholyte solution made up to 220 ml with water; the dense solution contained 37.5 ml of stock ampholyte solution and 100 g sucrose made up to 155 ml with water. A linear gradient was pumped into the column over the cathode solution using a device as described by Bock & Ling (1954), to form the gradient. When the column was approximately half filled, the gradient making was interrupted and the sucrose concentration at that point estimated using a hand refractometer (Bellingham & Stanley Ltd., London, N.15).

The sample (20-50 ml), which had previously been dialysed against mM-NH_4OH, was made up to the measured concentration by the addition of sucrose and pumped into the column. The gradient was then continued to 1 cm below the upper electrode and then the anode solution pumped into the column. The apparatus was maintained at 2-4 °C by water cooling. The power supply was initially regulated to supply 5 mA at 400 V and subsequently raised to 1 kV (maximum load 6W).

12. Desalting by Gel-Filtration

Pure cytochrome c preparations were concentrated by rotary evaporation and desalted by passing 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.

III. PROTEIN SEQUENCE DETERMINATION METHODS

1. Denaturation of Cytochrome c

Cytochrome c was denatured with ethanol, based on the method of Margoliash, Kimmel, Hill & Schmidt (1962). Cytochrome was dissolved in water (10 mg/ml) and oxidized 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 centrifuging with absolute ethanol, 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 Radiometer TTTlc 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 addition.

The rate of digestion was recorded by the titrator (Radiometer SBR2c), and digestion terminated by adjusting to pH 4-5 with acetic acid and freezing at -20 °C.

3. 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 S.W.3, 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) in the flat-plate apparatus.

b. Location of peptides on paper electrophoretograms.

Papers were treated with the detection reagents in the orders 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-arginyl-arginine, and at pH 1.9 from 1-dimethylaminonaphthalene-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

(i) Precipitation method

The digest, after adjusting to pH 4.5 with acetic acid, was stood overnight at 4°. The majority of the haem peptide was removed as a dark-brown precipitate by centrifugation.

(ii) Elution method

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

When both methods were used for isolating the haem peptide, all the material was pooled.

4. Quantitative Amino Acid Composition of Proteins and Peptides

Quantitative amino acid analyses of peptides and proteins were made using a Technicon automatic amino acid analyser, with a 133 cm column of Chromobead type A resin. The nine-chamber Autograd 18 h elution system was used and the recorder had a scale expander fitted. Protein (0.2-1.0 mg) or peptide (0.05-0.5 μmol) was hydrolysed at 110°C in evacuated Pyrex tubes with constant boiling 5.7 M-HCl (Moore & Stein, 1963).

For peptides containing or suspected to contain tryptophan, 5% (v/v) thioglycollic acid was added with the HCl (Matsubara & Sasaki, 1969).

5. 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, E.C.2.). 50 μl of constant boiling 5.7 M-HCl was

added and the tube sealed and heated at 105° for 20 h. The acid was removed in vacuo over NaOH and the free peptide amino acids labelled by the dansyl method (see Section III, 6c), but without final hydrolysis. The dansyl amino acids were identified by chromatography on polyamide layers (see Section III, 6d).

6. Peptide Sequencing Methods

a. N-terminal method

The dansyl-Edman procedure of Gray & 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 carboxypeptidase-A and the amino acids liberated then identified as their dansyl derivatives after labelling but without acid hydrolysis.

c. The dansyl procedure (Gray & Hartley, 1963b)

A 1-10 nmol peptide sample was transferred to a clean Durham tube (30 mm x 6 mm, A. Gallenkamp Ltd., London, E.C.2), and dried in vacuo over NaOH and conc. H_2SO_4. The Durham tubes were cleaned by washing or by baking at 550 °C (Gray & 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 40-50 µl of constant boiling 5.7 M-HCl was added, the dansyl peptide hydrolysed in the

sealed tube 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 & 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 of the sheet. Frames of the type described by Smith (1958) were used to support up to five sheets for simultaneous separations.

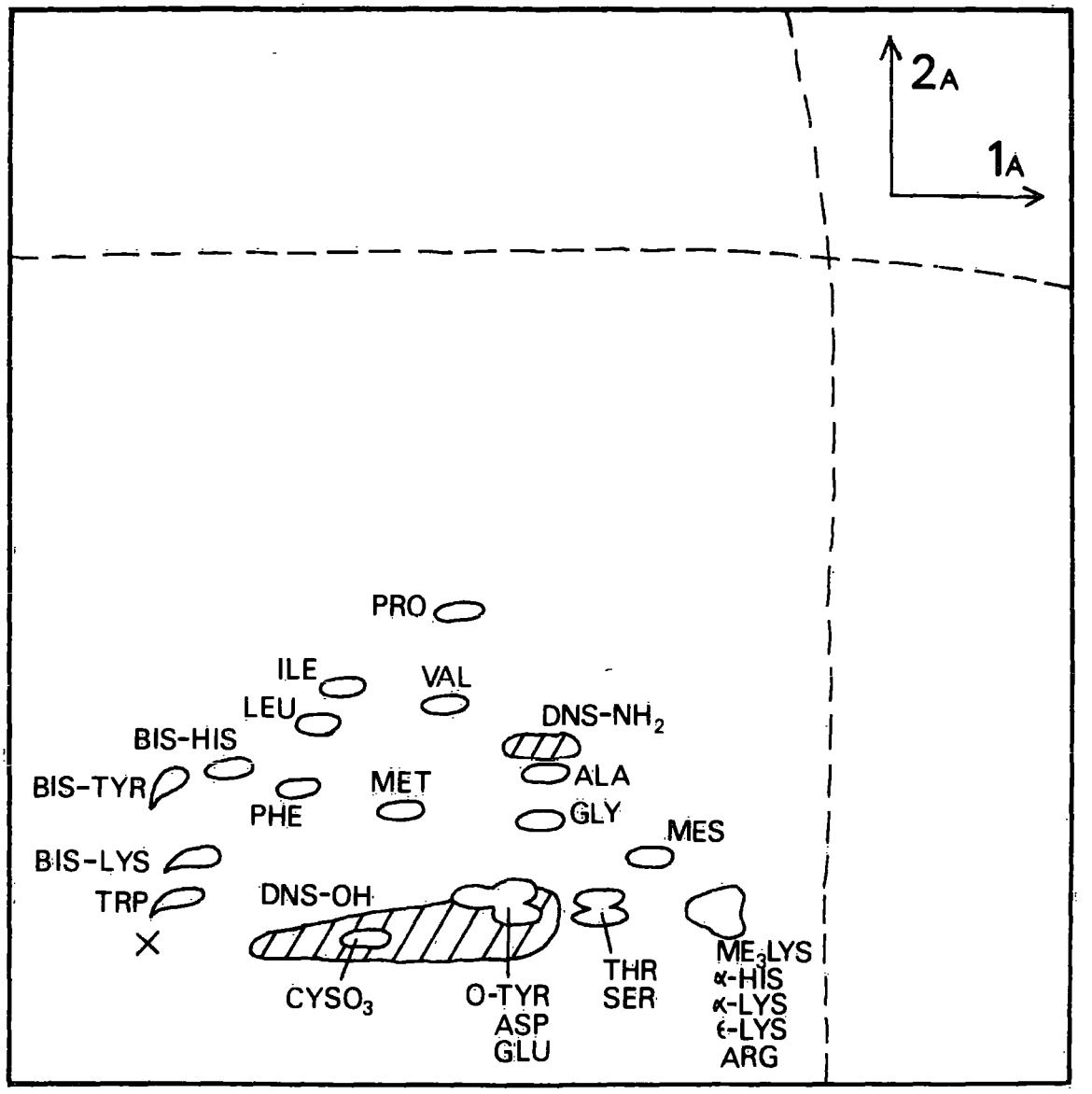
Chromatograms were initially developed in two dimensions. Solvent 1A was run for 40 min and the sheets dried for 30 min in a hot air draught. After cooling they were run in either solvent 2A or 2B for 40 min. The results of the two-dimensional separation were then recorded. Poorly resolved groups of 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. 1, 2 and 3. The cochromatography of samples and standards was used to help the identification of close pairs of derivatives.

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, Blomback, Edman & Hessel, 1966)

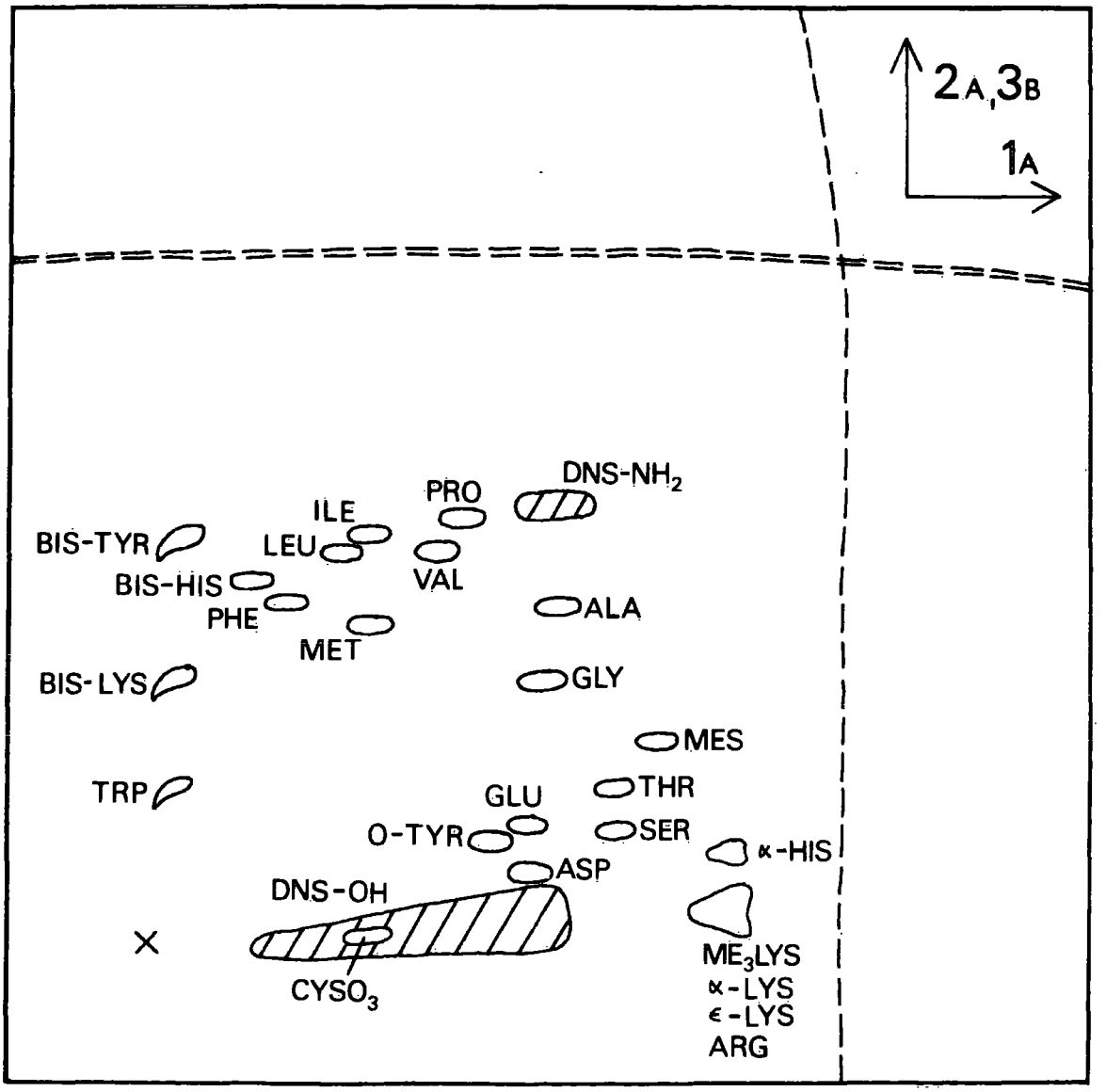
The peptide (0.1-0.5 µmol) was dissolved in 150 µl of

FIGURE 1. Chromatography of dansyl-amino acids on polyamide thin layers



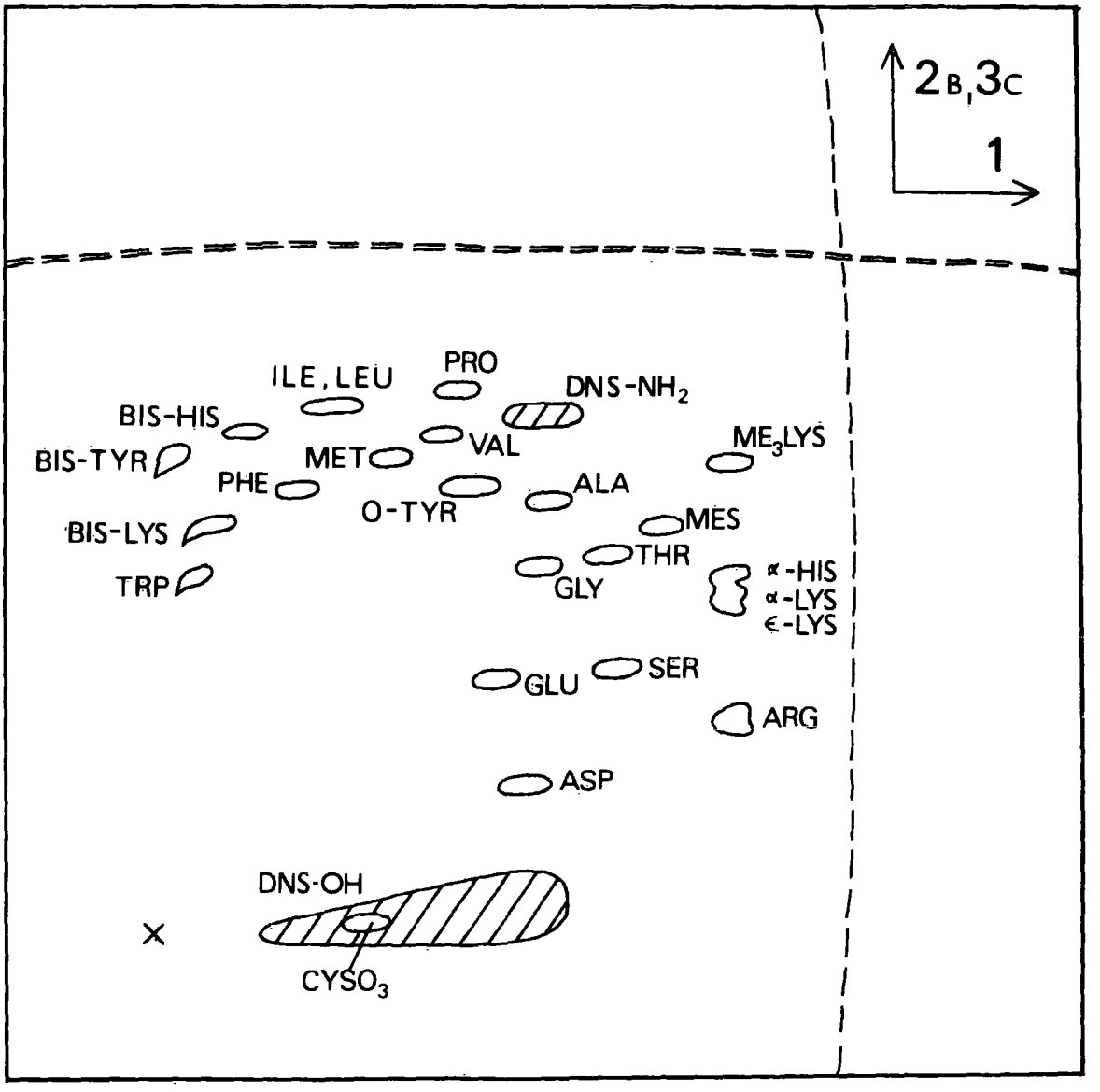
Development was by Solvent 1A in the first dimension, and Solvent 2A in the second dimension.

FIGURE 2. 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.

FIGURE 3. Chromatography of dansyl-amino acids on polyamide thin layers.



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

5% (v/v) redistilled phenylisothiocyanate in pyridine. The tube was flushed with oxygen-free nitrogen (Ilse & 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 & Toseland, 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 °. 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 °. Reaction was stopped by freezing and freeze-drying.

(iii) Papain digestion

Peptides were dissolved in 200 µl of 0.01 M-sodium phosphate buffer, pH 6.5, containing 20 µg of papain and 0.05% of 2,3-dimercaptopropan-1-ol (Smith & Kimmel, 1960). Digestion was at 37 ° and was terminated by freezing and freeze-drying.

(iv) 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° (Potts, Berger, Cooke & Anfinsen, 1962). Reaction was stopped by drying in vacuo.

g. Removal of haem moiety from haem peptides for sequence analysis

(i) Mercuric Chloride method (Ambler, 1963)

The method used was as described by Ambler (1963), with quantities scaled down proportionately for the smaller amounts of peptide used.

(ii) Direct performic oxidation method (Nolan & Margoliash, 1966)

The haem peptide was taken up in 200 µl of 90% (v/v) formic acid, and either two or three 10 µl portions of 30% (w/v) H_2O_2 were added at 10 min intervals to performic oxidize 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.

7. Determination of Amide Residues

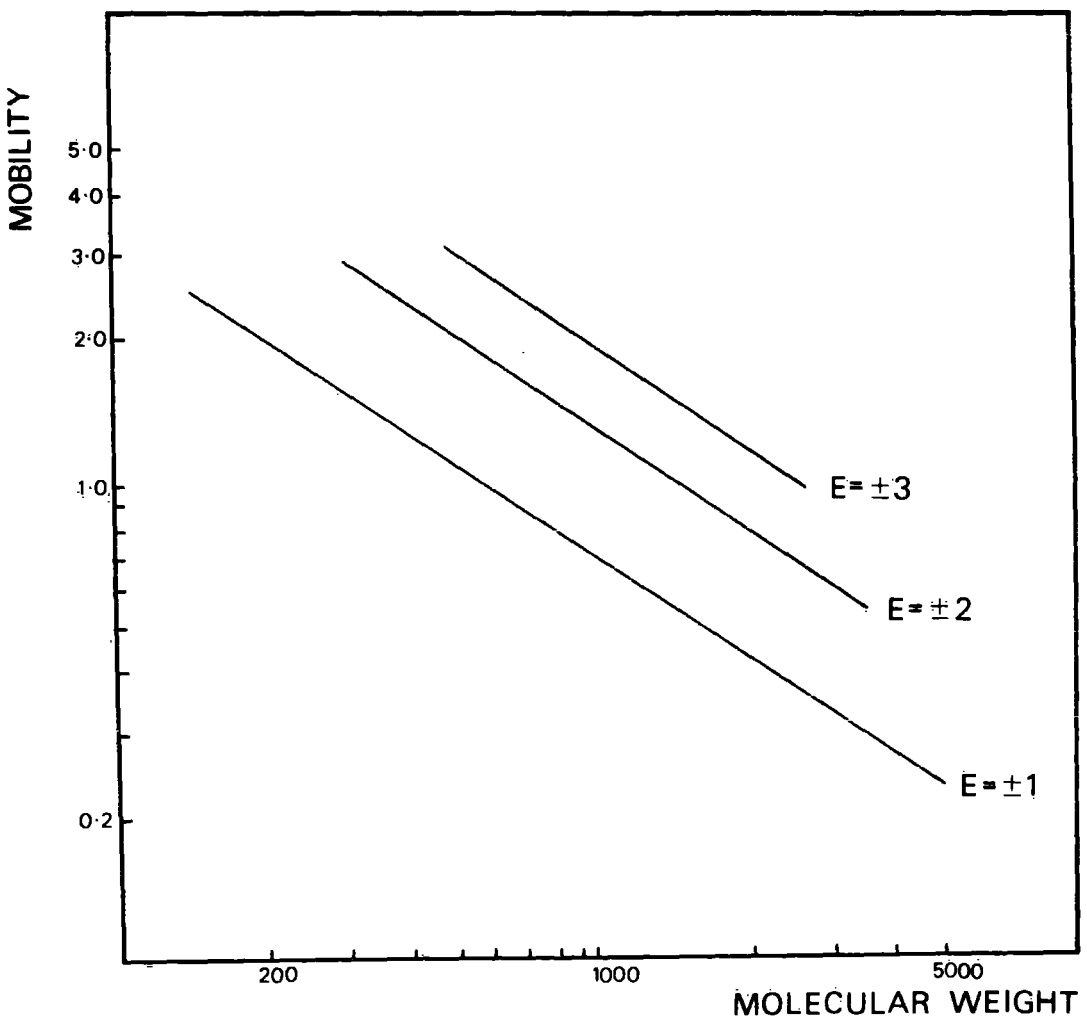
Amide residues were determined from peptide mobilities at pH 6.5, using the method of Offord (1966). A graph was drawn giving peptide charge from mobilities relative to the dansyl markers used (Fig. 4).

8. Determination of Acetyl Group

Acetyl groups were determined as 1-acetyl-2-dansyl-hydrazine derivatives (Schmer & Kriel, 1969).

The sample was dried in a Durham fermentation tube (6 x 30 mm). 25 µl of 0.1 M-HCl was added and dried in

FIGURE 4. Electrophoretic mobility of peptides.



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

vacuo over NaOH. 20 μl 95% + hydrazine was 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) was added and the tube sealed by parafilm, and heated at 35 °C for 16 h. After drying the 1-acyl-2-dansylhydrazine derivative was identified by chromatography on polyamide sheets.

9. Nomenclature Used to Describe Sequence Analysis

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, those by T, tryptic peptides, and those by P, papain peptides. Peptides derived by further cleavage are given a subscript to the parent peptide. All residue numbering is 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. Residues given in parentheses were determined from peptide composition, and the order determined by other evidence.

IV. CALCULATIONS BASED ON AMINO ACID SEQUENCE DATA

All complex calculations were performed on an I.B.M. 360/67 system computer, using appropriate programmes written in P.L.1.

1. Matrix of Differences 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 as described by Fitch & Margoliash (1967a) was used. Minimum mutation distances were calculated from the genetic code given by Dayhoff (1969).

2. Phylogenetic Tree Construction

a. Numerical matrix methods

Trees were constructed based on the methods of Fitch & Margoliash (1967a) and Lance & Williams (1966a).

b. Ancestral sequence method

Trees were constructed based on the method described by Dayhoff & Eck (1966).

RESULTS

1. The Purification of Mung Bean Cytochrome c

The first method used for preparing cytochrome c from mung bean was that of Laycock (1968).

Mung bean seeds were germinated for 4 days prior to extraction. The homogenate was filtered through paper and gave reasonably dry cakes after overnight filtration. After adjusting the filtrate to pH 8 the cytochrome was adsorbed on Amberlite resin in the NH_4^+ form, and then eluted batchwise. The eluates were pooled and stored at -20 °C until sufficient impure cytochrome had been collected. The total was then dialysed and adsorbed on to 3 x 20 cm columns of CM-52 cellulose resin equilibrated to pH 8 in the same buffer as was used for the dialysis. The red cytochrome c band was eluted with 0.5 M-NaCl and the total collected as a single fraction. It was then dialysed for 24 h against 1% (w/v) glycine and 24 h against mM-NH_3 solution; any precipitate which formed was removed by centrifugation. It was then purified using isoelectric fractionation by Ampholine electrofocussing. The cytochrome focussed as a single reduced band; commercially available purified horse cytochrome c separated into 4 bands. A more basic bright green protein moved below the cytochrome band and dispersed into the cathode solution. A large quantity of white protein focussed in the upper part of the column. As the electrofocussing proceeded, this material precipitated out and fell through the column disturbing the gradient and the other focussed protein bands. The column was fractionated after 110 h before focussing was complete, as the cytochrome band had nearly reached the cathode solution boundary. The cytochrome

fractions were pooled and dialysed exhaustively against distilled water; the preparation was then freeze-dried and stored at -20 °C.

An $E_{410}(ox)/E_{280}(ox)$ ratio for the preparation of 1.0 indicated that this method had not given a pure product.

In an attempt to further purify the cytochrome, it was redissolved and adsorbed on to a 3 x 8 cm column of CM-Sephadex. The column was washed through with I l. of 0.05 M Na/PO_4 buffer, pH 8, and then the cytochrome eluted by 0.5 M-NaCl. After exhaustive dialysis against distilled water the preparation was freeze-dried.

An $E_{410}(ox)/E_{280}(ox)$ ratio for the preparation of 3.1 indicated the purity was greatly increased, but still not complete.

The yields and purities of the cytochrome obtained using this method of purification are shown in Table 1.

The second method used for preparing cytochrome c from mung beans was based on the use of repeated gel filtration chromatography (Flatmark, 1964).

The cytochrome was extracted and concentrated using Amberlite resin by the same procedures as were used in the first method. Pooled eluates were dialysed and further concentrated by chromatography on CM-Sephadex. The second column used was washed through with buffer causing the cytochrome to migrate partly down the column. Resin adsorbing a black impurity was removed from the top of the column with a pipette, before the cytochrome band was eluted. The cytochrome was then purified by chromatography on Sephadex G-50 (Fig. 5); fractions with an $E_{410}(ox)/E_{280}(ox)$

TABLE 1

The purification of mung bean cytochrome c (method 1)

Step	Yield mg cytochrome/ kg starting material	Purity Ratio $E_{410}(ox)/E_{280}(ox)$
Amberlite CG-50	1.0-2.0	<0.01
CM-52 cellulose	0.8-1.0	0.1-0.2
Isoelectric focussing	0.6	1.0
Dialysis and freeze-drying	0.6	1.0
CM-Sephadex	0.4	3.1
Dialysis and freeze-drying	0.4	3.1

TABLE 2

The purification of mung bean cytochrome c (method 2)

Step	Yield mg cytochrome/ kg starting material	Purity Ratio $E_{410}(ox)/E_{280}(ox)$
Amberlite CG-50	1.0-2.0	< 0.01
CM-50 Sephadex	0.8-1.7	0.3-0.4
Sephadex G-50	0.6	0.8-1.0
Biogel P-30	0.4	2.5-4.0
Sephadex G-25	0.3	3.2

FIGURE 5. Chromatography of partially purified mung bean cytochrome c on Sephadex G-50.

A 3 cm x 30 cm column of resin was used; it was equilibrated and eluted by water.

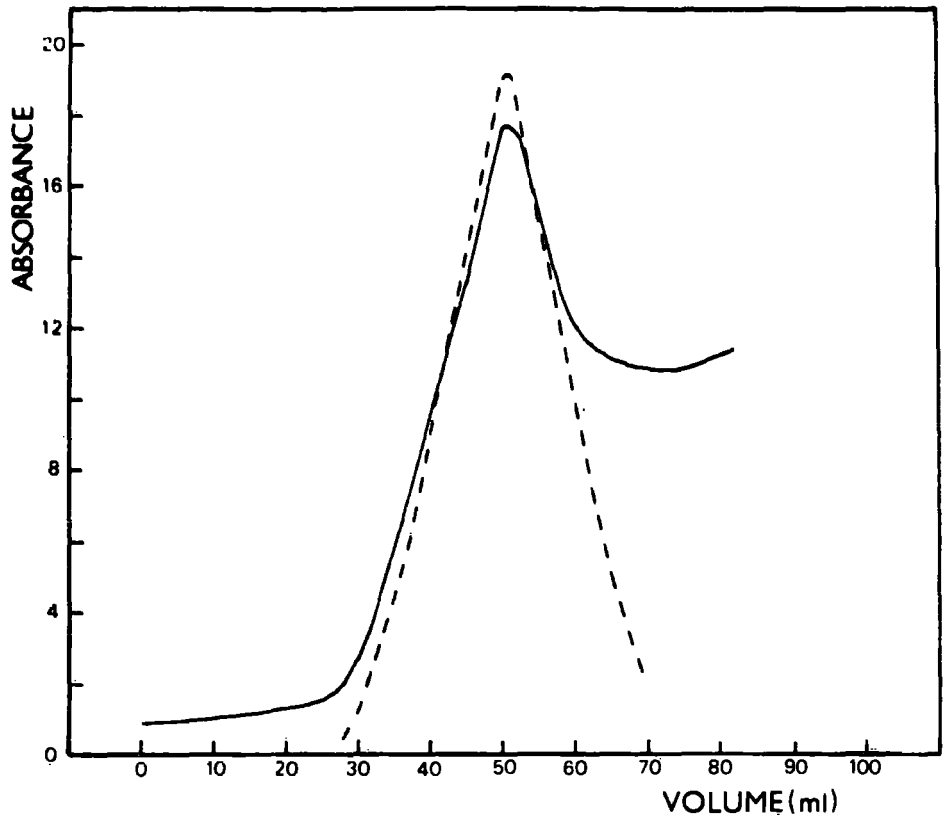
In Fig. 5 to Fig. 14 the following notation has been used:-

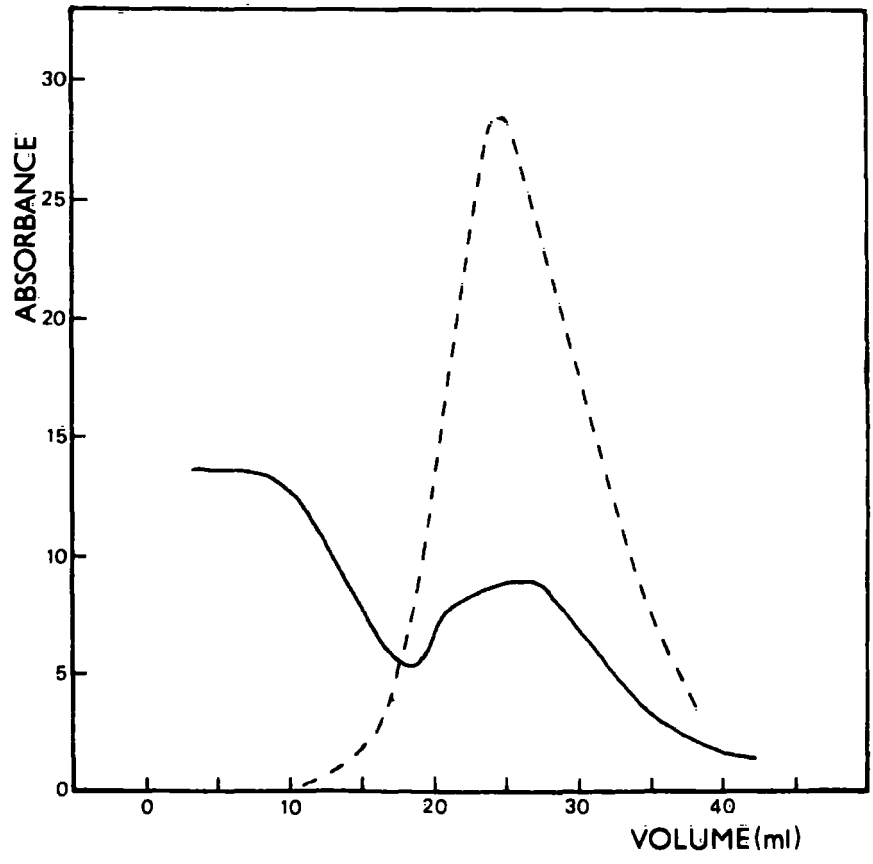
———	(continuous line)	Absorbance at 280 nm.
-------	(short dashes)	Absorbance at 410 nm.
— — — —	(long dashes)	Buffer concentration.

FIGURE 6. Chromatography of partially purified mung bean cytochrome c on Biogel P-30.

A 3 cm x 30 cm column of resin was used; it was equilibrated and eluted by 0.05 M sodium phosphate buffer, pH 8.0.

(See Fig. 5 for the notation used).


ABSORBANCE
20
16
12
8
4
0
0
10
20
30
40
50
60
70
80
90
100
VOLUME(ml)


ABSORBANCE
30
25
20
15
10
5
0
0
10
20
30
40
VOLUME(ml)

ratio greater than 0.8 were pooled. After rotary evaporation to reduce the volume the preparation was further purified by chromatography on a column of Biogel P-30 equilibrated and developed with 0.05 M Na/PO_4 buffer, pH 8 (Fig. 6).

Fractions with an $E_{410}(ox)/E_{280}(ox)$ ratio greater than 2.5 were pooled, rotary evaporated and desalted using Sephadex G-25. The $E_{410}(ox)/E_{280}(ox)$ ratio of the material was 3.2; the cytochrome was freeze-dried and stored at -20 °C.

Recycling of material through both the gel filtration steps could give material of higher purity ($E_{410}(ox)/E_{280}(ox)$ = 4.0), but with lower yields.

The yields and purities of the cytochrome obtained using this second method of purification for mung bean cytochrome c are shown in Table 2.

2. The Purification of Sunflower Cytochrome c

Sunflower seeds were germinated for 6-7 days before extraction. The homogenate was filtered through paper and adjusted to pH 8. The heavy precipitate which formed was removed by decantation and filtration. Amberlite resin in the NH_4^+ form was used to adsorb the cytochrome from the extract and was eluted batchwise. After dialysis the cytochrome was further concentrated by chromatography on CM-Sephadex. The columns were well washed with buffer which caused the cytochrome band to move down the column. Before the cytochrome band was eluted, a brown impurity which was left adsorbed at the top of the column was removed by pipetting out resin. The cytochrome was then desalted using a 3 cm x 30 cm column of G-50 Sephadex. The total cytochrome band

was collected and concentrated by rotary evaporation. The cytochrome was then purified by further chromatography on G-50 Sephadex. Fractions with an $E_{410}(ox)/E_{280}(ox)$ ratio greater than 2.5 were pooled (Fig. 7). A separation was achieved from two yellow proteins chromatographing in front of and behind the cytochrome band. Some of this material was used for sequence determination. The pooled cytochrome fractions were then adsorbed on to a column of CM-52 cellulose, which was eluted by an increasing cation gradient (Fig. 8). Five main cytochrome bands were eluted and each separately pooled; when analysed all were in the oxidised form. Band 1, which was fairly broad, was considered to consist of modified or deamidated forms of the protein. A preliminary tryptic digestion and amino acid analysis had shown that material purified by gel-filtration only had associated with it a glycine rich impurity. Material from each of bands 2-5 of the gradient elution was analysed on the amino acid autoanalyser and shown to still be contaminated by a glycine rich impurity. The amount of this impurity increased between bands 2 to 5, but no quantitative relationship between the different bands was established. Band 4, which accounted for half of the recovered cytochrome, was concentrated by rotary evaporation and desalted by chromatography through G-25. This preparation, which was not pure, was freeze-dried and used for chymotryptic digestion. The yields and purities of the cytochrome prepared from sunflower are shown in Table 3.

TABLE 3

The purification of sunflower cytochrome c

Step	Yield mg cytochrome/ kg starting material	Purity Ratio $E_{410}(ox)/E_{280}(ox)$
Amberlite CG-50	2.5-3.2	< 0.01
CM-50 Sephadex	2.0-2.5	0.05
Sephadex G-50	1.2-1.5	2.5-3.2
CM-52 Cellulose	0.6 (Band 4)	3.0-3.9
Desalting	0.5	3.9

FIGURE 7. Chromatography of partially purified sunflower cytochrome c on Sephadex G-50.

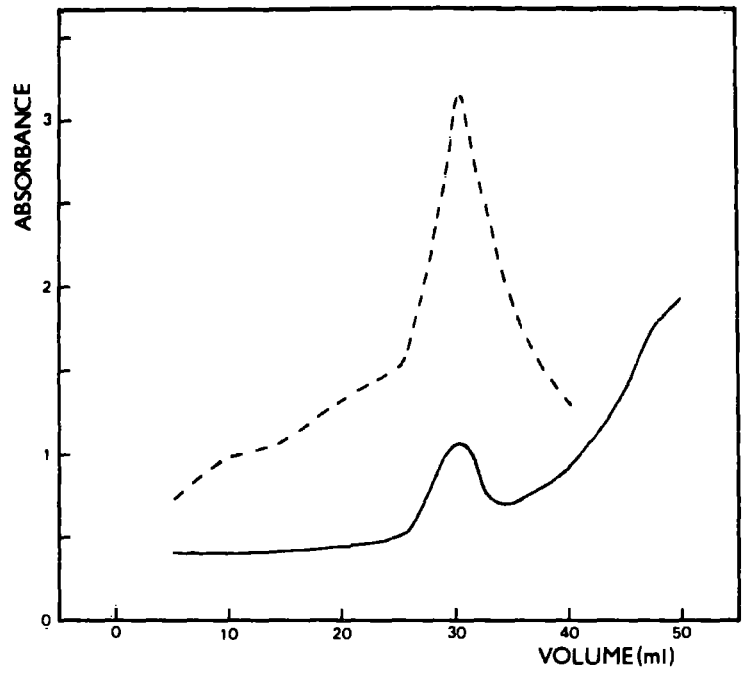
A 2 cm x 90 cm column of resin was used; it was equilibrated and eluted by water.

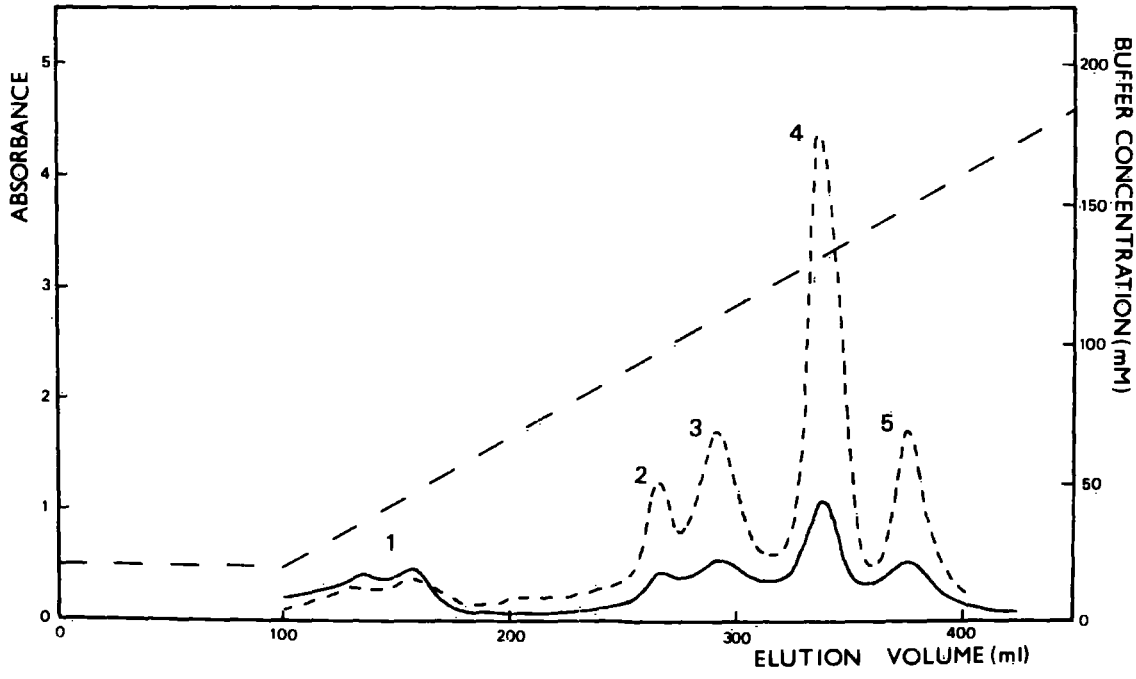
(See Fig. 5 for the notation used).

FIGURE 8. Chromatography of partially purified sunflower cytochrome c on CM-52 cellulose.

A 2 cm x 10 cm column of resin was used. The gradient was formed using 300 ml of each buffer.

(See Fig. 5 for the notation used).


ABSORBANCE
0
1
2
3
0
10
20
30
40
50
VOLUME(ml)


ABSORBANCE
0
1
2
3
4
5
0
100
200
300
400
ELUTION VOLUME (ml)
BUFFER CONCENTRATION(mM)
0
50
100
150
200
1
2
3
4
5

3. The Purification of Niger Cytochrome c

Niger seeds were germinated for 4 days prior to extraction. The homogenate was filtered by centrifugation and after adjusting to pH 8.0 filtered through the buchner funnels to give a clear yellow filtrate. Amberlite resin in the NH_4^+ form was used to concentrate the extract and the cytochrome eluted batchwise. After dialysis the eluates resulting from extracting 30 Kg seeds were concentrated by chromatography on CM-Sephadex. Ammonium sulphate fractionation was then used to further concentrate and purify the cytochrome; cytochrome was present in the 60% saturation pellet and was completely precipitated at 80% saturation. Although a large quantity of impurities were discarded, the purity of the cytochrome did not greatly increase by this procedure (see Table 4). The cytochrome pellets were dissolved in a minimum quantity of water and desalted and purified on a column of Biogel P-30 (see Fig. 9). The cytochrome separated from a yellow protein which moved more slowly through the column. All cytochrome fractions with an $E_{410}(ox)/E_{280}(ox)$ ratio greater than 1.8 were pooled and adsorbed on to a column of CM-52 cellulose. The cytochrome was eluted by an increasing cation gradient giving two main fractions (see Fig. 10). The first peak was mainly reduced cytochrome, together with modified oxidized cytochrome and the major second peak was oxidized cytochrome. All fractions from the oxidized peak having an $E_{410}(ox)/E_{280}(ox)$ ratio greater than 3.7, were pooled. After reducing the solution volume by rotary evaporation, the cytochrome was desalted on a column of Sephadex G-25, and then freeze-dried. Part of the cytochrome bound on to the resin considerably

TABLE 4

The preparation of Niger cytochrome c

Step	Yield mg cytochrome/ kg starting material	Purity Ratio $E_{410}(ox)/E_{280}(ox)$
Amberlite CG-50)) CM-50 Sephadex)	1.6-1.8	0.3
Ammonium Sulphate Fractionation	1.3	0.4
Biogel P-30	1.2	1.8-3.0 (2.7 pooled)
CM-52 Cellulose	0.7	3.8-4.3
Desalting	0.4	4.45

FIGURE 9. Chromatography of partially purified niger cytochrome c on Biogel P-30.

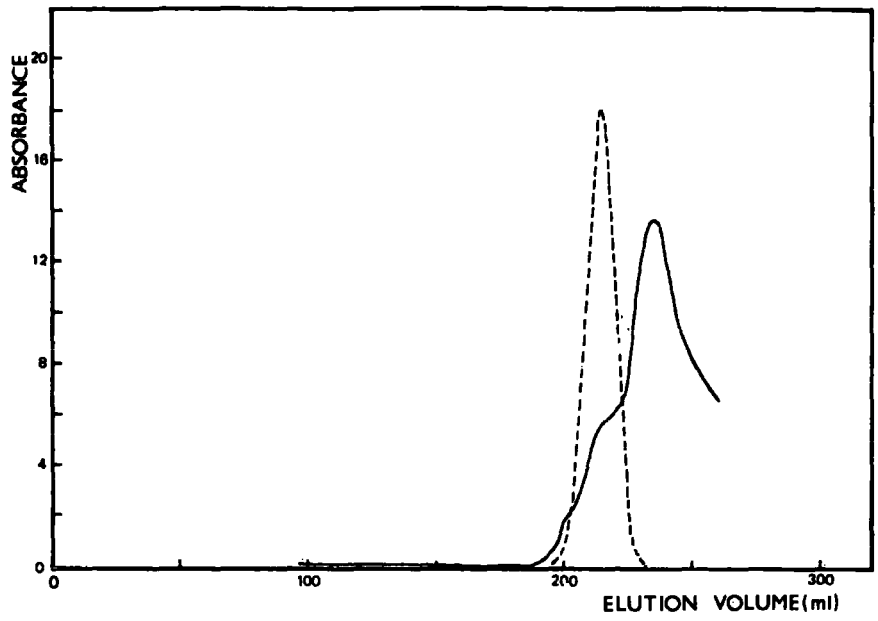
A 2 cm x 90 cm column of resin was used; it was equilibrated and eluted by water. The flow rate was 55 ml/hr.

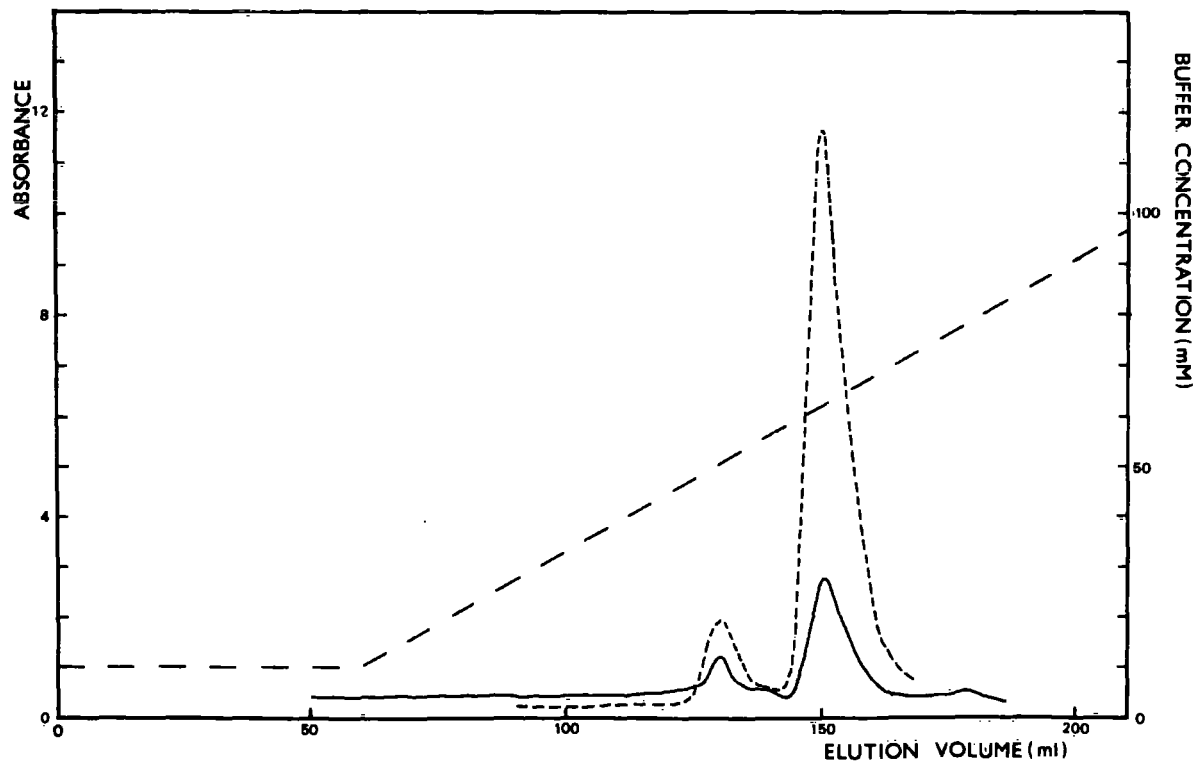
(see Fig. 5 for the notation used).

FIGURE 10. Chromatography of partially purified niger cytochrome c on CM-52 cellulose.

A 1 cm x 7 cm column of resin was used. The gradient was formed using 250 ml of each buffer. The flow rate was 20 ml/hr.

(See Fig. 5 for the notation used).


ABSORBANCE
20
16
12
8
4
0
0
100
200
300
ELUTION VOLUME(ml)


ABSORBANCE
12
8
4
0
0
50
100
150
200
ELUTION VOLUME(ml)
100
50
0
BUFFER CONCENTRATION(mM)

reducing the final yield of protein. The yields and purity of the cytochrome at the various stages in the purification are given in Table 4. The ratio $E_{550}(R)/E_{280}(ox)$ of the final preparation was 1.05, suggesting that the preparation was not totally pure. The wave-lengths of the main adsorption peaks were accurately determined on the Perkin-Elmer spectrophotometer:

α-band	(red)	550 nm
β-band	(ox)	520 nm
γ-band	(red)	415 nm
γ-band	(ox)	408 nm

4. The Purification of Black Gram Cytochrome c

Black gram seeds were germinated for 4 days prior to extraction. Filtration of the homogenate through paper was extremely slow and dry cakes were not obtained after 48 h filtration. Filtration using the centrifuge was also slow and this gave a cloudy filtrate which could not be clarified readily by repeated use of the centrifuge or by filtration on buchner funnels. Adjusting the filtrate to pH 8 produced more precipitate, and this solution was partially clarified by centrifugation. Amberlite resin in the NH_4^+ form was used to concentrate the extract and the cytochrome was eluted batchwise. The eluates from extracting 25-30 Kg of seeds, were pooled and stored at -20°. After dialysis, the total was concentrated by chromatography on CM-Sephadex.

The cytochrome was further purified by ammonium sulphate fractionation; cytochrome was present in the 50% saturation pellet, and was completely precipitated by 65% saturation. The cytochrome pellet was redissolved in a minimum volume of water and desalted by gel filtration on a column of Biogel

TABLE 5

The purification of Black Gram cytochrome c

Step	Yield mg cytochrome/ kg starting material	Purity Ratio $E_{410}(ox)/E_{280}(ox)$
Amberlite CG-50)) CM-50 Sephadex)	0.3-0.4	< 0.01
Ammonium Sulphate) Fractionation)) Biogel P-30)	0.2	0.6-1.5
CM-52 Cellulose	0.1	2.3-2.5

FIGURE 11. Chromatography of partially purified black gram cytochrome c on Biogel P-30.

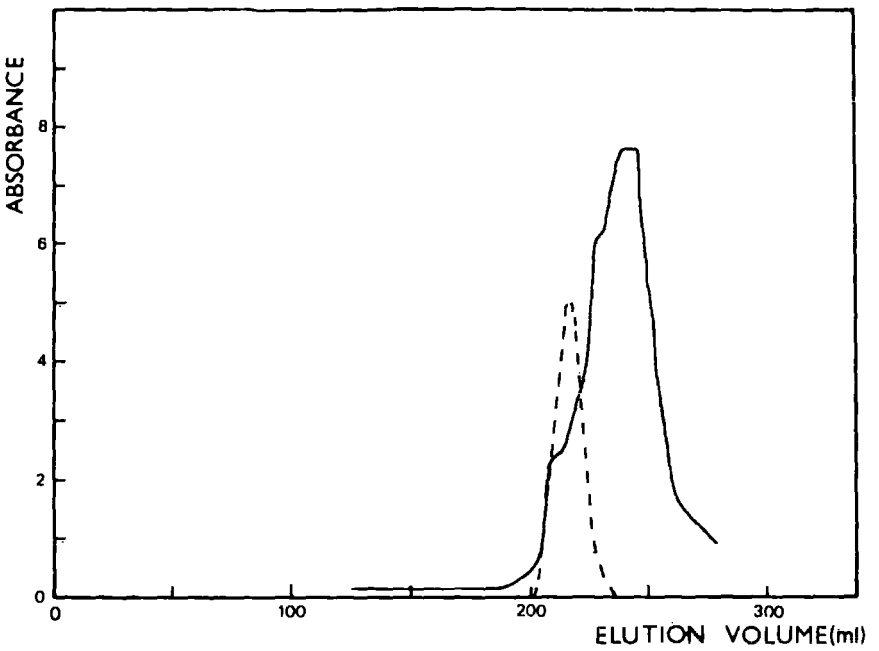
A 2 cm x 90 cm column of resin was used; it was equilibrated and eluted by water. The flow rate was 65 ml/hr.

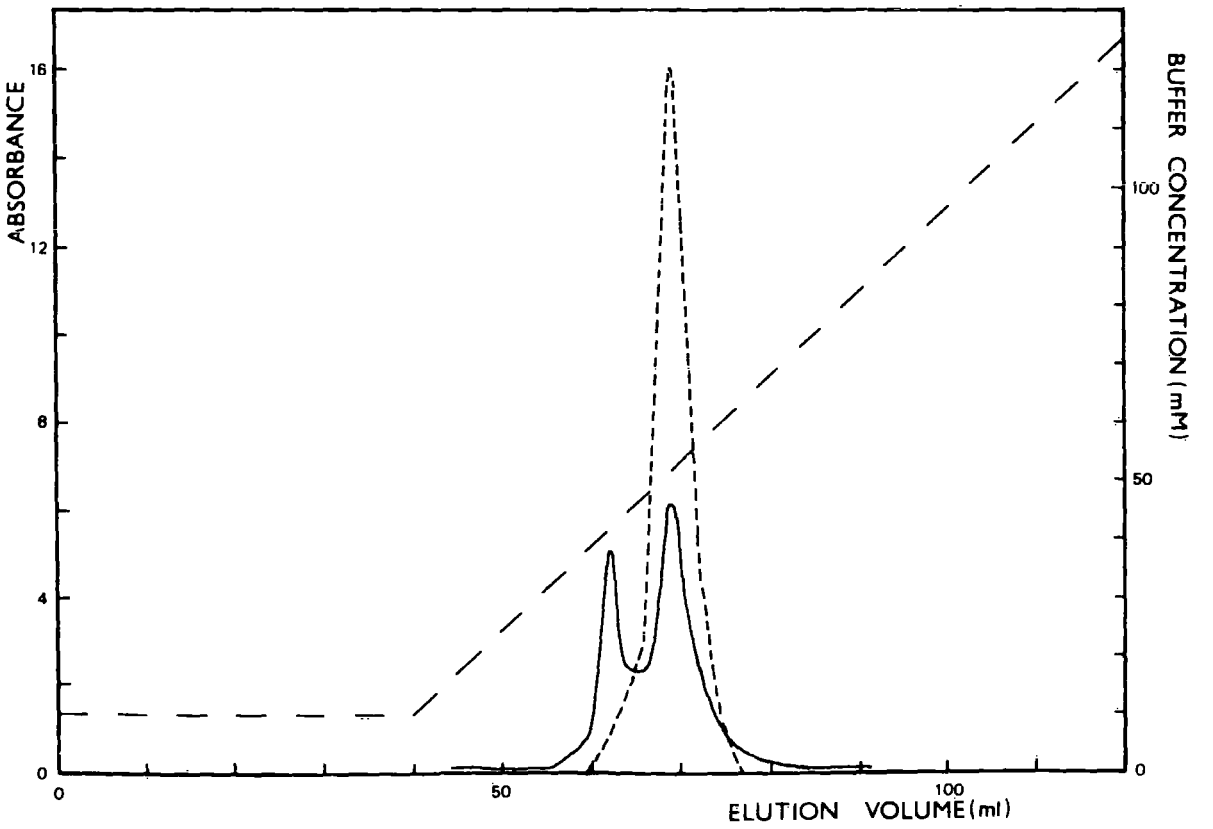
(See Fig. 5 for the notation used).

FIGURE 12. Chromatography of partially purified black gram cytochrome c on CM-52 cellulose.

A 0.8 cm x 6 cm column of resin was used. The gradient was formed using 100 ml of each buffer. The flow rate was 25 ml/hr.

(See Fig. 5 for the notation used).


ABSORBANCE
8
6
4
2
0
0
100
200
300
ELUTION VOLUME(ml)


ABSORBANCE
16
12
8
4
0
0
50
100
ELUTION VOLUME(ml)
BUFFER CONCENTRATION (mM)
100
50
0

P-30 (see Fig. 11). All fractions with an $E_{410}(ox)/E_{280}(ox)$ ratio greater than 0.6 were pooled and applied to a column of CM-52 cellulose. The cytochrome was eluted by an increasing cation gradient giving two peaks (see Fig. 12); the first peak was essentially reduced cytochrome, and the second and major peak oxidized cytochrome. All fractions from the major peak with an $E_{410}(ox)/E_{280}(ox)$ ratio greater than 2.3 were pooled.

Although the cytochrome was not completely pure, because insufficient material remained, no further purification steps were attempted. The yields and purity of the cytochrome at the various stages in the purification, are given in Table 5.

5. The Purification of Cephalosporium Cytochrome c

Cytochrome was extracted from acetone washed Cephalosporium mycelia by stirring in sodium chloride solution. Raising or lowering the pH of the solution or addition of organic solvents did not give any increase in the amount of material adsorbing at 410 nm which was extracted; this may not however, necessarily correlate with the amount of cytochrome c being extracted. Extraction of fresh cells by autolysis (Sherman, Taber & Campbell, 1965; Sherman et al., 1968), gave very low yields of cytochrome c (<5 mg/Kg), probably due to the large amount of oil associated with the cells from the medium.

Amberlite resin in the Na^+ form was used to concentrate the extract. A red band of cytochrome adsorbed to the top of the column and was eluted directly. This fraction represented only 30% of the material in the initial extract which adsorbed at 410 nm; the remainder was not bound by

the Amberlite and passed through the column. The majority of the 410 nm adsorbing material, which had passed through the Amberlite, was strongly bound on a column of DEAE-Sephadex. When eluted by 1 M-NaCl a strong red-brown band with adsorption at 560 nm was observed using the hand spectroscope. Pigments from the fungus may also be responsible for some adsorption at 410 nm in the extract (Tertzakian, Haskins, Slater & Nesbitt, 1964). The cytochrome solution eluted from the Amberlite had a strong adsorption at 260 nm, probably due to Cephalosporin c (Patterson, Van Meter & Bohonos, 1964), which masked the protein adsorption at 280 nm, so that no estimate of purity was obtained.

The material obtained from the trial and main extractions which had been stored at -20 °C was pooled and dialysed. This solution was concentrated by chromatography on CM-Sephadex and after dialysis further concentrated by rotary evaporation. It was then desalted and purified by gel-filtration on Biogel P-30 (see Fig. 13); fractions with an $E_{410}(ox)/E_{280}(ox)$ ratio greater than 2.5 were pooled. The cytochrome was adsorbed on to a column of CM-52 cellulose and eluted by an increasing cation gradient (see Fig. 14). Seven different cytochrome bands were separated; bands 1, 2 and 6 were all very small and were discarded. The material in each of the other bands was pooled (see Fig. 14), and the spectral characteristics of the major bands 4 and 7 examined. Band 7 appeared to be native cytochrome in the oxidized form; the wavelengths of the main spectral adsorption bands were determined using the Perkin Elmer spectrophotometer as, α-band 550 nm; β-band 521 nm; γ-band (red) 415 nm, and

TABLE 6

The purification of Cephalosporium cytochrome c

Step	Yield mg cytochrome/ kg starting material	Purity Ratio $E_{410}(ox)/E_{280}(ox)$
Amberlite CG-50	30-50	
CM-50 Sephadex	35	2.6
Biogel P-30	25	2.5-4.8
CM-52 Cellulose		
Band 4	10.5	4.4-5.6
Band 7	3.5	4.0-4.2

FIGURE 13. Chromatography of partially purified Cephalosporium cytochrome c on Biogel P-30.

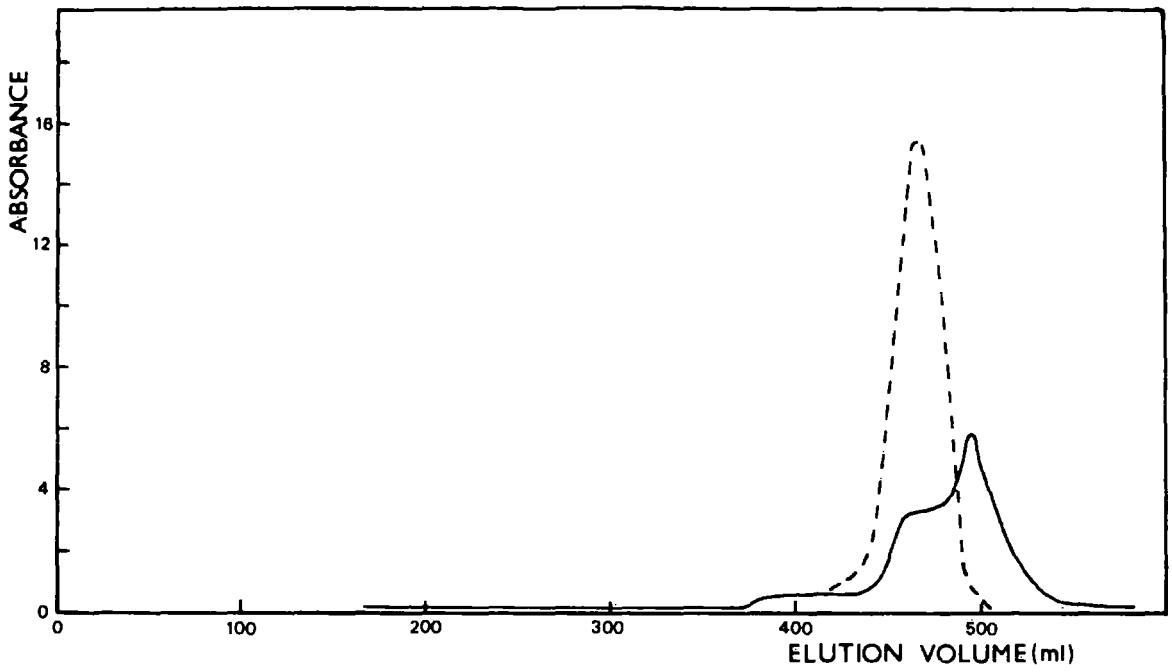
A 3 cm x 85 cm column of resin was used; it was equilibrated and eluted by water. The flow rate was 55 ml/hr.

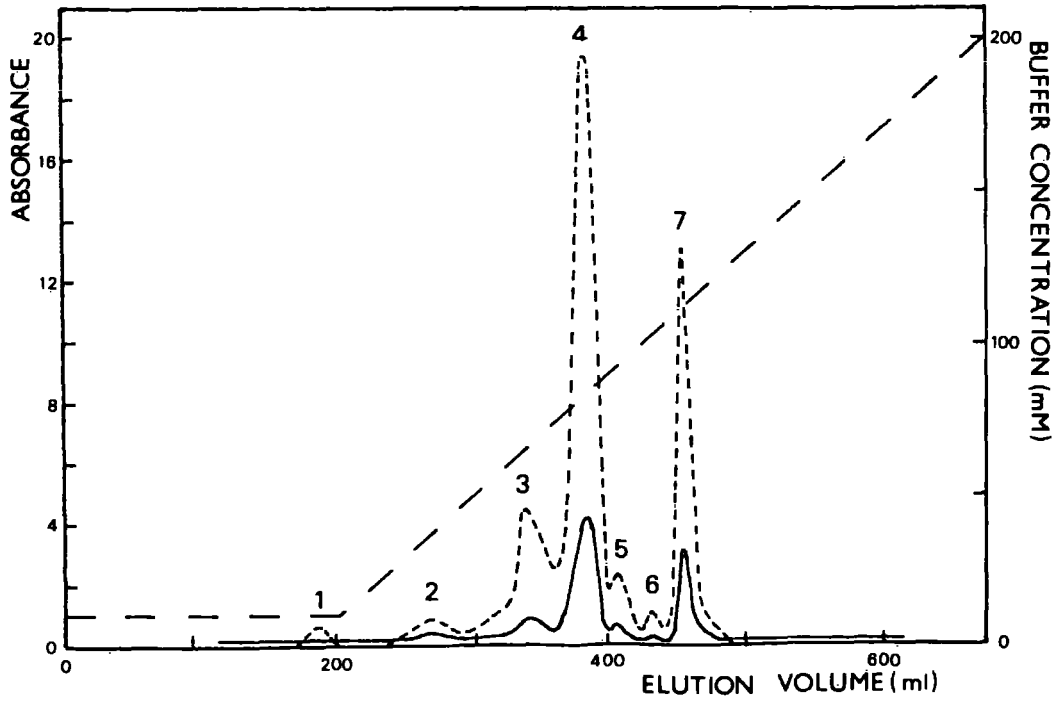
(See Fig. 5 for the notation used).

FIGURE 14. Chromatography of partially purified Cephalosporium cytochrome c on CM-52 cellulose.

A 2 cm x 12 cm column of resin was used. The gradient was formed using 300 ml of each buffer. The flow rate was 30 ml/hr.

(See Fig. 5 for the notation used).


ABSORBANCE
0
4
8
12
16
0
100
200
300
400
500
ELUTION VOLUME (ml)


ABSORBANCE
0
4
8
12
16
20
0
200
400
600
ELUTION VOLUME (ml)
BUFFER CONCENTRATION (mM)
0
100
200
1
2
3
4
5
6
7

γ-band (ox) 410 nm. Band 4, which was the major peak, appeared to be modified cytochrome in the oxidized form. It had a γ-band adsorption at 414 nm and was not reduced or spectrally altered on the addition of ascorbic acid. The cytochrome from each of these bands was lyophilised and then after re-dissolving in minimal quantities of water, desalted using Sephadex G-10. The preparations were freeze-dried and stored at -20 °C. The details of yields and purities of the cytochrome are given in Table 6.

6. The Amino Acid Sequence of Mung Bean Cytochrome c

The amino acid composition of mung bean cytochrome c is given in Table 7. The values obtained were in good agreement with those calculated from the complete sequence given in Fig. 15, although the values obtained for glycine and methionine were both high, while the values for aspartic acid and cysteine were both low.

The protein was readily denatured by ethanol. For chymotryptic digest 6.0 mg of denatured cytochrome was equilibrated at 40 °C; 2% (w/w) enzyme was added at zero time, an additional 1% (w/w) after 90 min and the digestion terminated after 120 min. For tryptic digestion 6.0 mg denatured cytochrome was equilibrated at 35 °C; 1% (w/w) enzyme was added at zero time, an additional 1% (w/w) after 60 min and the digestion terminated after 100 min. Separate digestions were used to provide peptides for amino acid sequence and composition analyses.

The results of the sequence analysis on chymotryptic peptides are shown in Table 8, and the peptide amino acid

TABLE 7

The amino acid composition of mung bean cytochrome c

	Mean values 20 h hydrolysis	Amino acid analysis (Integral values)	Sequence values
Asp	9.8	10	11
Thr	7.6	8	8
Ser	6.3	7	7
Glu	8.6	9	9
Pro	6.5	7	7
Gly	12.1	12	11
Ala	9.9	10	10
Val	3.2	3	3
Cysteine	1.2	1	2
Met	2.6	3	2
Ile	3.4	3	4
Leu	8.0	8	8
Tyr	5.9	6	6
Phe	4.3	4	4
Me_3Lys	1.9	2	2
Lys	11.5	12	12
His	2.2	2	2
Arg	2.3	2	2
Trp	*	*	1

The 'mean values' were obtained from analysis of seven samples of 0.1-0.4 mg cytochrome.

No correction was made for destruction of certain amino acids.

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

TABLE 8

Mung Bean Chymotryptic Peptides

Peptide	Mobility at pH 6.5	Mobility at pH 1.9	Dansyl-Edman Results
C1	-1.50	0	Does not react
C2	-0.65		Asx- Digested with papain for 5 h.
C2P1		0.30	Asx-Glx (Ala,Pro,Pro,Gly)
C2P2		1.10	Asx-Ser (Lys)
C2P3		0.70	Ser-Gly (Glx,Lys,Ile,Phe)
C3	0.70	0.84	Lys-Thr-Lys-$CySO_3$-Ala-Glx-$CySO_3$-His
C4	0.81	0.74	Lys-Glx-Gly-Pro-Asx-Leu-Asx
C5	0	0.58	Gly-Leu-Phe
C6	0.70	0.61	Gly-Arg-Glx-Ser-Gly-Thr-Thr-Ala-Gly-Tyr
C7	0	0.66	Ser-Tyr
C8	0.87	0.70	Ser-
C9	0	0.51	Ala-Val-Ile-Trp
C10	-0.91		Glx-Glx-Lys-Thr-Leu-Tyr
C11(a)	-1.65	0.50	Asx-Tyr
C11(b)	-1.37	0.47	Asx-Tyr-Leu
C11(c)	1.53		Leu-Leu-Asx-Pro-Me_3Lys-Lys-Tyr
C11(d)	1.68		Leu-Asx-Pro-Me_3Lys-Lys-Tyr
C12	1.02		Ile-Pro-Gly-Thr-Lys-Met
C13	0	0.47	Val-Phe-Pro-Gly-Leu
C14	0.70	1.01	Me_3Lys-Lys-Pro-Glx-Asx-(Arg)-Ala-Asx-Leu
C15	0	0.55	Ile-Ala-Tyr
C16	0	0.82	Leu-Lys-Glx-Ser-Thr-Ala

TABLE 9

The Amino Acid Analysis of Mung Bean Chymotryptic Peptides

Peptide	Composition (molar ratios)
C1	Ser 1.0(1), Ala 1.0(1), Phe 0.9(1)
C2	Not analysed
C3	$CySO_3$ 2.5(2), Asp 1.3(1), Thr 0.3(2), Glu 1.1(1), Gly 2.2(2), Ala 1.6(2), Val 0.8(1), Lys 2.9(3), His 1.8(2)
C4	Asp 2.2(2), Glu 1.0(1), Pro 0.8(1), Gly 1.2(1), Leu 0.9(1), Lys 1.1(1)
C5	Gly 1.1(1), Leu 1.0(1), Phe 0.8(1)
C6	Thr 1.6(2), Ser 0.8(1), Glu 0.9(1), Gly 3.3(3), Ala 0.8(1), Tyr 0.9(1), Arg 1.0(1)
C7	Ser 1.0(1), Tyr 0.8(1)
C8	Not analysed
C9	Not analysed
C10	Thr 0.8(1), Glu 1.8(2), Leu 1.0(1), Tyr 0.7(1), Lys 1.0(1)
C11(a)	Asp 1.0(1), Tyr 0.7(1)
C11(b)	Asp 1.3(1), Leu 0.8(1), Tyr 0.7(1)
C11(d)	Asp 1.0(1), Pro 0.6(1), Leu 1.1(1), Tyr 0.5(1), Me_3Lys 0.9(1), Lys 1.3(1)
C12	Thr 0.8(1), Pro 0.7(1), Gly 1.1(1), Met 0.5(1), Ile 0.6(1), Lys 1.0(1)
C13	Pro 0.8(1), Gly 1.0(1), Val 1.0(1), Leu 1.3(1), Phe 0.8(1)
C14	Asp 2.0(2), Glu 1.2(1), Pro 0.8(1), Ala 0.7(1), Leu 1.1(1), Me_3Lys 0.8(1), Lys 1.1(1), Arg 1.1(1)
C15	Ala 0.9(1), Ile 0.6(1), Tyr 0.3(1)
C16	Thr 0.9(1), Ser 0.8(1), Glu 1.0(1), Ala 1.0(1), Leu 1.0(1), Lys 1.1(1)

Values given in parentheses are the compositions finally deduced from the sequence.

Amino acids found in molar ratios of less than 0.3 are ignored.

TABLE 10

Mung Bean Tryptic Peptides

Peptide	Mobility at pH 6.5	Mobility at pH 1.9	Dansyl-Edman Results
T1	-1.24		Does not react; digested with chymotrypsin for 2 h.
T1C1			(Acetyl-Ala-Ser-Phe)
T1C2			Asx-Glx-Ala-Pro-Pro-Gly-Asx-Ser-
T2	0	0.87	Ser-Gly-Glx-Lys
T3	1.38		Ile-Phe-Lys
T4	1.75	1.45	Thr-Lys
T5	0	0.28	$CySO_3$-Ala-Glx-$CySO_3$-His-Thr-Val-Asx-Lys
T6	2.20		Gly-Ala-Gly-His-Lys
T7	*	0.47	Glx-Gly-Pro-Asx-Leu-Asx-
T8	*	0.39	Digested with chymotrypsin
T8C1		0.30	Glx-
T8C2		0.66	Ser-Tyr
T8C3		0.92	Ser-Thr-Ala-Asx-Lys
T9	-0.55		Asx-Met-Ala-Val-Ile-
T10	*	0.66	Thr-Leu-Tyr-Asx-
T11	1.00		Tyr-Ile-Pro-Gly-Thr-Lys
T12	1.15		Met-
T13	0	0.56	Ala-Asx-Leu-Ile-Ala-Tyr-Leu-Lys
T13(a)	-1.03		Ala-Asx-Leu-Ile-Ala-Tyr
T13(b)	1.70	1.39	Leu-Lys
T14	-1.45		Glx-Ser-Thr-Ala

*Peptides T7, T8 and T10 were not resolved at pH 6.5. A zone of mobility 0.50-0.65 was eluted.

TABLE 11

The Amino Acid Analysis of Mung Bean Tryptic Peptides

Peptide	Composition (molar ratios)
T1	Asp 2.2(2), Ser 1.6(2), Glu 1.2(1), Pro 2.2(2), Gly 1.2(1), Ala 2.4(2), Phe 1.1(1), Lys 1.1(1).
T2*	Ser (1), Glu (1), Gly (1), Lys (1).
T3	Ile 0.7(1), Phe 1.0(1), Lys 1.0(1).
T4	Thr 1.0(1), Lys 1.0(1).
T5*	$CySO_3$ (2), Asp (1), Thr (1), Glu (1), Ala (1), Val (1), Lys (1), His (1).
T6	Gly 1.9(2), Ala 0.9(1), His 1.0(1), Lys 1.0(1).
T7	Asp 1.9(2), Glu 0.8(1), Pro 0.8(1), Gly 3.1(3), Leu 2.0(2), Phe 1.0(1), Arg 1.0(1).
T8C1	Thr 1.8(2), Ser 0.8(1), Glu 1.0(1), Gly 2.1(2), Ala 0.9(1), Tyr 0.6(1).
T8C2*	Ser (1), Tyr (1)
T8C3*	Ala (1), Asp (1), Lys (1), Ser (1), Thr (1).
T9	Asp 1.0(1), Glu 2.2(2), Ala 0.8(1), Val 0.9(1), Met 1.0(1), Ile 0.8(1), Lys 1.3(1).
T10	Asp 2.1(2), Thr 0.8(1), Pro 1.0(1), Leu 3.3(3), Tyr 0.5(2), Me_3Lys 1.1(1), Lys 1.1(1).
T11	Thr 0.8(1), Pro 0.7(1), Gly 1.2(1), Ile 0.7(1), Tyr 0.6(1), Lys 1.1(1).
T12	Asp 1.0(1), Glu 1.0(1), Pro 1.7(2), Gly 1.5(1), Val 0.8(1), Met 0.5(1), Leu 1.0(1), Phe 1.0(1), Me_3Lys 0.6(1), Lys 0.9(1), Arg 1.3(1).
T13	Asp 1.0(1), Ala 1.8(2), Ile 0.7(1), Leu 1.9(2), Tyr 0.8(1), Lys 1.1(1).
T14	Thr 0.9(1), Ser 0.9(1), Glu 1.0(1), Ala 0.9(1).

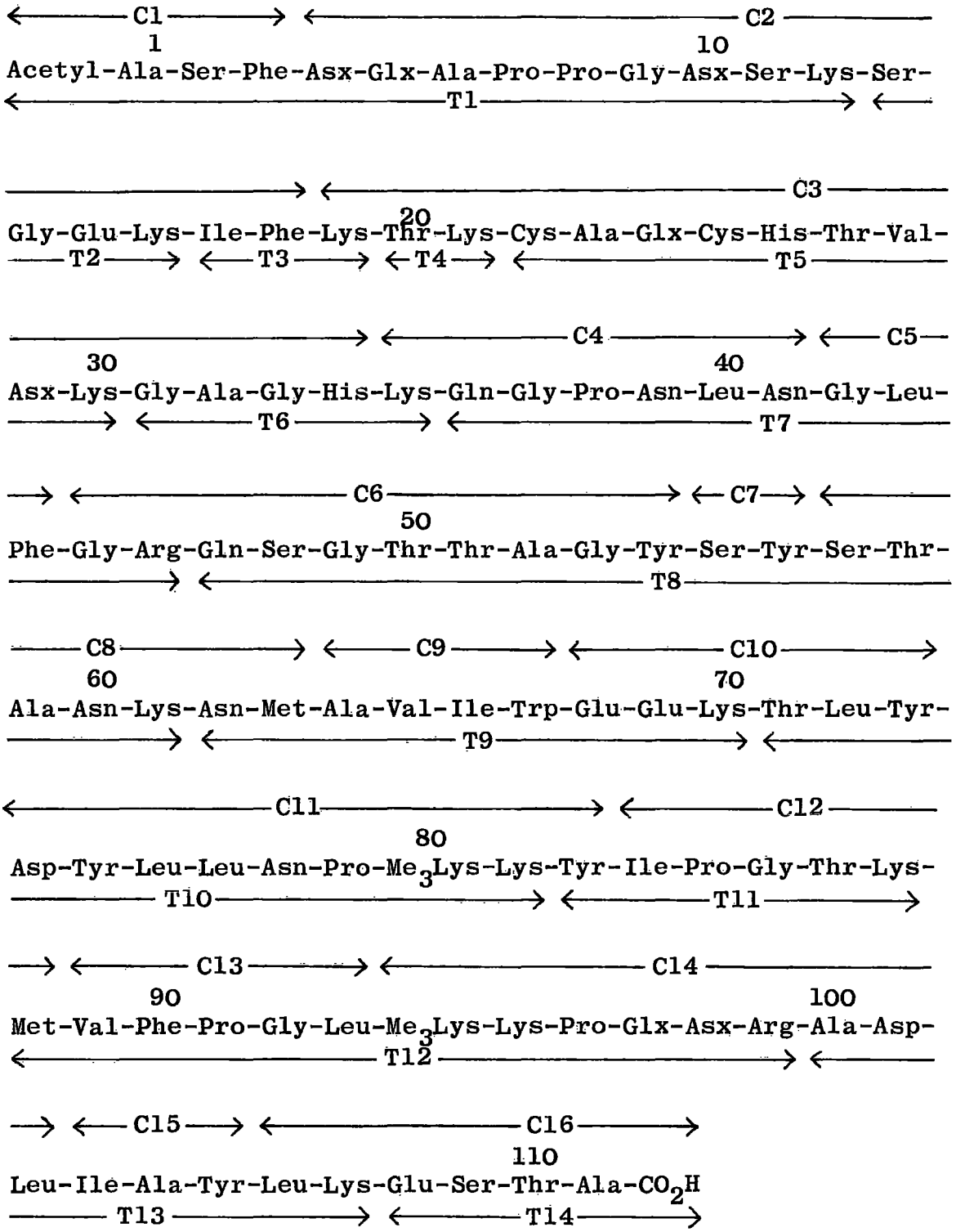
Values given in parentheses are the compositions finally deduced from the sequence.

Amino acids found in molar ratios of less than 0.3 are ignored.

*Composition shown by hydrolysis plus 'dansylation'.

FIGURE 15

The amino acid sequence of mung bean cytochrome c



analyses in Table 9. The sequence analysis results for the tryptic peptides are shown in Table 10, and the amino acid analyses in Table 11. The solvents used for separating dansyl amino acids were solvents 1A, 2A and 3A/3B.

Peptide C1. (Acetyl-Ala-Ser-Phe)

This peptide was ninhydrin negative and was detected by the starch-iodide method between peptides C11(a) and C11(b). It was purified as the only peptide neutral at pH 1.9, suggesting that the N-terminal amino group was blocked. Carboxypeptidase A digestion released phenylalanine and small amounts of serine after 3 h and equal amounts of phenylalanine and serine after 24 h.

Peptide C2. (Asx-Glx-Ala-Pro-Pro-Gly-Asx-Ser-Lys-Ser-Gly-Glu-Lys-Ile-Phe)

This peptide was found in low yields, probably due to incomplete cleavage at phenylalanine-3, and had aspartyl as the N-terminus. Papain digestion gave three fragments, which were separated at pH 1.9. Peptides T1, T2 and T3 were used to confirm the sequence. There was insufficient material for an accurate amino acid analysis.

Peptide C3. (Lys-Thr-Lys-Cys-Ala-Glx-Cys-His-Thr-Val-Asx-Lys-Gly-Ala-Gly-His)

The haem peptide was obtained by precipitating at pH 4.5 and elution from the paper by 20% (v/v) pyridine. The haem prosthetic group was removed by the mercuric chloride method. The complete sequence was obtained in conjunction with peptides T5 and T6. The amino acid analysis gave a low threonine value, possibly due to oxidation by performic acid during the removal of the haem. 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-Gln-Gly-Pro-Asn-Leu-Asn)

This gave a weak positive reaction with the Ehrlich reagent, but no tryptophan could be found in the peptide. The mobility at pH 6.5 showed a charge of 1 for a peptide of this molecular weight, which indicates that residues 36, 39 and 41 are all amides.

Peptide C5. (Gly-Leu-Phe)

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

Peptide C6. (Gly-Arg-Gln-Ser-Gly-Thr-Thr-Ala-Gly-Tyr)

This peptide was yellow with cadmium-ninhydrin reagent and was Sakaguchi positive. As the amino acid composition gave no lysine in the peptide, dansyl-arginine was observed with no masking by ϵ-dansyl-lysine, which co-chromatographed with dansyl-arginine in the solvents used. A charge of +1 was indicated by the mobility at pH 6.5, so residue 47 was glutamine.

Peptide C7. (Ser-Tyr)

Peptide C8. (Ser-Thr-Ala-Asn-Lys-Asn-Met)

This was platinic iodide positive and was only partially separated from peptide C5, which had almost identical mobilities at both pH 6.5 and pH 1.9 and occurred in much greater yield. N-terminal serine was found. Tryptic peptides T8C3 and T9 were used to obtain the sequence.

Peptide C9. (Ala-Val-Ile-Trp)

This was neutral at pH 6.5. It was Ehrlich positive and was not obtained in sufficient yield to allow an accurate amino acid analysis. Carboxypeptidase A digestion revealed tryptophan as the C-terminus.

Peptide C10. (Glu-Glu-Lys-Thr-Leu-Tyr)

The mobility at pH 6.5 indicated a charge of -1, which means that both residues 68 and 69 were glutamic acid.

Peptide C11. (Asp-Tyr-Leu-Leu-Asn-Pro-Me_3Lys-Lys-Tyr)

The mobilities at pH 6.5 of peptides C11(a) and C11(b) indicated that residue 74 was aspartic acid, and of peptides C11(c) and C11(d) that residue 78 was asparagine.

Peptides C11(c) and C11(d) were each shown to contain one residue of both ϵ-N-trimethyllysine and lysine.

Bis-dansyl-lysine was observed for position 81; α-N-dansyl-ϵ-N-trimethyllysine co-chromatographed with ϵ-dansyl-lysine and hence ϵ-N-trimethyllysine was not observed directly, but was placed from the amino acid analysis.

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

This peptide was platinic iodide positive.

Peptide C13. (Val-Phe-Pro-Gly-Leu)

Peptide C14. (Me_3Lys-Lys-Pro-Glx-Asx-Arg-Ala-Asp-Leu)

This peptide was Sakaguchi positive. The N-terminus gave a strong ϵ-dansyl-lysine,α-dansyl——ϵ-N-trimethyllysine or dansyl-arginine, but no other amino acid, and amino acid analysis showed that the peptide contained lysine, ϵ-N-trimethyl-lysine and arginine. Lysine was shown in position 95 by

di-dansyl-lysine, but arginine and ϵ-N-trimethyllysine could not be placed. The mobility at pH 6.5 indicated a charge of 1, showing two of the positions 89, 90 and 93 were amides.

Peptide C15. (Ile-Ala-Tyr)

Peptide C16. (Leu-Lys-Glu-Ser-Thr-Ala)

As the peptide was neutral at pH 6.5, position 108 was glutamic acid. The observed C-terminal amino acid, alanine, is not normally susceptible to chymotryptic cleavage, indicating that this is the C-terminal peptide for the protein.

Peptide T1. (Acetyl-Ala-Ser-Phe-Asx-Glx-Ala-Pro-Pro-Gly-Asx-Ser-Lys)

This ninhydrin positive peptide gave no N-terminus by the dansyl method. It was digested with chymotrypsin; since peptide T1C1 does not react in the dansyl-Edman the two fragments were not separated. The electrophoretic mobility at pH 6.5 indicated the presence of one amide residue which cannot be placed.

Peptide T2. (Ser-Gly-Glu-Lys)

Peptide T3. (Ile-Phe-Lys)

Peptide T4. (Thr-Lys)

Peptide T5. (Cys-Ala-Glx-Cys-His-Thr-Val-Asx-Lys)

The haem peptide was obtained by combining material precipitated at pH 4.5 with that from eluting the paper by 20% (v/v) pyridine. The peptide was recovered in low yield after removal of the haem prosthetic group by the mercuric

chloride method. The electrophoretic mobilities could not be used, because of the cysteic acid and histidine residues, for deduction of the amide content.

Peptide T6. (Gly-Ala-Gly-His-Lys)

Peptide T7. (Gln-Gly-Pro-Asn-Leu-Asn-Gly-Leu-Phe-Gly-Arg)

Insufficient peptide was available for complete analysis and peptides C5 and C6 give the complete sequences. The electrophoretic mobility at pH 6.5 indicates that positions 36, 39 and 41 are all amides.

Peptide T8. (Gln-Ser-Gly-Thr-Thr-Ala-Gly-Tyr-Ser-Tyr-Ser-Thr-Ala-Asn-Lys)

This peptide was digested with α-chymotrypsin. This gave three peptides, which were separated at pH 1.9. These results, together with evidence from the sequences of chymotryptic peptides gave the sequence of this peptide. The mobility at pH 6.5 indicated that positions 47 and 60 were both amides.

Peptide T9. (Asn-Met-Ala-Val-Ile-Trp-Glu-Glu-Lys)

This was strongly Ehrlich positive and was also pink in the trifluoroacetic acid stage of the first Edman degradation (Uphaus, Grossweiner, Katz & Kopple, 1959), indicating the presence of tryptophan. The peptide was also platinic iodide-positive. The mobility at pH 6.5 indicates a charge of -1 and, as positions 68 and 69 were both glutamic acid (see peptide C10), position 62 was asparagine.

Peptide T10. (Thr-Leu-Tyr-Asp-Tyr-Leu-Leu-Asn-Pro-Me_3Lys-Lys)

Amino acid analysis and three Edman degradations, together with chymotryptic peptide C11, gave the sequence for this

peptide. The mobility at pH 6.5 indicated one amide and one acid residue. The former was placed in position 78 by peptides C11(c) and C11(d), and the latter in position 74 by peptides C11(a) and C11(b).

Peptide T11. (Tyr-Ile-Pro-Gly-Thr-Lys)

Peptide T12. (Met-Val-Phe-Pro-Gly-Leu-Me_3Lys-Lys-Pro-Glx-Asx-Arg)

This was both Sakaguchi and platinic iodide positive, and the N-terminus was methionine. As with peptide C14, the mobility at pH 6.5 indicated one acidic and one amide residue but these could not be placed.

Peptide T13. (Ala-Asp-Leu-Ile-Ala-Tyr-Leu-Lys)

Peptides T13(a) and T13(b) were the result of chymotryptic activity causing a split at tyrosine -105. This was the only evidence of any chymotryptic activity in the tryptic digest. Position 101 is aspartic acid as peptide T13 was neutral at pH 6.5.

Peptide T14. (Glu-Ser-Thr-Ala)

This peptide was acidic, the mobility showing glutamic acid in position 108. No ϵ-dansyl-lysine was observed in the 'dansyl' method, so this was considered to be the C-terminal peptide in the protein as it was not Sakaguchi-positive.

The overlapping chymotryptic and tryptic peptides give the sequence of mung bean cytochrome c given in Fig. 15; this is supported by the determined peptide and protein amino acid compositions. All residues were identified during the

sequence analysis except for ε-N-trimethyllysine-80, ε-N-trimethyllysine-94 and arginine-99, which were identified from peptide compositions. The order given for the latter two residues in the sequence was based on homology with other cytochromes c (Dayhoff, 1969). The nature of the N-terminal blocking group was not determined and is assumed to be an acetyl group by homology with plant and other cytochromes c.

Two additional peptides were identified during the analysis, which did not belong to the complete sequence. These were Gly-Met-Lys and Lys-Ile-Phe. As the spectral ratios of the cytochrome used suggested that it was not totally pure, these peptides may have been derived from the impurity present.

The electrophoretic mobilities of the peptides at pH 6.5 (Table 8 and Table 10), indicated the presence of 10 amide residues in the sequence. The presence of one amide in the haem peptide was established by comparison of the haem peptide mobilities with those of horse haem peptides of known amide content (Margoliash et al., 1961). Of the amide residues, 7 may be placed directly; these are asparagine in positions 39, 41, 60, 62 and 78, and glutamine in positions 36 and 47. Homology with other cytochromes c (Dayhoff, 1969) suggests that positions 24 and 97 are both glutamine and homology with plant cytochromes c that residue 10 is asparagine.

The observed enzyme specificities were consistent with those expected (Smyth, 1967). Chymotrypsin also gave cleavage at both methionine residues, histidine-34, asparagine-41, and leucine residues 76, 93 and 102. There was no tryptic digestion at either ε-N-trimethyllysine residue but tryptic

breaking at tyrosine-105 did occur.

7. The Amino Acid Sequence of Sunflower Cytochrome c

Sunflower cytochrome c was readily denatured by ethanol and 6 mg was digested at 40 °C with chymotrypsin; 2% (w/w) enzyme was added at zero time, an additional 2% (w/w) after 100 min and the digestion terminated after 150 min. Separate digests were used to obtain peptides for sequence analysis and for amino acid composition determination. The results of the sequence analysis are shown in Table 12, and the quantitative amino acid analyses in Table 13. The solvents used in separating dansyl amino acids were solvents 1A, 2A and 3C.

Peptide C1. (Acetyl-Ala-Ser-Phe)

This was ninhydrin negative, but total analysis by the dansyl method gave the composition as (Ala,Ser,Phe). Carboxypeptidase A digestion released phenylalanine and traces of serine after 3 h and equal amounts of phenylalanine and serine after 18 h. The N-terminal block was determined as an acetyl group.

Peptide C2. (Ala-Glu-Ala-Pro-Ala-Gly-Asp-Pro-Thr-Thr-Gly-Ala-Lys-Ile-Phe).

This gave an N-terminal alanine. The peptide was then digested with papain and only two major peptides were located after electrophoresis at pH 1.9. The original peptide had a charge of -1 at pH 6.5, indicating that residues 5 and 10 are both acidic. Insufficient material was available for an amino acid analysis of this peptide.

TABLE 12

Sunflower Chymotryptic Peptides

Peptide	Mobility at pH 6.5	Mobility at pH 1.9	Dansyl-Edman Results
C1	-1.50	0	Does not react. (Acetyl-Ala) Ser-Phe
C2	-0.67		Ala- Digested with papain for 2 h.
C2 P1		0.47	Ala-Glx-Ala-Pro-Ala-Gly
C2 P2		0.52	Asx-Pro-Thr-Thr-Gly-Ala-Lys-Ile-Phe
C3	0.54		Lys- Performic oxidized and digested with trypsin for 1 h.
C3 T1		1.61	Lys-Thr-Lys
C3 T2		1.36	Thr-Lys
C3 T3		0.24	$CySO_3$-Ala-Glx-$CySO_3$-His-Thr-Val-Glx-Lys
C3 T4		1.14	Gly-Ala-Gly-His
C4	0.86	0.67	Lys-Glx-Gly-Pro-Asx-Leu-Asx
C5	0	0.60	Gly-Leu-Phe
C6	0.65	0.60	Gly-Arg-Glx-Ser-Gly-Thr-Thr-Ala-Gly-Tyr
C7	0	0.63	Ser-Tyr
C8(a)	0.90	0.73	Ser-Ala-Ala-Asx-Lys-Asx-Met
C9	0	0.46	Ala-Val-Ile-
C10	-1.78	0.32	Glx-Glx-Asx-Thr-Leu-Tyr
C11	-1.35	0.44	Asx-Tyr-Leu
C11(a)	-1.65	0.49	Asx-Tyr
C12	1.70	1.02	Leu-Asx-Pro-Me_3Lys-Lys-Tyr
C13	1.02	0.78	Ile-Pro-Gly-Thr-Lys-Met
C14	0	0.42	Val-Phe-Pro-Gly-Leu
C15	1.70	1.31	Me_3Lys-Lys-Pro-Glx-Glx-Arg
C16	-0.98		Ala-Asx-Leu-Ile-Ala-Tyr
C17	1.02	0.84	Leu-Lys-Thr-Ser-Thr-Ala

TABLE 13

The Amino Acid Analysis of Sunflower Chymotryptic Peptides

Peptide	Composition (molar ratios)
C1	Ser 0.85(1), Ala 1.12(1), Phe 1.00(1),
C2	Not analysed.
C3	Thr 2.06(2), Glu 2.24(2), Gly * (2), Ala 1.92(2), Val 0.98(1), Cys 0.83(2), Lys * (3), His 2.00(2).
C4	Asp 1.87(2), Glu 1.13(1), Pro * (1), Gly 1.25(1), Leu 1.00(1), Lys 1.32(1).
C5	Gly 1.10(1), Leu 1.00(1), Phe 0.94(1).
C6	Thr 1.92(2), Ser 0.90(1), Glu 1.04(1), Gly * (3), Ala 0.97(1), Tyr 0.92(1), Arg 1.00(1).
C7	Ser 1.00(1), Tyr 0.79(1).
C8(b)	Asp 0.98(1), Ser 1.00(1), Ala 1.64(2), Lys 1.10(1).
C10	Asp 0.95(1), Thr 0.68(1), Glu 1.76(2), Leu 1.00(1), Tyr * (1).
C11(a)	Asp 1.00(1), Tyr 0.65(1).
C12	Asp 1.03(1), Pro 0.97(1), Leu 1.00(1), Tyr 0.91(1), Me_3Lys 0.96(1), Lys 1.05(1).
C13	Thr 0.85(1), Pro 0.73(1), Gly 1.25(1), Met 1.00(1), Ile 0.76(1), Lys 1.02(1).
C14	Pro 1.09(1), Gly 1.06(1), Val 0.73(1), Leu 1.00(1), Phe 0.95(1).
C15	Glu 1.92(2), Pro 1.00(1), Me_3Lys 0.76(1), Lys * (1), Arg 1.24(1).
C16	Asp 1.00(1), Ala 1.22(2), Ile 0.79(1), Leu 1.23(1), Tyr * (1).
C17	Thr 1.66(2), Ser 0.89(1), Ala 0.94(1), Leu 1.00(1), Lys 1.02(1).

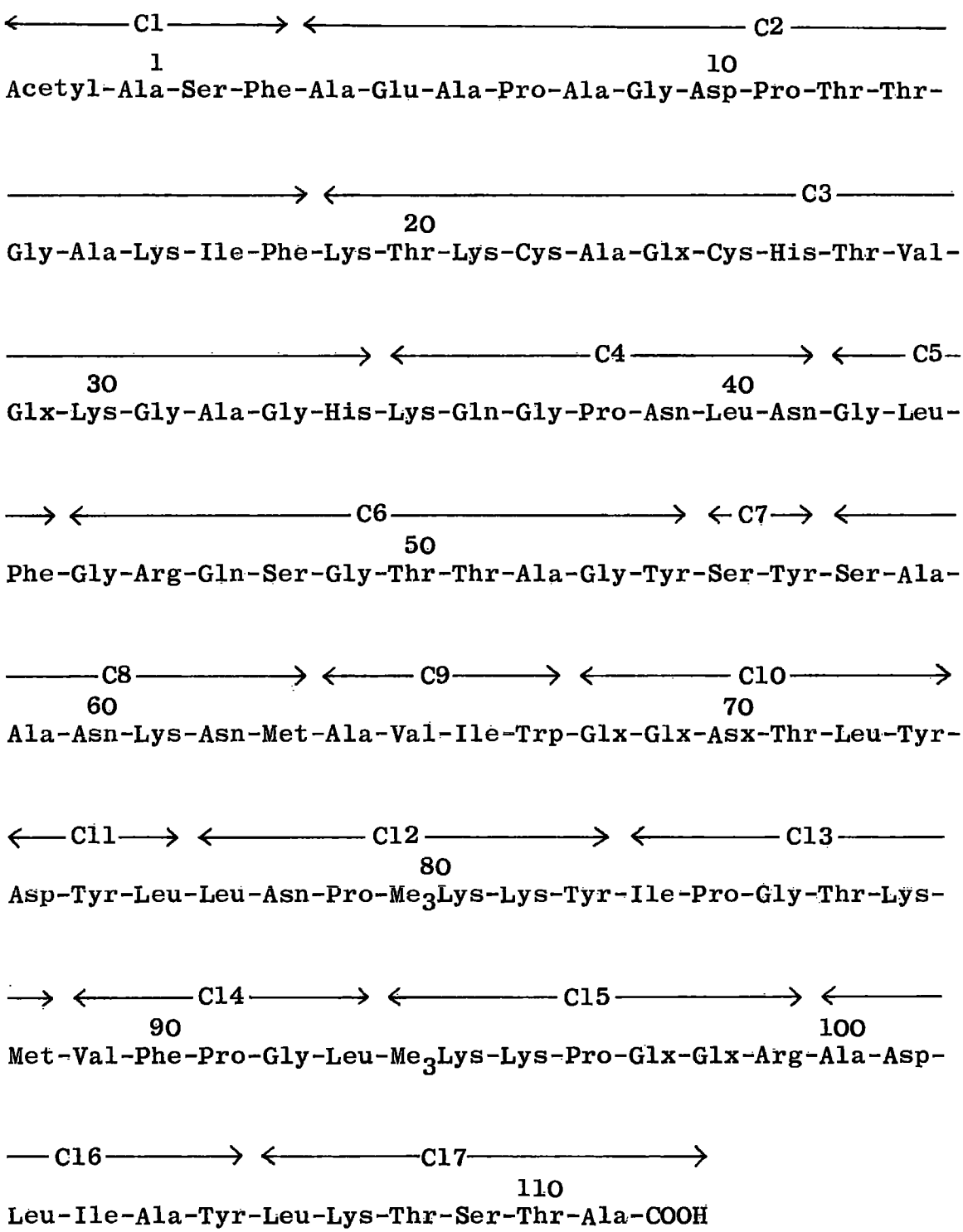
Values given in parentheses are the compositions deduced from the sequence.

Amino acids found in molar ratios of less than 0.25 are ignored.

*Amino acid present but the peak unsuitable for integration.

FIGURE 16.

The amino acid sequence of sunflower cytochrome c



Peptide C3. (Lys-Thr-Lys-Cys-Ala-Glx-Cys-His-Thr-Val-Glx-Lys-Gly-Ala-Gly-His)

The haem peptide was eluted by 20% (v/v) pyridine and gave an N-terminal lysine. It was digested with trypsin and then performic oxidized. After electrophoresis at pH 1.9, four major peptides were located, together with free lysine. The amide content could not be inferred directly because of the histidine and cysteic acid residues present.

Peptide C4. (Lys-Gln-Gly-Pro-Asn-Leu-Asn).

The charge at pH 6.5 of +1 indicated that residues 36, 39 and 41 are all amides. A peptide due to partial cleavage at asparagine-41 was also found in reasonable yield.

Peptide C5. (Gly-Leu-Phe)

Peptide C6. (Gly-Arg-Gln-Ser-Gly-Thr-Thr-Ala-Gly-Tyr)

A charge of +1 at pH 6.5 indicated that residue 47 was glutamine.

Peptide C7. (Ser-Tyr)

Peptide C8. (Ser-Ala-Ala-Asn-Lys-Asn-Met)

The peptide was difficult to separate from peptide C13 and was recovered in only low yield. Sufficient was available for dansyl-Edman analysis only. A peptide resulting from partial cleavage of residue lysine-61 was also located and this was used to determine the amino acid composition on the automatic analyser. The charge of both these peptides at pH 6.5, +1, indicated that both residues 60 and 62 were amides.

Peptide C9. (Ala-Val-Ile-Trp)

This was the only Ehrlich-positive peptide, and gave a

pink colour in the trifluoroacetic acid stage of the first Edman degradation, indicating tryptophan (Uphaus et al., 1969). Sufficient peptide could not be separated from peptide C14 for an amino acid analysis.

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

The charge of -2 at pH 6.5 indicates that one of the three acidic residues is in the amide form.

Peptide C11. (Asp-Tyr-Leu).

The charge of -1 at pH 6.5 indicates residue 74 is aspartic acid. The peptide breaking after tyrosine-75 was obtained in major yield.

Peptide C12. (Leu-Asn-Pro-Me_3Lys-Lys-Tyr)

The charge of +2 at pH 6.5 indicates residue 78 is asparagine.

Peptide C13. (Ile-Pro-Gly-Thr-Lys-Met)

This peptide was strongly platinic iodide-positive.

Peptide C14. (Val-Phe-Pro-Gly-Leu)

Peptide C15. (Me_3Lys-Lys-Pro-Glx-Glx-Arg)

The charge of +2 at pH 6.5 indicates one amide residue, which cannot be placed. The break at arginine-99 was the major break although a peptide in minor yield breaking at leucine-102 was also found.

Peptide C16. (Ala-Asp-Leu-Ile-Ala-Tyr)

This charge of -1 at pH 6.5 indicates residue 101 is aspartic acid. A low yield of alanine in the peptide analysis was probably due to incomplete hydrolysis of the Ile-Ala bond.

A peptide Ile-Ala-Tyr was also found in low yield.

Peptide C17. (Leu-Lys-Thr-Ser-Thr-Ala)

This peptide was suspected to be the C-terminal peptide of the protein since alanine is not a residue normally susceptible to chymotryptic cleavage.

The sequence of sunflower cytochrome c is given in Fig. 16. As only chymotryptic peptides were examined, no overlap peptides exist to enable the order of the peptides to be deduced directly. Many of the peptides were identical to those examined during the sequence determinations on other plant cytochromes; the remainder differed only by single amino acid substitutions. Thus, it was possible to align the peptides with the other plant cytochrome c sequences and derive the complete sequence. No additional chymotryptic peptides were found to those giving the sequence. A small quantity of core material from the digest remained on the loading origin after pH 6.5 electrophoresis.

The electrophoretic mobilities of the peptides at pH 6.5 (Table 12) indicated the presence of 10 amide residues. The presence of one amide in the haem peptide was established by comparison of the mobility of the haem peptide with the mobility of horse chymotryptic haem peptide for which the amide content is known (Margoliash et al., 1961). Of these amide residues, 7 can be unambiguously placed; these are glutamine in positions 36 and 47, and asparagine in positions 39, 41, 60, 62 and 78. Homology with other cytochromes (Dayhoff, 1969) suggest that the remaining amide residues in peptides C3, C10 and C15 are glutamine-24, asparagine-62

and glutamine-89.

The observed chymotryptic specificity was as expected (Smyth, 1967), except digestion at tyrosine-75 was incomplete. Cleavage also occurred at both methionine residues, leucine residues 76 and 93, and at histidine-34. Partial splits were observed at asparagine-41 and leucine-102. A major tryptic type specificity was observed at arginine-99, and to a lesser extent at lysine-61.

Amino acid composition determinations gave results with abnormally high glycine values, indicating the presence of an impurity in the preparation. This meant that a reliable amino acid composition was not available. No evidence of the impurity present was found during the examination of the chymotryptic digest, except for the small quantity of core material, and it was therefore, possible to determine the complete sequence.

8 mg of denatured cytochrome was digested at 35 °C for 75 min with 2% (w/w) trypsin. The amount of alkali used by the autotitrator in following the digestion was about four times greater than expected. One-third of the peptides soluble at pH 5 were separated by electrophoresis at pH 6.5. Three major peptide zones with mobilities centred at 0.70, 1.00 and 1.35, were observed together with a few minor peptide bands. These three major zones were further separated by electrophoresis at pH 1.9, but no recognisable cytochrome peptides were found. All the peptides separated had glycine as N-terminus, and several were analysed by the dansyl-Edman method; all these gave the sequence Gly-Gly-Gly-Gly-. Two of these peptides had their amino acid compositions

determined on the autoanalyser; this gave Ser 3-4, Gly 10-12, Lys 1, His 1 and Thr 1, Ser 2, Gly 10-12, Lys 2, for the compositions. From the minor peptide zones only Gly-Arg, Gly-Ala-Gly-His-Lys and Tyr-Ile-Pro-Gly-Thr-Lys were strong enough to be identifiable as cytochrome peptides. Because of the severe contamination, no further examination was made of the tryptic digest.

The compositions of the impurity peptides accounted for the high glycine value observed for the total amino acid composition of the protein. In the peptides examined there were no residues normally susceptible to chymotryptic cleavage, which may account for there being no interference with the chymotryptic digest.

8. The Amino Acid Sequence of Ginkgo Cytochrome c

The amino acid composition of Ginkgo cytochrome c is given in Table 14. The values obtained were in good agreement with those calculated from the complete sequence given in Fig. 17, although the value obtained for aspartic acid in the 20 h hydrolysate was consistently high and the cysteine recoveries low in all hydrolysates; no other additional amino acids were present.

The protein was readily denatured by ethanol. For chymotryptic digestion 3.0 mg denatured cytochrome was equilibrated at 35 °C; 2% (w/w) enzyme was added at zero time, an additional 2% (w/w) after 100 min and the digestion terminated after 150 min. For tryptic digestion 2.8 mg denatured cytochrome was equilibrated at 35 °; 2% (w/w) enzyme was added at zero time, an additional 1% (w/w) after

TABLE 14

The Amino Acid Composition of Ginkgo Cytochrome c

	Mean Values 20 h hydrolysis	Mean Values 70 h hydrolysis	Mean Corrected Values	Amino acid Analysis (integral values)	Sequence values
Asp	8.09	7.17	7.63	8	7
Thr	8.24	5.67	9.26*	9	9
Ser	5.80	4.14	6.46*	6	6
Glu	13.92	13.69	13.80	14	14
Pro	8.4	8.4	8.4	8	8
Gly	12.71	13.30	13.01	13	13
Ala	8.69	8.81	8.75	9	9
Val	3.20	3.23	3.21	3	3
Cysteine	1.61	1.34	1.47	1	2
Met	1.00	0.96	0.98	1	1
Ile	3.03	3.00	3.02	3	3
Leu	7.95	7.88	7.92	8	8
Tyr	5.56	5.59	5.58	6	6
Phe	3.87	4.49	4.18	4	4
Me_3Lys	1.54	1.79	1.66	2	2
Lys	12.37	11.51	11.94	12	12
His	2.76	3.12	2.94	3	3
Arg	2.10	1.98	2.04	2	2
Trp	-	-	-	+	1

The 'mean values' were obtained from analysis of two samples of 0.1 mg cytochrome at each hydrolysis time. The 'mean corrected values' were obtained as an average of the 4 determinations with corrections made for the destruction of certain amino acids.

*Calculated from 20 h and 70 h values assuming first-order kinetics for destruction (Moore & Stein, 1963).

+Trp was not determined; the best spectral ratios of the cytochrome (Richardson, Richardson, Ramshaw, Thompson & Boulter, 1971), indicate one residue is present.

TABLE 15

Ginkgo Chymotryptic Peptides

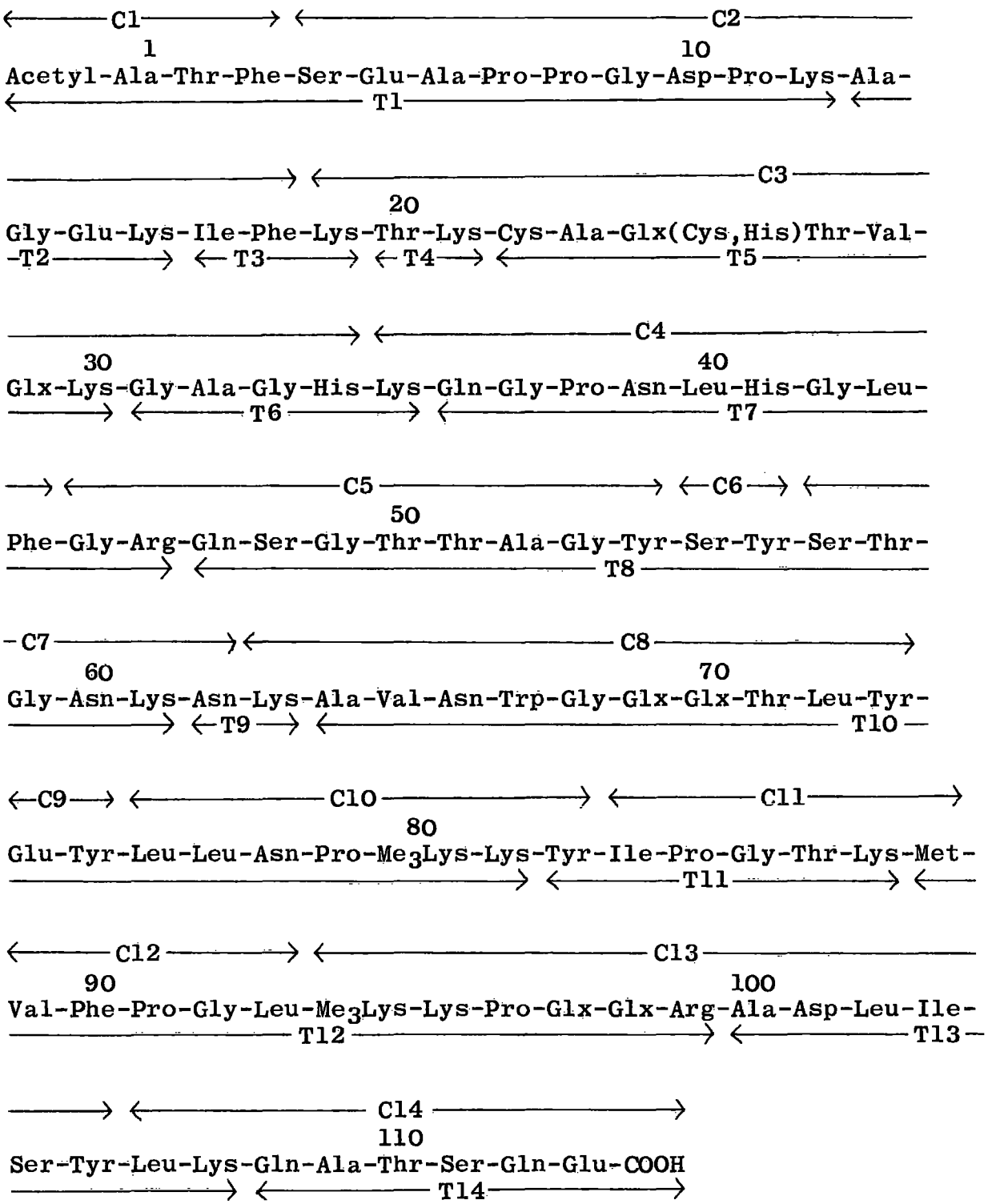
Peptide	Mobilities at pH 6.5	Mobilities at pH 1.9	Dansyl-Edman Results
C1	-1.35	0	Blocked N-terminus. Did not react. (Acetyl-Ala)-Thr-Phe
C2	-0.58		Ser- Digested with papain.
C2P1		0.60	Ser-Glx-Ala-(Pro,Pro)-Gly-Asx-Pro-Lys
C2P2		0.95	Asx-Pro-Lys-Ala
C2P3		0.82	Gly-Glx- (ε-lys present)
C3	0.65		Lys- Digested by trypsin and performic oxidized.
C3T1		1.43	Lys
C3T2		1.53	Lys-Thr-Lys
C3T3		0.27	$CySO_3$-Ala-Glx($CySO_3$,His,Thr,Val)-Glx-Lys
C3T4		1.12	Gly-Ala-Gly-His
C4	1.25		Lys-Glx-Gly-Pro-Asx-Leu-His-Gly-Leu-Phe
C4(a)	(1.70)		Lys-Glx-Gly(Pro,Asx,Leu,His)
C4(b)	0	0.56	Gly-Leu-Phe
C5	0.69		Gly-Arg-Glx-Ser-Gly-Thr-Thr-Ala-Gly-Tyr
C6	0	0.58	Ser-Tyr
C7	(0.90)		Ser- * - * -Asx-Lys-Asx
C8(a)	(-0.2)		Lys-Ala-Val-Asx- * -Gly-Glx-Glx-
C8(b)	-1.75		Gly-Glx-Glx-Thr-Leu-Tyr
C9	-1.53	0.51	Glx-Tyr
C10(a)	1.70		Leu-Leu-Asx-Pro-Me_3Lys-Lys-Tyr
C10(b)	1.82		Leu-Asx-Pro-Me_3Lys- * -Tyr
C11	1.13		Ile-Pro-Gly-Thr-Lys-Met
C12	0	0.41	Val-Phe-Pro-Gly-Leu
C13(a)	(0.70)		Me_3Lys- (o-Tyr present)
C13(b)	(1.00)		Me_3Lys- * - * -Glx-Glx- * -Ala-Asx-Leu
C13(c)	0	0.50	Ile- * -Tyr
C14	(0)	-	Leu-Lys-Glx-Ala-Thr-Ser-Glx-Glx

TABLE 16

Peptide	Mobilities at pH 6.5	Mobilities at pH 1.9	Dansyl-Edman Results
T1	-1.28		Blocked N-terminus. Digested with chymotrypsin for 1 h.
T1C1			Not isolated. (Acetyl,Ala,Thr,Phe)
T1C2			Ser-Glx-Ala-Pro-Pro-Gly-Asx-Pro-Lys
T2	0	1.28	Ala-Gly-Glx-Lys
T3	1.41		Ile-Phe-Lys
T4	1.85		Thr-Lys
T5	0		Performic oxidized. $CySO_3$-Ala-Glx-($CySO_3$,His) Thr-Val-Glx(Lys)
T6	2.43		Gly-Ala-Gly-His(Lys)
T7	1.17		Glx- Digested with chymotrypsin for 1 h.
T7C1		0.65	Glx-Gly-Pro-Asn-Leu-His-Gly-
T7C2		1.47	Gly-Arg
T8	0.53		Glx- Digested with chymotrypsin for 1 h.
T8C1		0.32	Glx-Ser-Gly-Thr-Thr-Ala-Gly-Tyr
T8C2		0.58	Ser-Tyr
T8C3		0.93	Ser-Thr-Gly-Asx(Lys)
T9	1.75		Asx-Lys
T10	0		Ala- Digested with chymotrypsin for 1 h.
T10C1	-1.16		Ala- Digested with chymotrypsin for 20 h.
T10C1C1	0		Ala-(Val,Asx) Trp
T10C1C2	-1.75		Gly-(Glx,Glx,Thr,Leu,Tyr)
T10C2	-1.61		Glx-Tyr
T10C3	1.82		Leu-Leu-Asx-Pro-Me_3Lys-Lys
T11	0.98		Tyr-Ile-Pro-Gly-Thr-Lys
T12	1.17		Met- Digested with chymotrypsin for 1 h.
T12C1		0.37	Met-Val-Phe-Pro-Gly-Leu
T12C2		1.34	Me_3Lys-Lys-Pro-Glx-Glx-Arg
T13	0	0.62	Ala-Asx-Leu-Ile-Ser-Tyr-Leu-Lys
T14	-0.98		Glx-Ala-Thr-Ser-Glx-Glx

FIGURE 17

The Amino Acid Sequence of Ginkgo Cytochrome c



60 min and the digest terminated after 75 min.

The results of the sequence analysis on the chymotryptic peptides are shown in Table 15, and on the tryptic peptides in Table 16. The solvents used for separating the dansyl-amino acids were solvents 1A, 2B and 3C.

Peptide C1. (Acetyl-Ala-Thr-Phe)

This peptide was ninhydrin negative and was located on the origin after electrophoresis of peptide C9 at pH 1.9. Its composition was determined by the dansyl method and its sequence by carboxypeptidase-A digestion. After 1 h phenyl-alanine and traces of threonine were released and after 24 h both amino acids were present in roughly equal amounts. The acetyl blocking group was identified as the 1-acetyl-2-dansyl-hydrazine derivative.

Peptide C2. (Ser-Glu-Ala-Pro-Pro-Gly-Asp-Pro-Lys-Ala-Gly-Glu-Lys-Ile-Phe)

This peptide with N-terminal serine was digested with papain for 200 min. After electrophoresis at pH 1.9, three fragments were separated, one of which, peptide C2P3, was not present in sufficient yield for analysis. Proline residues were not identified at positions 7 and 8, and these positions were established using peptide T1C1. The electrophoretic mobility of the parent peptide at pH 6.5 showed no amide residues were present.

Peptide C3. (Lys-Thr-Lys-Cys-Ala-Glx-Cys-His-Thr-Val-Glx-Lys-Gly-Ala-Gly-His)

The haem peptide was eluted from the paper by 20% (v/v) pyridine and had lysine at its N-terminal. It was digested with trypsin for 120 min and then performic oxidized. Peptide

C3T3 was present in low yield and gave weak spots on the dansyl-Edman analysis. Determination of its composition was used to obtain evidence for the identity of the 4th to 7th residues, and the remainder of the peptide used for obtaining identification of the last two residues by the dansyl-Edman method. The mobility of the haem peptide could not be used directly to obtain information on amide residues, as it contained histidine and cysteic acid residues.

Peptide C4. (Lys-Gln-Gly-Pro-Asn-Leu-His-Gly-Leu-Phe)

The histidine residue was identified as the α-dansyl-His derivative and from the composition of the peptide. Peptides due to partial cleavage at histidine-41 were found; C4(a) was present only in trace quantity trailing peptide C10(a). The electrophoretic mobility at pH 6.5 indicated that both residues 36 (glutamine) and 39 (asparagine) were amides.

Peptide C5. (Gly-Arg-Gln-Ser-Gly-Thr-Thr-Ala-Gly-Tyr)

The electrophoretic mobility at pH 6.5 indicated that residue 47 (glutamine) was an amide.

Peptide C6. (Ser-Tyr).

Peptide C7. (Ser-Thr-Gly-Asn-Lys-Asn)

This peptide was not identified as a distinct ninhydrin positive peptide after pH 6.5 electrophoresis, but was located in the tailings of peptide C11. Evidence for the threonine-58 and glycine-59 was obtained from the tryptic peptide T8C3. The electrophoretic mobility indicated that residues 60 and 62 are both asparagine.

Peptide C8. (Lys-Ala-Val-Asn-Trp-Gly-Glx-Glx-Thr-Leu-Tyr)

Two Ehrlich positive zones were found after pH 6.5 electrophoresis. One, peptide C8(a), was located at the acidic edge of the neutral peptide zone. It had a faint blue fluorescence and did not require re-electrophoresis at pH 1.9. On the trifluoroacetic acid stage of the first Edman degradation, a pink colour was observed indicating tryptophan (Uphaus et al., 1969). The dansyl-Edman analysis did not establish the full length of the peptide so no definite evidence on amide content was obtained. Peptide C8(b), due to partial cleavage at tryptophan-67, had an electrophoretic mobility which indicated that both residues 69 and 70 were glutamic acid (see peptide T10).

The other Ehrlich positive zone was located at the basic edge of the neutral peptide zone. However, no peptide was found during the dansyl-Edman analysis of these neutral peptides, which corresponded to the tryptophan sequence in the protein.

Peptide C9. (Glu-Tyr)

The electrophoretic mobility at pH 6.5 indicated that residue 74 was glutamic acid.

Peptide C10. (Leu-Leu-Asn-Pro-Me_3Lys-Lys-Tyr)

Two peptides were found due to partial cleavage at leucine-76. The mobilities of both at pH 6.5 showed that residue 78 was asparagine.

Peptide C11. (Ile-Pro-Gly-Thr-Lys-Met)

This peptide was strongly platinic iodide positive.

Peptide C12. (Val-Phe-Pro-Gly-Leu)

Peptide C13. (Me_3Lys-Lys-Pro-Glx-Glx-Arg-Ala-Asp-Leu-Ile-Ser-Tyr)

Peptide C13(a) was present in very low yield and was eluted from the electrophoretogram with the major peptide C5. A partial split at leucine-102 was observed. Peptide C13(b) was eluted with the shorter peptide C11. After six steps of Edman degradation only this peptide remained, allowing identification of the remaining residues.

Peptide C14. (Leu-Lys-Gln-Ala-Thr-Ser-Gln-Glu)

This neutral peptide was separated at the basic edge of the neutral peptide zone and did not require re-electrophoresis at pH 1.9. To avoid confusion which can be caused by incomplete Edman degradation, the length of the peptide was established during the examination of the last three residues. At each of the fifth and subsequent steps of Edman degradation, dansylated samples were examined with and without hydrolysis. The dansyl-peptides from the fifth and sixth steps did not co-chromatograph with dansyl-serine or with either dansyl-glutamic acid or dansyl-glutamine, which were the derivatives identified from the hydrolysed dansyl-peptide samples. The dansyl-'peptide' sample from the seventh step of degradation co-chromatographed with the dansyl-glutamic acid and did not contain any dansyl-glutamine. No dansyl derivatives were present after an eighth step in the dansyl-Edman procedure. The electrophoretic mobility of the peptide indicated that one acidic and two amide residues were present. As the C-terminal residue of the peptide was shown to be glutamic acid-113, residues 108 and 112 were both glutamine.

Peptide T1. (Acetyl-Ala-Thr-Phe-Ser-Glu-Ala-Pro-Pro-Gly-Asp-Pro-Lys)

This peptide gave only ϵ-dansyl-lysine on N-terminal analysis. After digestion for 1 h with chymotrypsin a clean serine N-terminus was obtained. This peptide was analysed directly by the dansyl-Edman method without being separated from the blocked N-terminal fragment, whose presence was confirmed by total dansyl analysis. The mobility of the peptide indicated that both residues 5 and 10 were acidic.

Peptide T2. (Ala-Gly-Glu-Lys)

This peptide was neutral at pH 6.5 indicating that residue 15 was glutamic acid.

Peptide T3. (Ile-Phe-Lys)

Peptide T4. (Thr-Lys)

Peptide T5. (Cys-Ala-Glx-Cys-His-Thr-Val-Glx-Lys)

The haem peptide was eluted by 20% (v/v) pyridine and then performic oxidized to remove the haem moiety before dansyl-Edman analysis. The presence of cysteine-25, histidine-26 and lysine-30 residues followed from the composition of the peptide. No information on amide content could be derived because of the histidine and cysteic acid residues present.

Peptide T6. (Gly-Ala-Gly-His-Lys)

This peptide was yellow with ninhydrin. The histidine-34 was identified as the α-dansyl-His derivative. Lysine-35 was shown to be present from the composition of the peptide.

Peptide T7. (Gln-Gly-Pro-Asn-Leu-His-Gly-Leu-Phe)

This peptide did not separate from peptide T12 after

electrophoresis at pH 6.5. It was digested for 1 h with chymotrypsin giving two fragments. The mobility of the peptide indicated that residue 36 was glutamine and residue 39 was asparagine.

Peptide T8. (Gln-Ser-Gly-Thr-Thr-Ala-Gly-Tyr-Ser-Tyr-Ser-Thr-Gly-Asn-Lys)

This peptide T8C3 enabled positive identification of residues threonine-58 and glycine-59, which were not identified in the chymotryptic peptide C7. Lysine-61 was obtained from the composition of the peptide. The electrophoretic mobility indicated that both residues 47 (glutamine) and 60 (asparagine) were amides.

Peptide T9. (Asn-Lys)

This peptide gave a brown colour with ninhydrin, which is characteristic of asparagine as the N-terminal residue; the electrophoretic mobility confirmed asparagine-62.

Peptide T10. (Ala-Val-Asn-Trp-Gly-Glx-Glx-Thr-Leu-Tyr-Glu-Tyr-Leu-Leu-Asn-Pro-Me3Lys-Lys)

A single Ehrlich positive peptide located in the neutral peptide zone was found after pH 6.5 electrophoresis. N-terminal analysis gave alanine, and composition of the peptide was the same as that expected from the analysis of the chymotryptic peptides C8, C9 and C10. The peptide was digested for 1 h with chymotrypsin and the fragments separated by electrophoresis at pH 6.5. Peptides T10C2 and T10C3 were analysed by the dansyl-Edman method, and shown to have overall charges of -1 and +2 at pH 6.5, consistent with those expected from the chymotryptic peptides C9 and C10. This indicated

that residue 78 was asparagine. Peptide T1OC1 had N-terminal alanine, and its composition indicated that it contained residues 64-73, giving an overall charge of -2 at pH 6.5. Thus, the sum of the charges of the T1OC peptides did not equal the charge of the parent peptide T1O.

Digestion of peptide T1OC1 with chymotrypsin overnight gave two peptides, T1OC1C1 and T1OC1C2, which were separated on electrophoresis at pH 6.5. Peptide T1OC1C1 was Ehrlich positive and neutral which indicated that residue 66 was asparagine. Carboxypeptidase-A digestion for 3 h released tryptophan which was directly identified as its dansyl derivative. The sequences were confirmed by compositions as insufficient material remained for dansyl-Edman analysis. Peptide T1OC1C2 had an overall charge of -2, the same as C8(b), which indicated that no amide residues were present. The discrepancy in charge between the T1OC peptides and the peptide T1O was probably due to one of the three residues 69, 70 and 74 being present as an amide in peptide T1O, but deamidated during the second digestion. It was assumed that the neutral mobilities of the original peptides T1O and C8(a) represented the state when no deamidation had occurred. From its composition, the neutral peptide C8(a) was either residues 63-73 or 63-75 and contained the amide residue asparagine-66 indicated by peptide T1OC1C1. If peptide C8(a) was residues 63-73 then in order to be neutral either residue 69 or 70 would have been an amide. In peptides C8(b) and T1OC1C2 both those residues were acidic. If peptide C8(a) was residues 63-75 then in order to be neutral two of the residues 69, 70 and 74 would have been amides.

This was not consistent with the mobility of peptide T10 which indicated that only one of these three residues was an amide. An amide at residue 69 or 70 and a second acidic residue at position 74 was consistent with the mobilities of the peptides C8(a) and T10. However, no peptide equivalent to either C8(b) or T10C1C2 with one amide residue present was found.

Peptide T11. (Tyr-Ile-Pro-Gly-Thr-Lys)

Peptide T12. (Met-Val-Phe-Pro-Gly-Leu-Me_3Lys-Lys-Pro-Glx-Glx-Arg)

This peptide had methionine as its N-terminal residue and was located after pH 6.5 electrophoresis in approximately the same position as peptide T7. It was digested for 1 h with chymotrypsin without prior separation from the other peptide and the resultant fragments (T12C1 and T12C2) separated by electrophoresis at pH 1.9.

The mobility of the original peptide indicated the presence of one amide residue, and this will occur in either position 97 or 98.

Peptide T13. (Ala-Asp-Leu-Ile-Ser-Tyr-Leu-Lys)

Dansyl-Edman analysis of this peptide confirmed that position 104 was serine; this position had been uncertain in the chymotryptic peptide C13(c). The electrophoretic mobility at pH 6.5 indicated that residue 101 was aspartic acid.

Peptide T14. (Gln-Ala-Thr-Ser-Gln-Glu)

This peptide was suspected as the C-terminal peptide, as

there was no ε-dansyl-lysine derivative on N-terminal determination and lysine and arginine were both absent. The composition agreed with the C-terminal region determined from the chymotryptic peptides. The length of the peptide was determined and glutamic acid shown to occupy position 113 using the same methods as those used to examine peptide C14. The electrophoretic mobility at pH 6.5 indicated one acid residue, which was placed as glutamic acid-113 by dansylation without hydrolysis. There was no evidence for glutamine in this position and hence, residues 108 and 112 must both be glutamine.

The overlapping chymotryptic and tryptic peptides give the sequence of Ginkgo cytochrome c as in Fig. 17; this sequence is in accord with the determined amino acid composition. Every residue was positively identified during the sequence analysis except for cysteine-25 and histidine-26, which were identified in peptide compositions. Their order given in the sequence, Fig. 17, is based on homology with all other cytochromes c (Dayhoff, 1969), and on their known role in the structure of the protein (Dickerson et al., 1971).

The electrophoretic mobilities of the peptides at pH 6.5 (Table 15 and Table 16) indicated the presence of 12 amide residues in the sequence. The presence of one amide in the haem peptide region was established by comparison of the haem peptide mobilities with those of horse haem peptides, for which the amide content is known (Margoliash et al., 1961). Of the amide residues, 9 may be directly placed; these are asparagine in positions 39, 60, 62, 66 and 78, and glutamine in positions 36, 47, 108 and 112. Homology with other

cytochromes c (Dayhoff, 1969) suggests that the other amide residues are glutamine residues in positions 24, 70 and 97.

The observed enzyme specificities were consistent with those expected (Smyth, 1967), except that the susceptibility of tryptophan-67 to chymotryptic cleavage was considerably less than expected. Full cleavage with chymotrypsin also occurred at histidine-34, asparagine-62, methionine-88 and leucine-93; partial cleavage occurred at histidine-41 and leucines-76 and 102. No tryptic digestion was observed at either of the ϵ-N-trimethyllysine residues.

The amide residue present in peptides C8 and T10 in either position 69 or 70 appeared particularly susceptible to deamidation.

9. The Amino Acid Sequence of Niger Cytochrome c

The complete sequence of niger cytochrome c is given in Fig. 18, and was deduced from the sequences of the chymotryptic and tryptic peptides.

The protein was readily denatured by ethanol. For chymotryptic digestion 3.5 mg denatured cytochrome was equilibrated at 37 °; 2% (w/w) enzyme was added at zero time, an additional 2% (w/w) after 80 min and the digestion terminated after 120 min. For tryptic digestion 2.5 mg denatured cytochrome was equilibrated at 37 °; 2% (w/w) enzyme was added at zero time, an additional 2% (w/w) after 60 min and the digestion terminated after 90 min.

The results of the sequence analysis on the chymotryptic peptides are shown in Table 17 and on the tryptic peptides in Table 18. The solvents used for separating the dansyl

TABLE 17

Niger Chymotryptic Peptides

Peptide	Mobility at pH 6.5	Mobility at pH 1.9	Dansyl-Edman Results
C1	-1.30	0	Blocked N-terminus. Does not react. (Acetyl-Ala)-Ser-Phe
C2	-0.66		Ala-Glx-Ala-Pro-Ala-Gly-Asx-
C2(a)	0	0.67	Ile-Phe
C3	0.65		Lys Performic oxidized and digested with chymotrypsin for 2 h.
C3C1		0.61	Lys-Thr-Lys-($CySO_3$)-Ala-Glx-($CySO_3$-His)
C3C2		0.98	Thr-Val-Glx-Lys-Gly-Ala-Gly-(His)
C4	0.90+	0.55	Lys-Glx-Gly-Pro-Asx-Leu-Asx-Gly-
C4(a)	0	0.59	Gly-Leu-Phe
C5	0.90+	0.50	Gly-Arg-Glx-Ser-Gly-Thr-Thr-Ala-Gly-Tyr
C6	0	0.64	Ser-Tyr
C7	1.09		Ser-Ala-Ala-Asx- * -Asx
C8	0	0.45	Ala-Val-Ala-Trp
C9	-1.67	0.34	Glx-Glx-Asx-Ser-Leu-Tyr
C10	(0.60)	0.71	Asx-Tyr-Leu-Leu-Asx-Pro-Me_3Lys- * -Tyr
C10(a)	-1.53	0.51	Asx-Tyr
C10(b)	-1.67	0.45	Asx-Tyr-Leu
C10(c)	1.70	0.99	Leu-Asx-Pro- * -Lys-Tyr
C11	1.09	0.83	Ile-Pro-Gly-Thr-Lys-Met
C12	0	0.47	Val-Phe-Pro-Gly-Leu
C13	(0.60)		Me_3Lys- * -Pro-Glx-Glx-Arg-Ala-Asp-Leu-Ile-Ala
C13(a)	(1.70)		Me_3Lys- * - * -Glx-Glx-Arg
C13(b)	0		Ile-Ala-Tyr
C14	(1.00)	0.85	Leu-Lys-Ala-Ser-Thr-Ala

+Leading edge of zone.

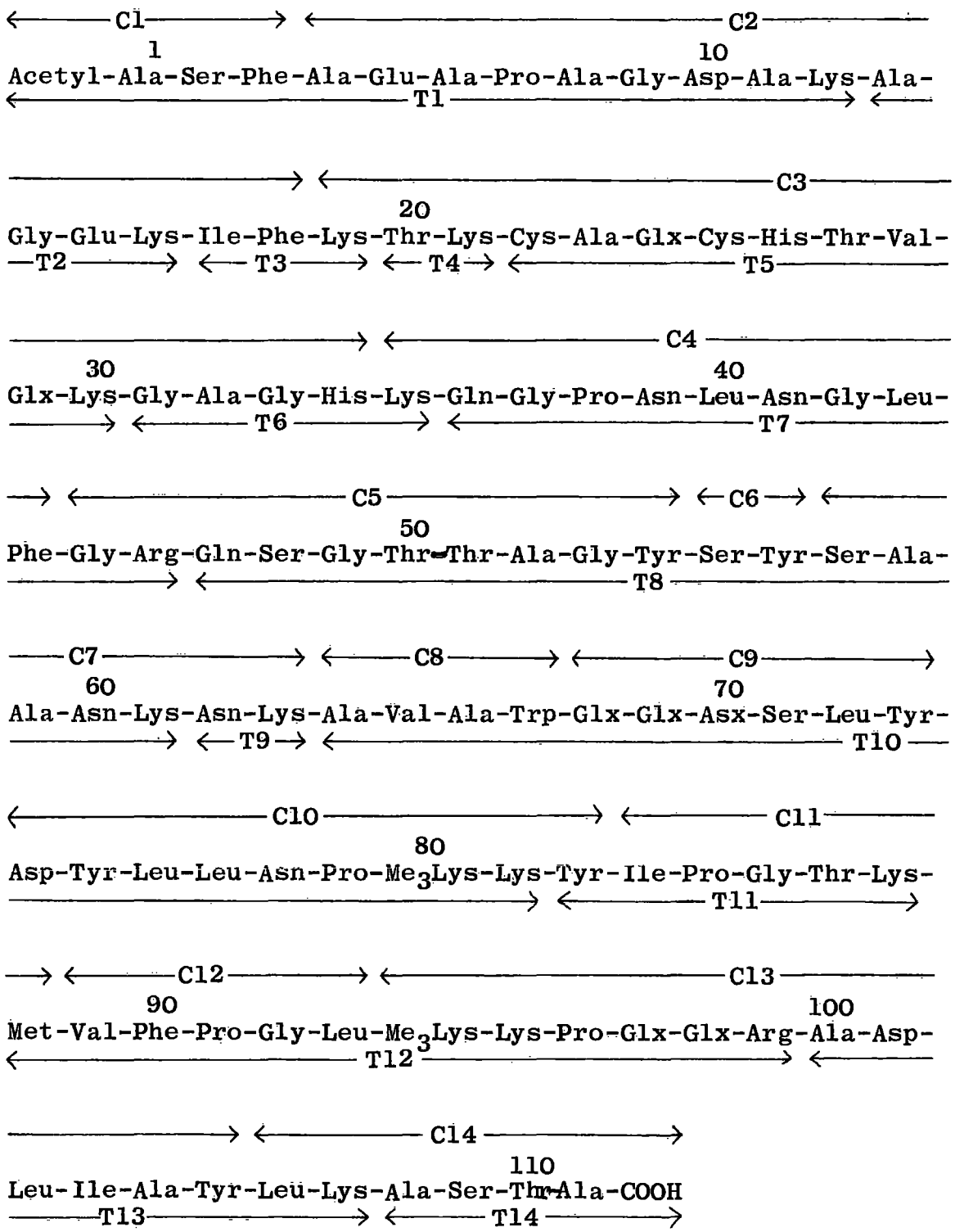
TABLE 18

Niger Tryptic Peptides

Peptide	Mobility at pH 6.5	Mobility at pH 1.9	Dansyl-Edman Results
T1		0.28	No N-terminus detected. Digested with chymotrypsin for 1 h.
T1C1		0	Acetyl(Ala,Ser,Phe)
T1C2		0.53	Ala-Glx-Ala-Pro-Ala-Gly-Asx-Ala-Lys
T2	0	1.06	Ala-Gly-Glx-Lys
T3	1.30		Ile-Phe-Lys
T4	1.69		Thr-Lys
T5	0		Performic oxidized.
		0.31	$CySO_3$-Ala-Glx($CySO_3$)His-Thr-
T6	2.18		Gly-Ala-Gly-His(Lys)
T7	0.64	0.53	Glx- Digested with chymotrypsin for 1 h.
T7C1		0.30	Glx-Gly-Pro-Asx-
T7C2		0.58	Gly-Leu-Phe
T7C3		1.46	Gly-Arg
T8	0.64	0.45	Glx- Digested with chymotrypsin for 1 h.
T8C1		0.32	Glx(Ser,Gly,Thr,Thr,Ala,Gly,Tyr)
T8C2		0.58	Ser-Tyr
T8C3		0.95	Ser-Ala-Ala(Asx,Lys)
T9	1.69		Asx-Lys
T10	0		Not isolated.
T11	0.91		Tyr-Ile-Pro-Gly-Thr-Lys
T12	1.10	(0.86)	Digested with chymotrypsin for 1 h.
T12C1		0.44	Met-Val-Phe-Pro-Gly-Leu
T12C2		1.26	Me_3Lys-Lys-Pro(Glx,Glx,Arg)
T13	0	0.63	Ala-Asx-Leu-Ile-Ala-Tyr-Leu-Lys
T14	0	0.56	Ala-Ser-Thr-Ala

FIGURE 18

The Amino Acid Sequence of Niger Cytochrome c



amino acids were solvents 1A, 2B and 3C.

Peptide C1. (Acetyl-Ala-Ser-Phe)

This peptide was ninhydrin negative and was located on the origin after electrophoresis of peptides C10(b) and C10(c) at pH 1.9. It was purified by electrophoresis at pH 6.5 and detected by the starch-iodide reagent. Its composition was determined by the dansyl method and its sequence by carboxypeptidase-A digestion. After 3 h phenylalanine was released and after 24 h both phenylalanine and serine were released. The acetyl blocking group was identified as the 1-acetyl-2-dansyl-hydrazine derivative.

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

This acidic peptide did not form a sharp zone at pH 6.5 electrophoresis. Its acidic mobility showed no amide residues were present. Peptides T1, T2 and T3 gave the complete sequence for this region. A peptide due to cleavage at lysine-16 was found.

Peptide C3. (Lys-Thr-Lys-Cys-Ala-Glx-Cys-His-Thr-Val-Glx-Lys-Gly-Ala-Gly-His)

The haem peptide was eluted from the paper by 20% (v/v) pyridine and had lysine as its N-terminal. It was performic oxidized and then digested with chymotrypsin. Two fragments were recovered in low yield after electrophoresis at pH 1.9 resulting from cleavage at histidine-26.

Peptide C4. (Lys-Gln-Gly-Pro-Asn-Leu-Asn-Gly-Leu-Phe)

Peptide C4(a) was found from partial cleavage at asparagine-41. The electrophoretic mobility at pH 6.5

indicated that residues 36 (glutamine) and both 39 and 41 (asparagine) were all amides.

Peptide C5. (Gly-Arg-Gln-Ser-Gly-Thr-Thr-Ala-Gly-Tyr)

This peptide was ninhydrin yellow and its electrophoretic mobility at pH 6.5 indicated that residue 47 (glutamine) was an amide.

Peptide C6. (Ser-Tyr)

Peptide C7. (Ser-Ala-Ala-Asn-Lys-Asn)

The peptide examined was due to cleavage at asparagine-62. The electrophoretic mobility at pH 6.5 indicated that residues 60 and 62 were both amides. There was no evidence of any chymotryptic peptide involving lysine-63.

Peptide C8. (Ala-Val-Ala-Trp)

This was an Ehrlich positive neutral peptide separated after electrophoresis at pH 1.9. The tryptophan residue was positively identified after carboxypeptidase-A digestion for 90 min.

Peptide C9. (Glx-Glx-Asx-Ser-Leu-Tyr)

The electrophoretic mobility at pH 6.5 indicated the presence of one amide residue which could not be placed directly.

Peptide C10. (Asp-Tyr-Leu-Leu-Asn-Pro-Me_3Lys-Lys-Tyr)

Peptides due to partial cleavage at tyrosine-75 and leucine-76 were found. The electrophoretic mobilities at pH 6.5 of these smaller peptides, indicate that residue 74 is acidic (aspartic acid) and residue 78 an amide (asparagine).

Peptide C11. (Ile-Pro-Gly-Thr-Lys-Met)

Peptide C12. (Val-Phe-Pro-Gly-Leu)

Peptide C13. (Me_3Lys-Lys-Pro-Glx-Glx-Arg-Ala-Asp-Leu-Ile-Ala-Tyr)

Partial cleavage at arginine-99 and leucine-102 occurred. Only two of the possible fragments were identified however. The electrophoretic mobilities of peptide C13(a) indicated that one of residues 97 or 98 was an amide residue, and therefore, from the mobility of peptide C13 residue 101 is aspartic acid.

Peptide C14. (Leu-Lys-Ala-Ser-Thr-Ala)

A sample taken after the fifth Edman degradation was dansylated, but not hydrolysed; this sample was found to co-chromatography with standard dansyl-alanine. This shows alanine to be the C-terminus of the peptide. Since alanine is not normally susceptible to chymotryptic cleavage, this peptide was taken as the C-terminal peptide of the protein.

Peptide T1. (Acetyl-Ala-Ser-Phe-Ala-Glu-Ala-Pro-Ala-Gly-Asp-Ala-Lys)

This peptide was not clearly defined at pH 6.5 and was isolated after electrophoresis of the acidic peptides at pH 1.9. It was digested with chymotrypsin and the fragments separated at pH 1.9. The blocked N-terminal peptide T1C1 was pooled with peptide C1 for the determination of the acetyl group.

Peptide T2. (Ala-Gly-Glu-Lys)

Peptide T3. (Ile-Phe-Lys)

Peptide T4. (Thr-Lys)

Peptide T5. (Cys-Ala-Glx-Cys-His-Thr-Val-Glx-Lys)

The haem peptide was eluted by 20% (v/v) pyridine, performic oxidized and purified by electrophoresis at pH 1.9.

Peptide T6. (Gly-Ala-Gly-His-Lys)

This peptide was ninhydrin positive. Lysine-35 followed from the composition of the peptide.

Peptide T7. (Gln-Gly-Pro-Asn-Leu-Asn-Gly-Leu-Phe-Gly-Arg)

This peptide was digested with chymotrypsin and three fragments separated by electrophoresis at pH 1.9. The mobility at pH 6.5 indicated that residues 36 (glutamine), 39 and 41 (asparagines) were all amides.

Peptide T8. (Gln-Ser-Gly-Thr-Thr-Ala-Gly-Tyr-Ser-Tyr-Ser-Ala-Ala-Asn-Lys)

This peptide was digested with chymotrypsin and three fragments separated at pH 1.9. The mobility at pH 6.5 indicated that residues 47 (glutamine) and 60 (asparagine) were both amides.

Peptide T9. (Asn-Lys)

The mobility at pH 6.5 indicated that residue 62 was asparagine.

Peptide T10. (Ala-Val-Ala-Trp-Glx-Glx-Asx-Ser-Leu-Tyr-Asp-Tyr-Leu-Leu-Asn-Pro-Me_3Lys-Lys)

A single Ehrlich positive peptide which was neutral was observed after electrophoresis at pH 6.5. It was not isolated after pH 1.9 electrophoresis. The sequence for this region is given by peptides C8, C9 and C10.

Peptide T11. (Tyr-Ile-Pro-Gly-Thr-Lys)

Peptide T12. (Met-Val-Phe-Pro-Gly-Leu-Me_3Lys-Lys-Pro-Glx-Glx-Arg)

This peptide was digested with chymotrypsin. Two fragment peptides and undigested peptide were recovered after electrophoresis at pH 1.9. The mobility at pH 6.5 indicated the presence of one amide residue, which could not be directly placed.

Peptide T13. (Ala-Asp-Leu-Ile-Ala-Tyr-Leu-Lys)

The neutral mobility on electrophoresis at pH 6.5 indicated that residue 101 is aspartic acid.

Peptide T14. (Ala-Ser-Thr-Ala)

After three steps of Edman degradation, a dansylated sample which was not hydrolysed co-chromatographed with dansyl-alanine. As alanine was not normally susceptible to tryptic digestion, this peptide was assumed to be the C-terminal peptide of the protein; see peptide C14.

Every residue in the sequence of niger cytochrome c (Fig. 18) was positively identified during the dansyl-Edman analysis of the chymotryptic or tryptic peptides, except for alanine-1, cysteine-25 and lysine-61, which were identified from peptide compositions. No additional peptides to those given in Tables 17 and 18 were identified.

The electrophoretic mobilities of the peptides at pH 6.5 indicated the presence of 10 amide residues in the sequence. The presence of one amide residue in the haem peptide region was established by comparison of the haem peptide mobilities with those of horse heart haem peptide mobilities, for which the amide content is known (Margoliash et al., 1961). Of the amide residues, 7 can be placed directly; these are asparagine in positions 39, 41, 60, 62 and 78, and glutamine

in positions 36 and 47. Homology with other cytochromes c (Dayhoff, 1969), suggests that the other amide residues are glutamine in positions 24 and 97 and asparagine in position 70.

The observed enzyme specificities were consistent with those expected (Smyth, 1967). Full chymotryptic cleavage also occurred at histidine-34, asparagine-62, lysine-63 and methionine-88; partial cleavage occurred at lysine-16, asparagine-41, tyrosine-75, leucine-76, arginine-99 and leucine-102. No tryptic digestion was observed at either trimethyllysine residue.

10. Matrix of Difference Calculations

The matrices of amino acid differences and minimum mutation distances between all the cytochrome c sequences so far examined, are given in Appendix 1 and Appendix 2. The sequences used for the construction of these matrices are given in Appendix 3. Amide residues which could not be directly placed were assigned positions by homology with other sequences (see footnote to Appendix 3).

The mean variations between the members of the major taxonomic groups have been calculated (Table 19). In cases in which identical cytochromes c occur in more than one species, a single value was used in calculating means and deviations. Using a paleontologically estimated time period of 280 million years for the divergence of the mammalian and avian lines of descent (Simpson, 1964), the two 'unit evolutionary periods' have been calculated. The 'unit evolutionary period' is the average time taken for one variation to occur between two diverging lines of descent (Zuckerkandl &

TABLE 19

Comparisons between cytochromes c of different taxonomic groups

Groups compared	Average number of variant residues	Time since divergence of lines (millions of years)	Corrected* time since divergence of lines (millions of years)	Average of the minimum mutation distance	Time since divergence of lines (millions of years)	Corrected* time since divergence of lines (millions of years)
Mammals - Birds	9.9 ± 1.7(36)	280+	280+	13.7 ± 2.4(36)	280+	280+
Mammals, Birds, Reptiles - Fish	17.8 ± 3.2(80)	500	520	26.4 ± 4.9(80)	540	570
Vertebrates - Invertebrates	26.6 ± 3.1(84)	750	830	33.6 ± 4.9(84)	690	740
Animals - Plants	47.3 ± 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 ± 3.7(64)	1400	1770

The standard deviation and the number of comparisons are given for each average difference between groups.

*Correction was made according to the formula of Feller (1950). See Fig. 19 .

+Fixed time estimated from paleontological evidence (Simpson, 1964).

FIGURE 19. 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 for the time since the 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). To obtain n for amino acid changes, of the 113 positions in the sequences considered, 32 were statistically calculated to be invariant (Fitch & Markowitz, 1970). The value of n used for nucleotide changes was corrected as described by Margoliash & Fitch (1968).

The major group divergences indicated are:-

A = Mammals to Birds.

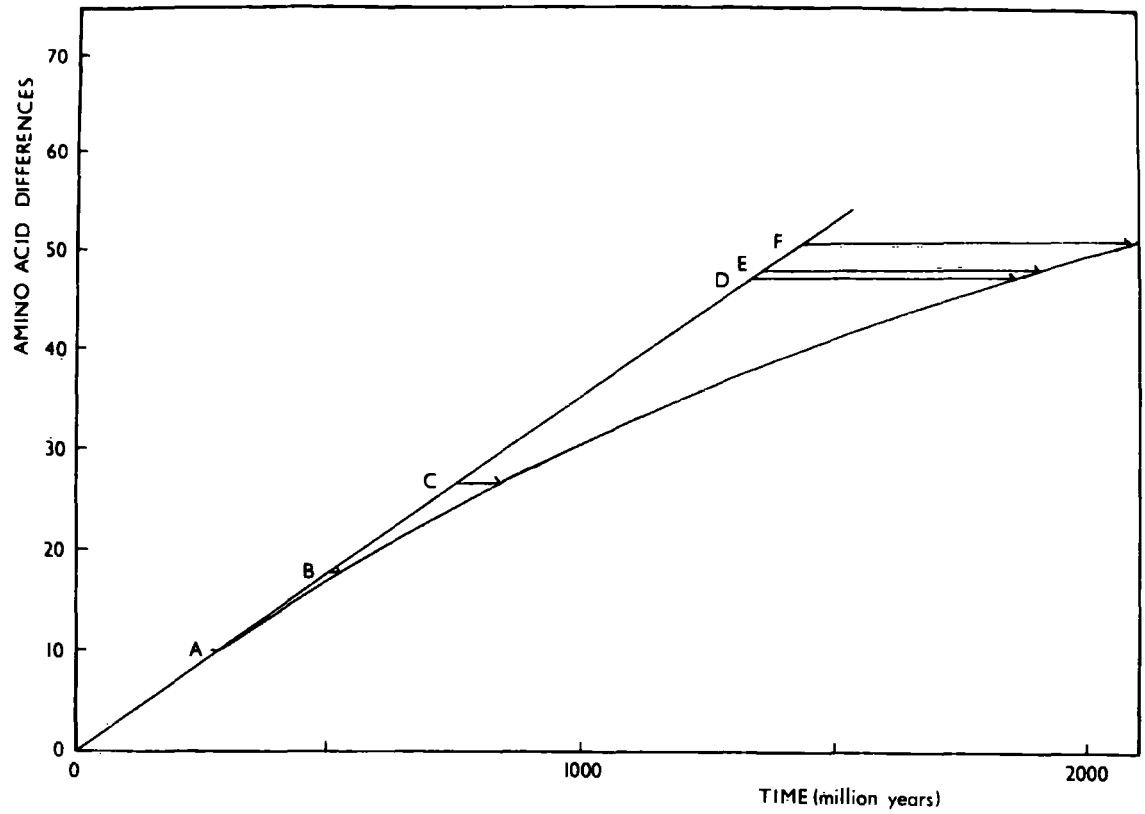
B = Mammals, Birds and Reptiles to Fish.

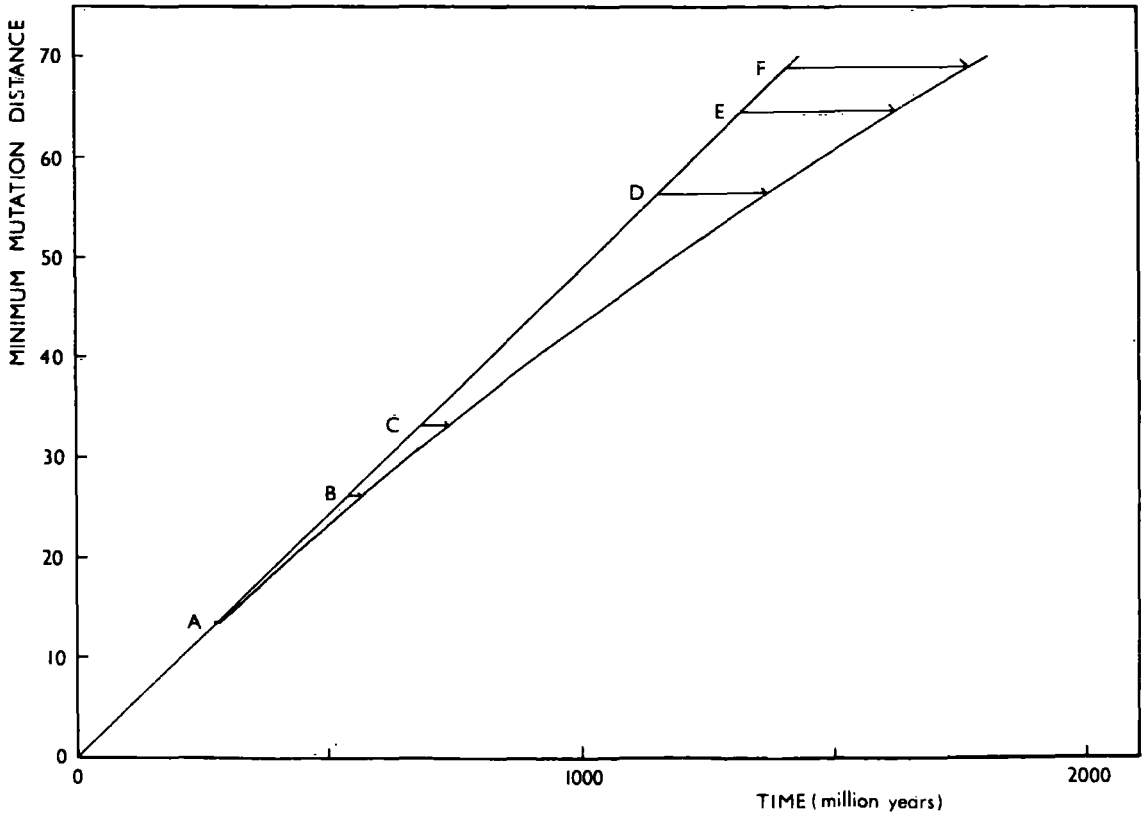
C = Vertebrates to Invertebrates.

D = Animals to Plants.

E = Animals to Fungi.

F = Plants to Fungi.


AMINO ACID DIFFERENCES
70
60
50
40
30
20
10
0
0
1000
2000
TIME (million years)
A
B
C
D
E
F


MINIMUM MUTATION DISTANCE
70
60
50
40
30
20
10
0
0
1000
2000
TIME (million years)
A
B
C
D
E
F

Pauling, 1962; Margoliash & Smith, 1965). The value of the 'unit evolutionary period' obtained is 28.3 million years for an amino acid substitution and 20.5 million years for a nucleotide substitution. Linear extrapolations have been made from the fixed time point using these unit evolutionary periods (Fig. 19), and the minimum times since the divergences of other taxonomic groups determined (Table 19). Further time estimates were made using a correction for the probability of the occurrence of multiple evolutionary effective mutations in the same codon or multiple mutations at the same nucleotide position in the structural gene (Fig. 19 and Table 19) (Margoliash & Smith, 1965).

The average amino acid differences between the gymnosperm Ginkgo and the angiosperms, was calculated as 17.4. The amino acid 'unit evolutionary period' leads to an approximate minimum time for the divergence between these lines of descent of 500 million years.

11. Phylogenetic Trees Constructed by Numerical Methods

The tree topologies relating all the higher plant cytochromes c, shown in Figs. 20 and 21, were constructed using the fusion strategy described by Fitch & Margoliash (1967a), using minimum mutation distance (Fitch & Margoliash, 1967a; Fig. 20), and amino acid differences (Fig. 21) as the similarity measures.

The trees shown in Figs. 22 to 25 were those constructed using the 'flexible' method described by Lance & Williams (1967a), using amino acid differences and minimum mutation distance as similarity measures. Figs. 22 and 23 show the

phylogenies obtained from both measures when $\beta = -0.24$; this value of β is considered to be close to an ideal 'space conserving' fusion (Lance & Williams, 1967a). Figs. 24 and 25 indicate the changes in topology obtained when changing from a 'space contracting' (β positive) to a 'space dilating' (β negative) system.

12. Phylogenetic Trees Constructed by the Ancest ral Sequence Method

Numerically derived trees can be assessed by their 'percent standard deviation' (Fitch & Margoliash, 1967a). When several alternative trees are considered, a range of values for this parameter are obtained from which a minimum value is chosen corresponding to the best tree. With the ancest ral sequence method, trees are assessed by the minimum number of mutations occurring through the topology and the best tree is considered as that for which this value is a minimum. Since the number of mutations on a tree is an integral value, it is possible that no unique minimum may exist and that several equally probable alternatives may be found.

This situation of no unique minimum was found to exist when the 16 higher plant cytochrome c sequences (see Appendix 3) were used to construct a phylogenetic tree. Fig. 26 shows the result of an attempt to find a 'basic tree' relating a few of the plant cytochromes c. The aim of this 'basic tree' methodology was to find the maximum number of species for which a unique best tree exists; this tree must be reasonably compatible with those trees obtained when all 16 species are considered and with biological information.

Alternative trees obtained from the species used in the basic tree, when a tolerance of one mutation above the minimum value is allowed, are shown in Fig. 27. The addition of each of the remaining species in turn to this 'basic tree' did not give a larger tree, for which an acceptable unique minimum existed. Figs. 28 and 29 show the alternative positions where these additional species were located on the 'basic tree' when added individually.

Some of the trees constructed from the 16 species are shown in Figs. 30a, b and c; no unique solution was found. Several of the alternatives which exist are caused by changes among certain small groups of species which do not lead to major changes in the topology of the tree concerned; these group changes are indicated in the respective figures.

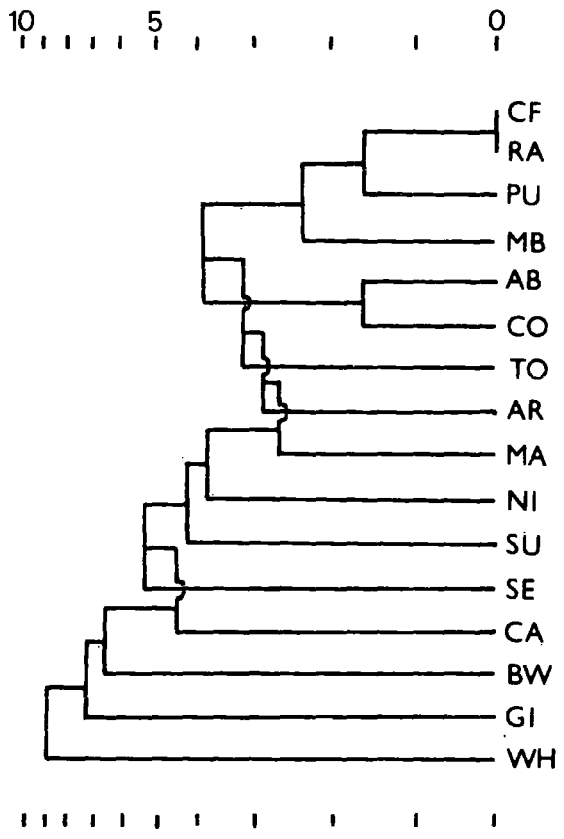
The higher plant cytochromes *c* were also considered together with the one algal (Meatyard, 1971), the fungal and the animal cytochromes *c* and the cytochrome c_2 from the bacterium *Rhodospirillum Rubrum* (Dus, Sletten & Kamen, 1968). No unique best solution tree was found and again general alternatives within the plant group exist. One of the alternative trees is shown in Fig. 31. No clear ordination for the three eukaryotic kingdoms was found; the differences between the trees equivalent to the possible alternative ordinations of these kingdoms are small and the values obtained are summarised in Fig. 32. The early points of divergence within the plant group were established (see Fig. 31), and the ancestoral sequences at these points and at the time of the divergence of the eukaryotic kingdoms established (Fig. 33).

FIGURE 20. A phylogenetic tree relating sixteen plant species.

The tree was constructed by the method of Fitch & Margoliash (1967a) using minimum mutation distance between species.

FIGURE 21. A phylogenetic tree relating sixteen plant species.

The tree was constructed by the method of Fitch & Margoliash (1967a) using amino acid differences between the sequences.


10
5
0
CF
RA
PU
MB
AB
CO
TO
AR
MA
NI
SU
SE
CA
BW
GI
WH

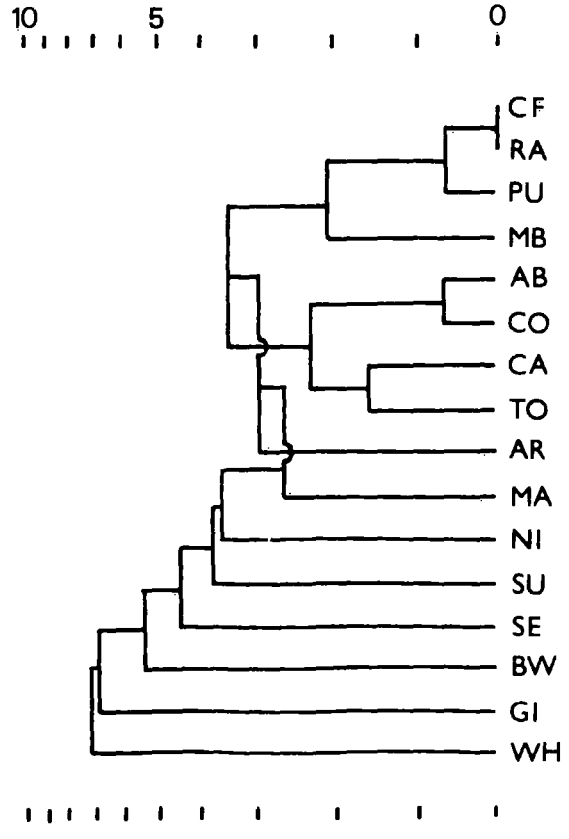
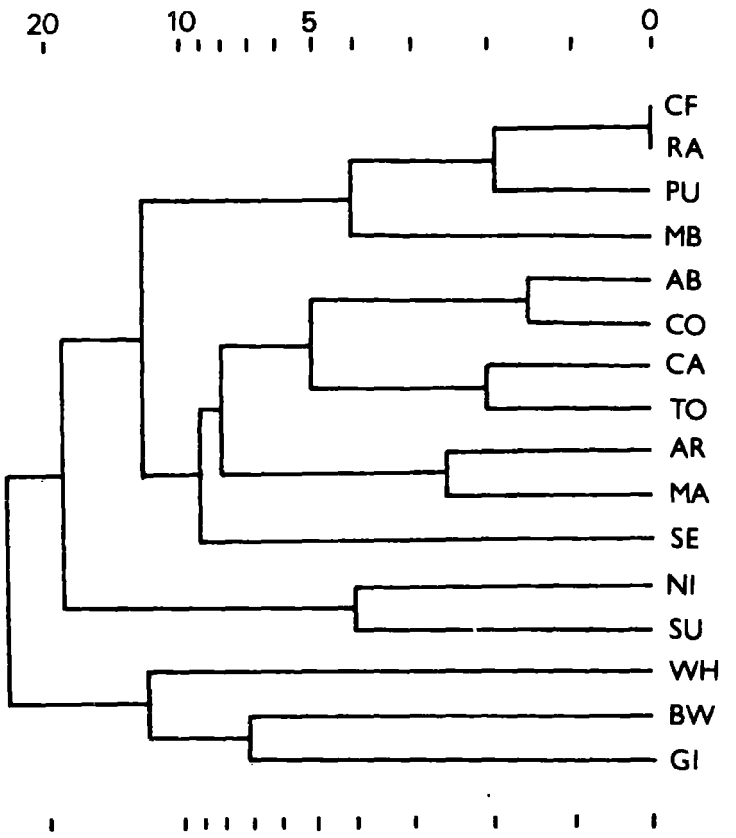

10
5
0
CF
RA
PU
MB
AB
CO
CA
TO
AR
MA
NI
SU
SE
BW
GI
WH

FIGURE 22. A phylogenetic tree relating sixteen plant species.

The tree was constructed by the 'flexible' method of Lance & Williams (1967a) with ß = -0.24 using amino acid differences between the sequences.

FIGURE 23. A phylogenetic tree relating sixteen plant species.

The tree was constructed by the 'flexible' method of Lance & Williams (1967a) with ß = -0.24 using minimum mutation distance between species.


20
10
5
0
CF
RA
PU
MB
AB
CO
CA
TO
AR
MA
SE
NI
SU
WH
BW
GI

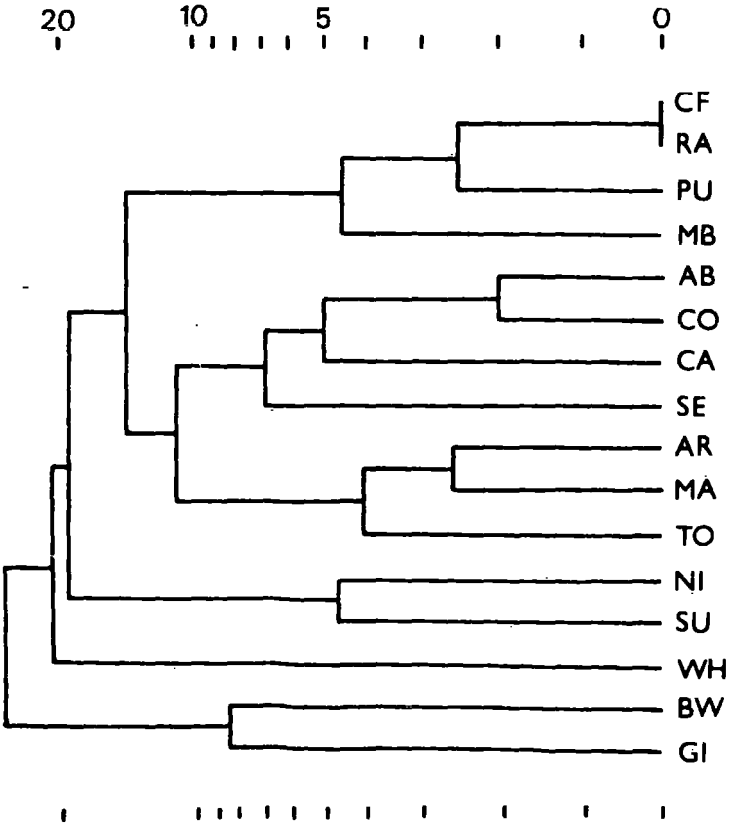

20
10
5
0
CF
RA
PU
MB
AB
CO
CA
SE
AR
MA
TO
NI
SU
WH
BW
GI

FIGURE 24. A series of phylogenetic trees relating sixteen plant species.

The trees were constructed by the 'flexible' method of Lance & Williams (1967a) using minimum mutation distance between species. The values of ß which were used, are indicated.

FIGURE 26. A basic phylogenetic tree relating ten plants constructed using the ancestral sequence method.

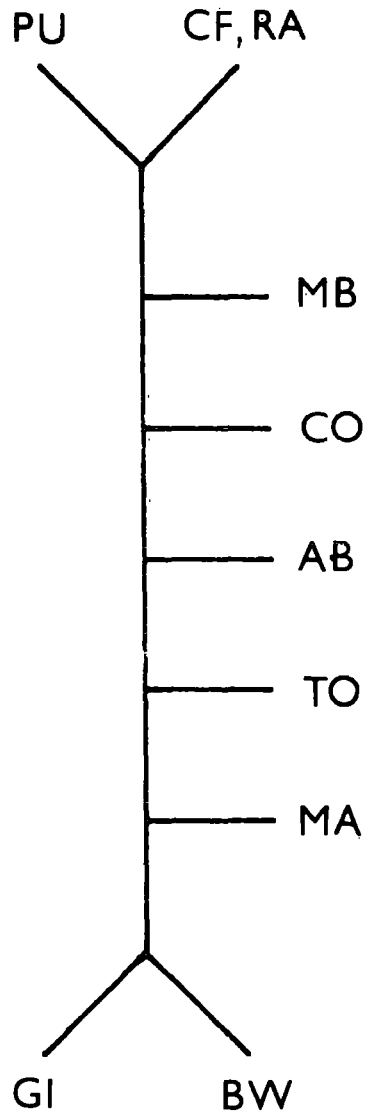
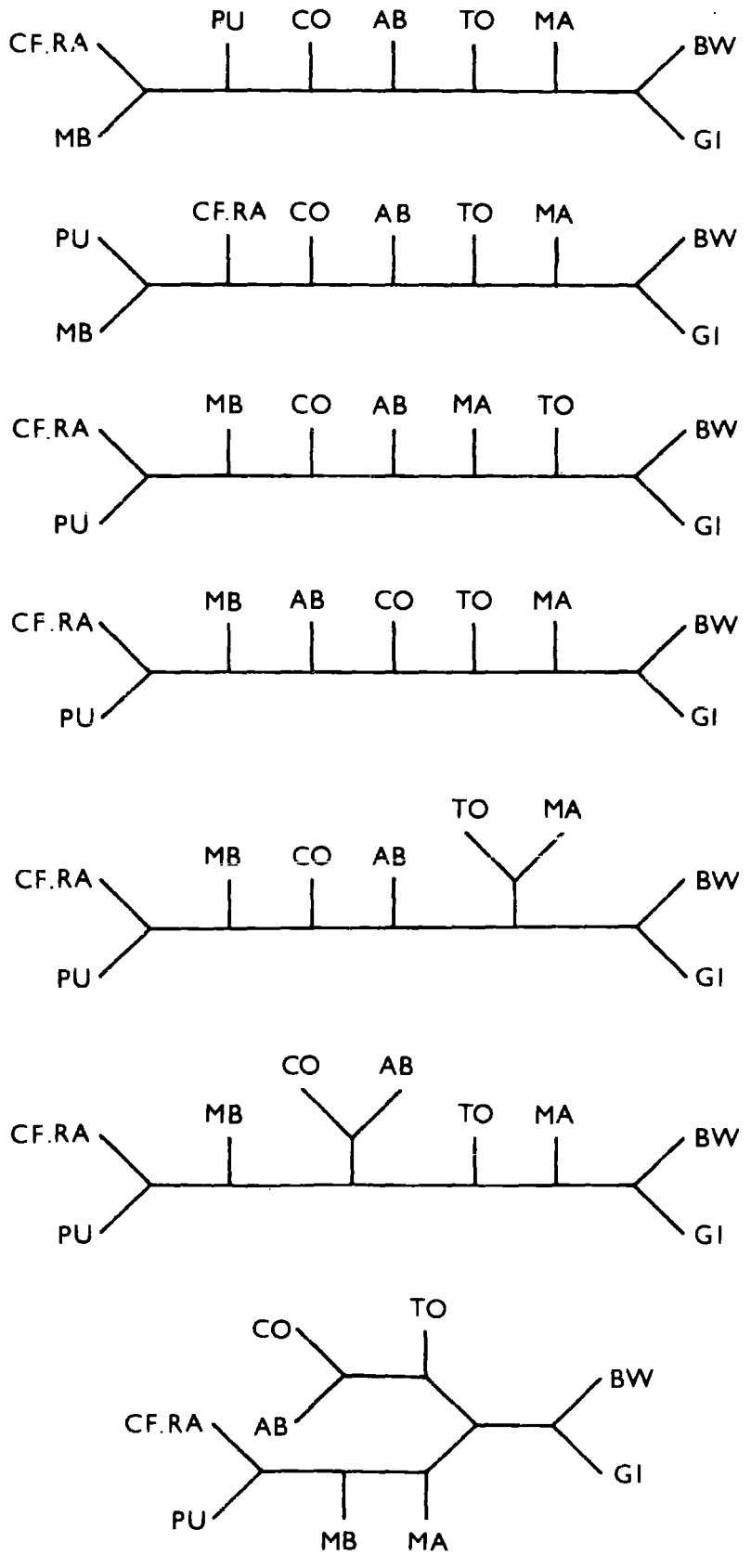


FIGURE 27. Alternative trees obtained from the basic tree when one mutation more than the basic tree minimum value is allowed.



FIGURES 30 (a, b, c). Alternative phylogenetic trees relating sixteen plant species, constructed using the ancestral sequence method.

FIGURE 28. The alternative positions where additional species, when added singly, are located on the basic tree which do not alter the basic tree topology.

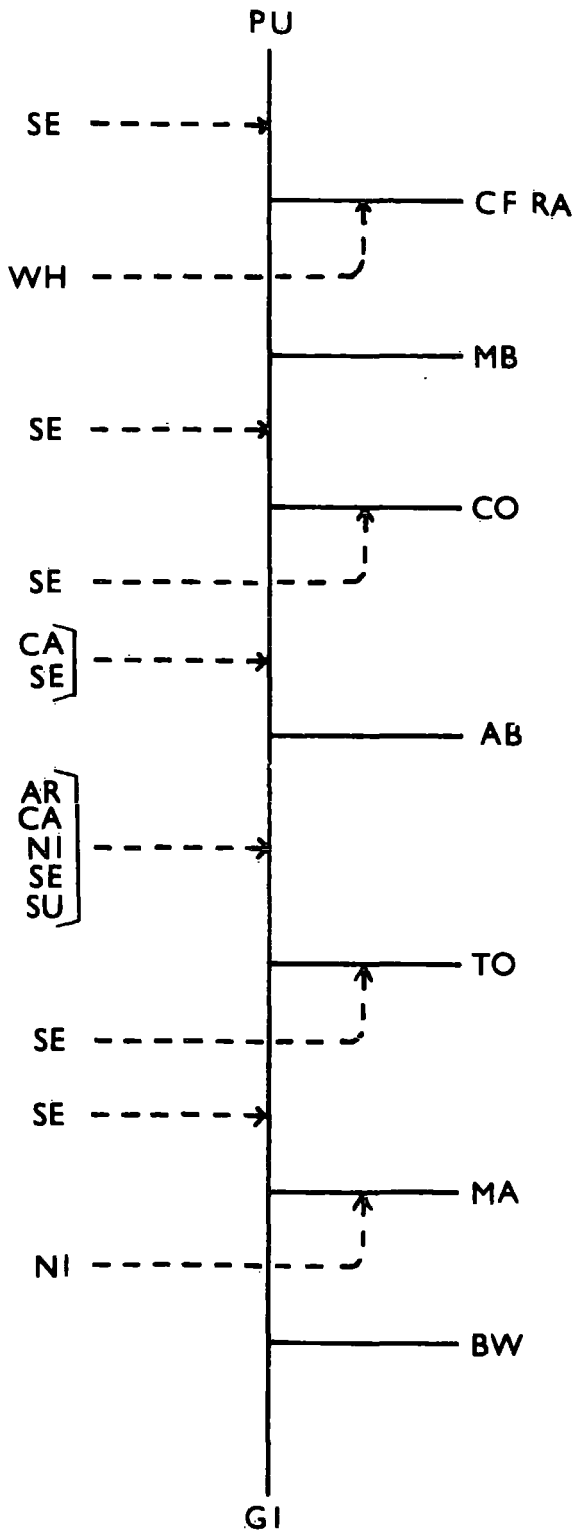
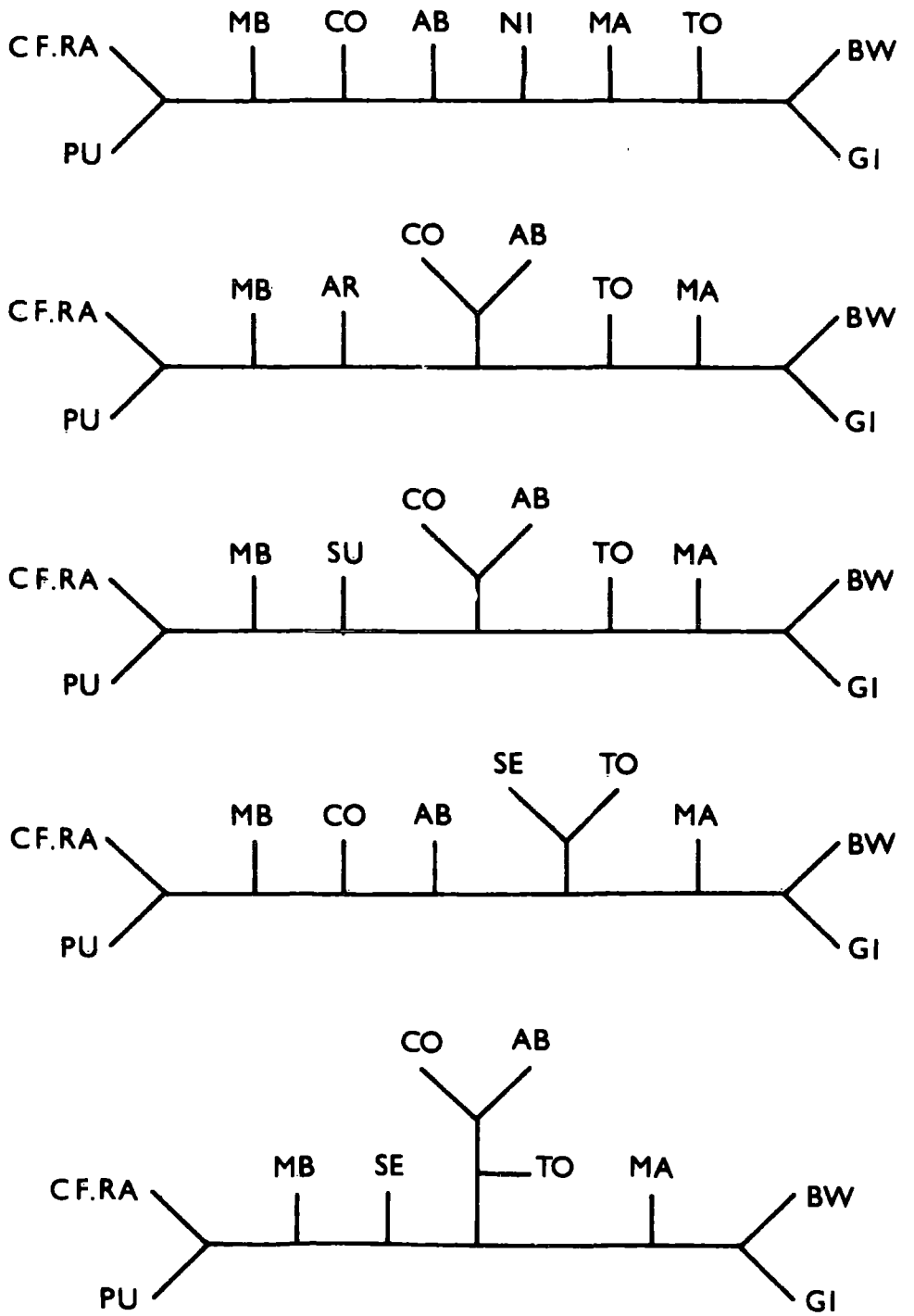
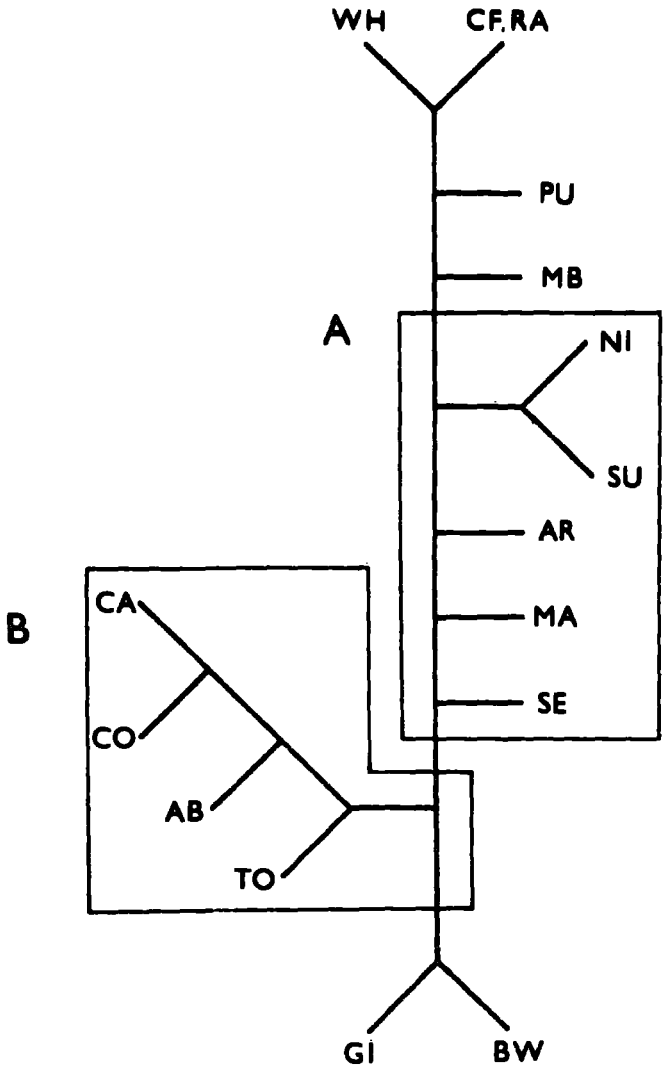
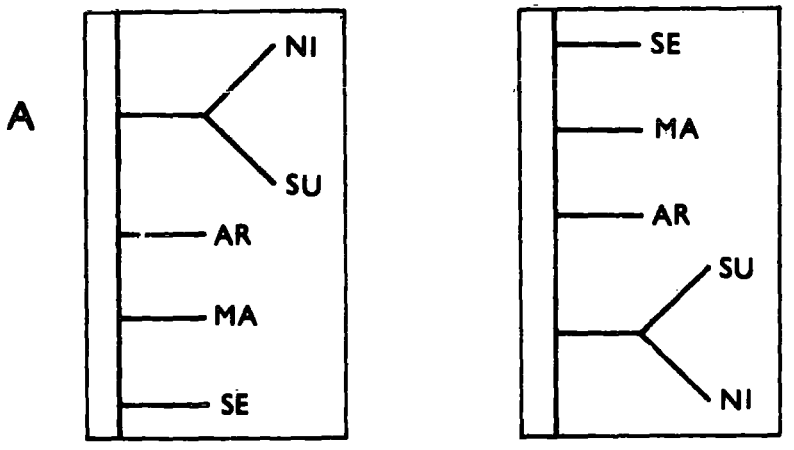
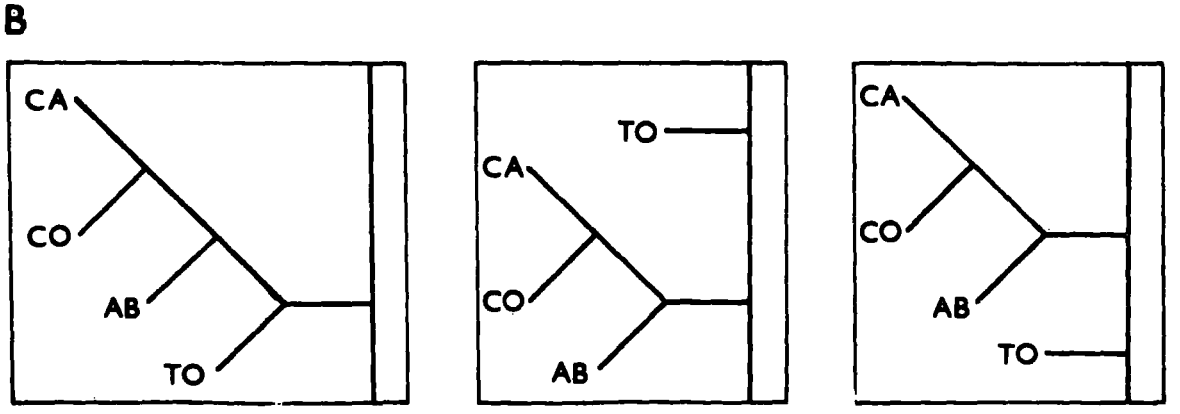


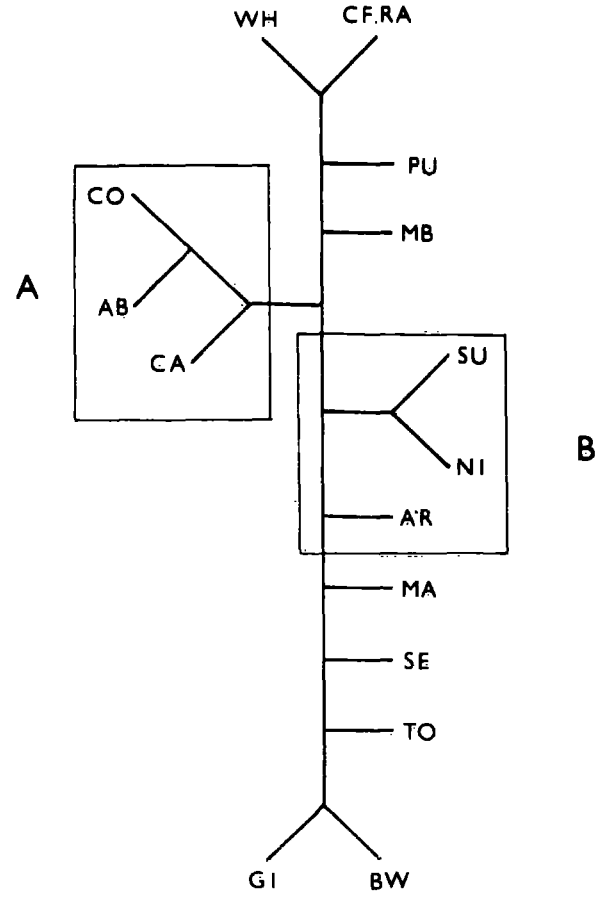
FIGURE 29. The alternative positions where additional species, when added singly, are located on the basic tree which alter the basic tree topology.

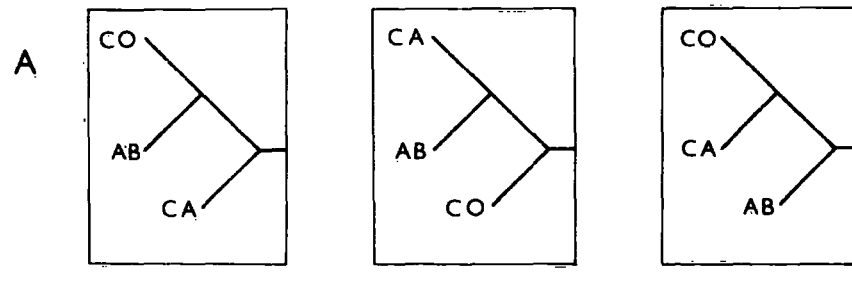


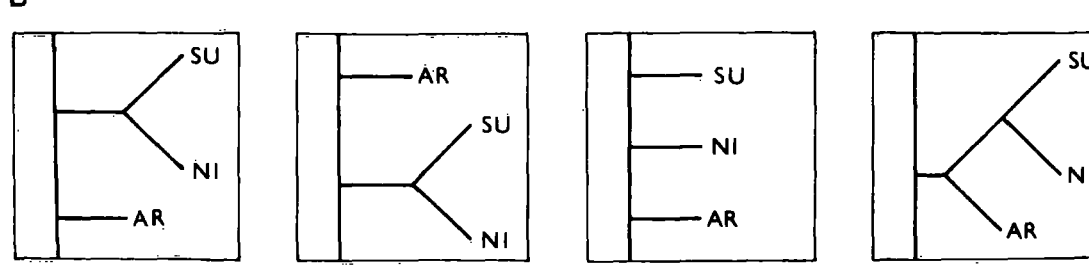

WH
CF.RA
PU
MB
A
NI
SU
AR
MA
SE
B
CA
CO
AB
TO
GI
BW

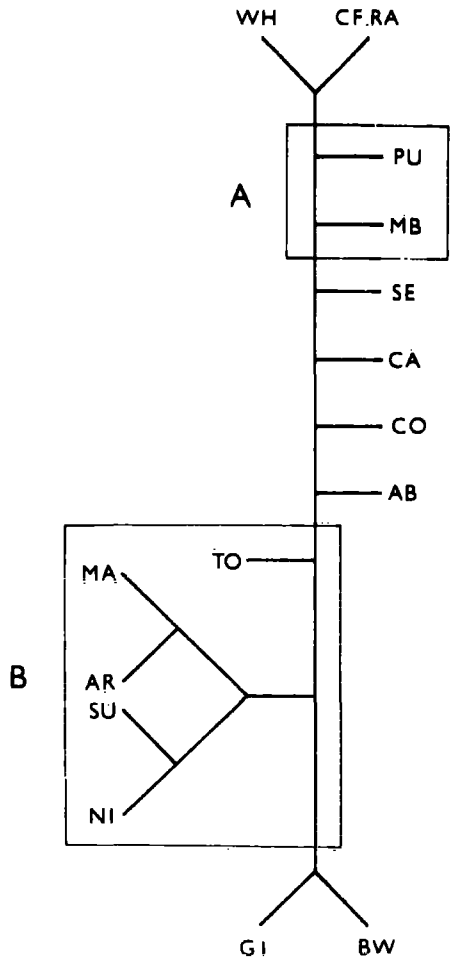

A
NI
SU
AR
MA
SE
SE
MA
AR
SU
NI


B
CA
CO
AB
TO
TO
CA
CO
AB
CA
CO
AB
TO


WH
CF.RA
PU
MB
A
CO
AB
CA
SU
NI
B
AR
MA
SE
TO
GI
BW


A
CO
AB
CA
CA
AB
CO
CO
CA
AB


B
SU
NI
AR
AR
SU
NI
SU
NI
AR
SU
NI
AR


WH
CF.RA
A
PU
MB
SE
CA
CO
AB
B
TO
MA
AR
SU
NI
G1
BW

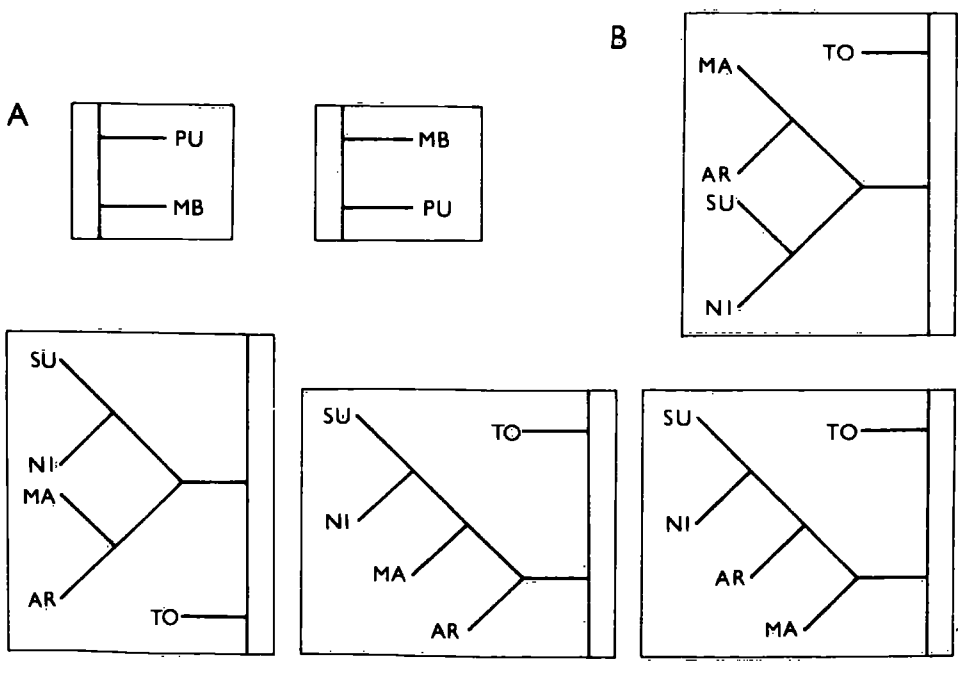

A
PU
MB
MB
PU
B
TO
MA
AR
SU
NI
SU
NI
MA
AR
TO
SU
TO
NI
MA
AR
SU
TO
NI
AR
MA

FIGURE 31. A phylogenetic tree relating fifty species.

The animal and fungal topology used was based on Dayhoff (1969); sequences not included or more recent have been incorporated into the topology. The plant topology used was one of the best alternatives obtained for the sixteen species (see Fig. 30(a)).

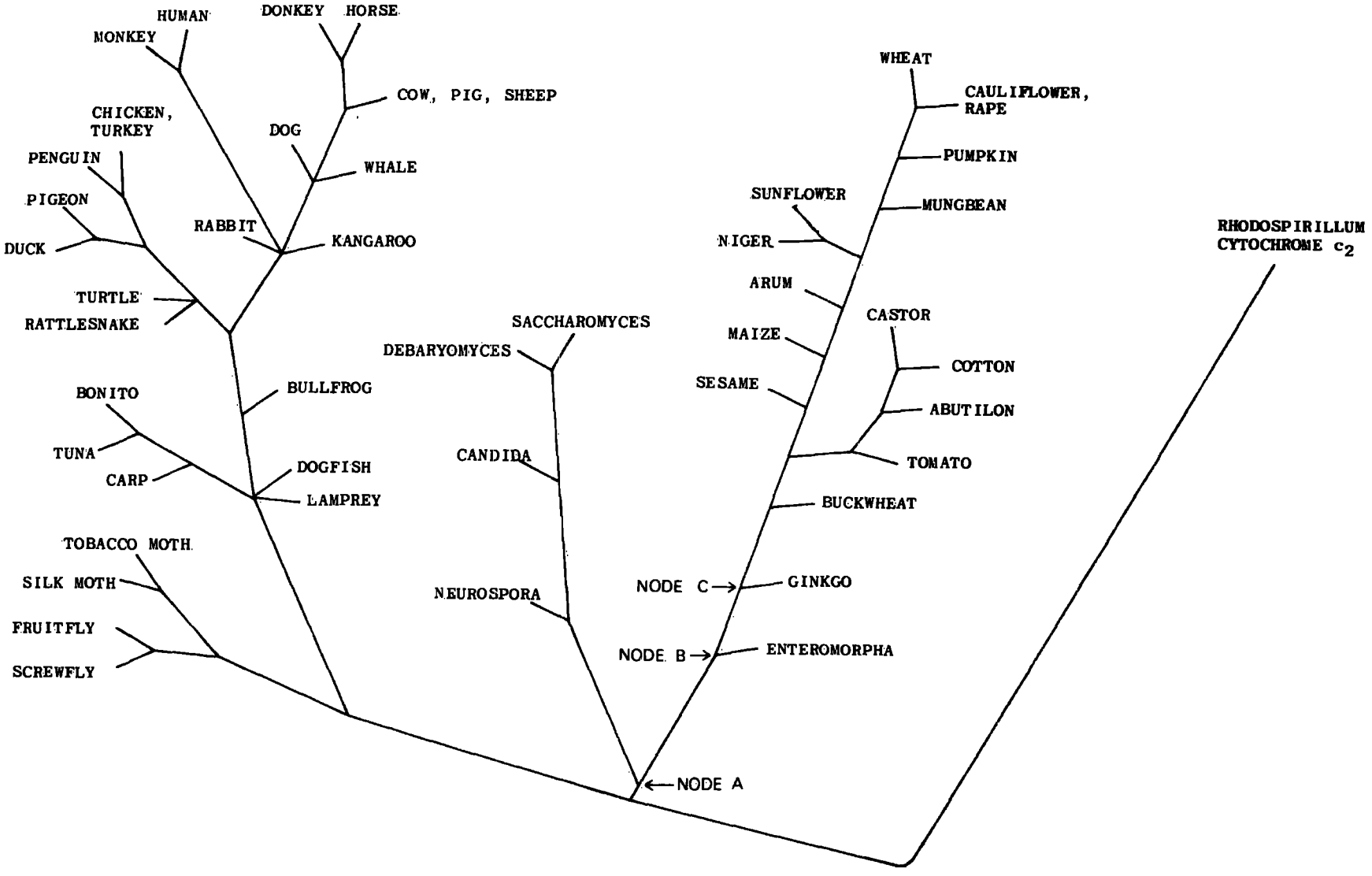
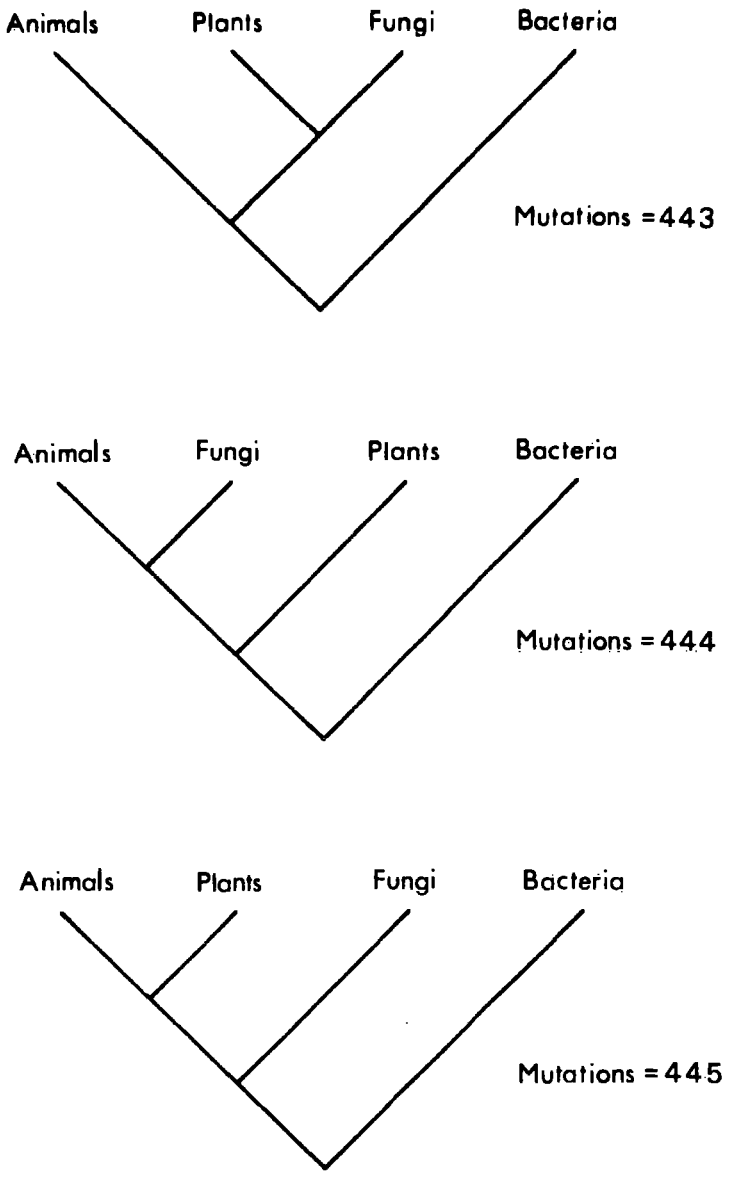

HUMAN
MONKEY
DONKEY
HORSE
COW, PIG, SHEEP
CHICKEN, TURKEY
DOG
WHALE
PENGUIN
PIGEON
DUCK
RABBIT
KANGAROO
TURTLE
RATTLESNAKE
BULLFROG
BONITO
TUNA
CARP
DOGFISH
LAMPREY
TOBACCO MOTH
SILK MOTH
FRUITFLY
SCREWFLY
SACCHAROMYCES
DEBARYOMYCES
CANDIDA
NEUROSPORA
WHEAT
CAULIFLOWER, RAPE
PUMPKIN
SUNFLOWER
MUNGBEAN
NIGER
ARUM
CASTOR
MAIZE
COTTON
SESAME
ABUTILON
TOMATO
BUCKWHEAT
NODE C
GINKGO
NODE B
ENTEROMORPHA
NODE A
RHODOSPIRILLUM CYTOCHROME c_2

FIGURE 32. The alternative ordinations of the tree kingdoms.



The 'mutation' values given are the scores obtained when the three alternative trees were evaluated using the ancestral sequence method. The sequences used are those given in Appendix 3. The topology of the animal kingdom was based on that given by Dayhoff (1969), and the plant topology is that given in Fig. 30(a). The best tree is shown in Fig. 31.

FIGURE 33.

Reconstructed plant ancestral cytochrome c sequences

```
                                                                  10
NODE  'A'                    ? - ? - ? - ? -Ala-Gly-Asp-Ala- ? -
NODE  'B'     ? -Thr-Phe- ? - ? -Ala-Pro-Pro-Gly-Asp-Pro- ? -
NODE  'C'    Ala-Thr-Phe-Ser-Glu-Ala-Pro-Pro-Gly-Asp-Pro-Lys-

                                  20
Lys-Gly-Ala-Lys-Ile-Phe-Lys-Thr-Lys-Cys-Ala-Gln-Cys-His-Thr-
Lys-Gly-Ala-Lys-Ile-Phe-Lys-Thr-Lys-Cys-Ala-Gln-Cys-His-Thr-
Ala-Gly-Glu-Lys-Ile-Phe-Lys-Thr-Lys-Cys-Ala-Gln-Cys-His-Thr-

         30                                                  40
Val-Glu-Lys-Gly-Gly- ? -His-Lys-Val-Gly-Pro-Asn-Leu-His-Gly-
Val-Glu-Lys-Gly-Ala-Gly-His-Lys-Gln-Gly-Pro-Asn-Leu- ? -Gly-
Val-Glu-Lys-Gly-Ala-Gly-His-Lys-Gln-Gly-Pro-Asn-Leu- ? -Gly-

                                  50
Leu-Phe-Gly-Arg-Thr- ? -Gly-Gln-Ala-Gly-Pro-Tyr-Ser-Tyr-Ser-
Leu-Phe-Gly-Arg-Thr-Ser-Gly-Thr-Ala-Ala-Gly-Tyr-Ser-Tyr-Ser-
Leu-Phe-Gly-Arg-Gln-Ser-Gly-Thr-Thr-Ala-Gly-Tyr-Ser-Tyr-Ser-

         60                                                  70
Asp-Ala-Asn-Lys- ? -Lys-Gly- ? -Thr-Trp- ? -Glu-Asn-Thr-Leu-
Ala-Ala-Asn-Lys-Asn-Lys- ? - ? - ? -Trp- ? -Glu-Asn-Thr-Leu-
Ala-Ala-Asn-Lys-Asn-Lys-Ala-Val- ? -Trp-Gly-Glu-Asn-Thr-Leu-

                                  80
Phe-Glu-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr-Ile-Pro-Gly-Thr-Lys-
Tyr-Glu-Tyr-Leu-Leu-Asn-Pro- * -Lys-Tyr-Ile-Pro-Gly-Thr-Lys-
Tyr-Glu-Tyr-Leu-Leu-Asn-Pro- * -Lys-Tyr-Ile-Pro-Gly-Thr-Lys-

         90                                                 100
Met-Val-Phe-Ala-Gly-Leu-Lys-Lys- ? -Gly-Asp-Arg-Ala-Asp-Leu-
Met-Val-Phe-Ala-Gly-Leu-Lys-Lys-Pro-Gln-Asp-Arg-Ala-Asp-Leu-
Met-Val-Phe-Pro-Gly-Leu- * -Lys-Pro-Gln-Glu-Arg-Ala-Asp-Leu-

                                 110
Ile-Ala-Tyr-Leu-Lys- ? -Ala-Thr-Ala
Ile-Ala-Tyr-Leu-Lys- ? -Ala-Thr-Ala
Ile-Ala-Tyr-Leu-Lys- ? -Ala-Thr-Ala
```

LEGEND: The sequences given are the reconstructed ancestral sequences for the early plant nodes as labelled in Fig. 31.

*Synthesised as Lys; subsequently trimethylated.

DISCUSSION

Keilin (1925, 1927) showed that cytochrome c was distributed not only among members of the animal kingdom, as had been demonstrated by MacMunn (1884, 1886, 1887), but was also present in plants. He showed its presence in 'monocotyledonous bulbs, in beans and in pollen from several different flowers' (Keilin, 1927). At the beginning of this investigation cytochrome c of plant origin had been prepared only from wheat (Goddard, 1944; Hagihara, Tagawa, Morikawa, Shin & Okunuki, 1958, 1959; Wassermann, Garver & Burris, 1963), corn pollen (Potgieter, 1964), soybean (Fridman, Lis, Sharon & Katchalski, 1968) and rice (Morita & Ida, 1968; Ida & Morita, 1969). This work had shown that the yields of cytochrome c from plants were very low, ~1 mg/Kg, especially when compared with the yields obtained from animal sources, ~ 250 mg/Kg (see Margoliash & Schejter, 1966).

The methods which have been used for the preparation of cytochrome c during this investigation were varied with the sources with the aim of improving the purity and yield. The different groups of methods employed were combinations of techniques previously used during other cytochrome c preparations (see Margoliash & Schejter, 1966).

The plant tissue chosen for extraction was dark germinating seedlings, since seeds are readily available, can be stored, and are biochemically conservative. The seeds were germinated in the dark to obtain maximal metabolic activity while avoiding the complications which may arise during purification steps if photosynthetic cytochromes were also present. The yields of cytochrome c obtained when seeds of low percentage germination or of low metabolic activity are used, are much lower (Richardson,

Richardson, Ramshaw, Thompson & Boulter, 1971b), as is also the case when mature or green tissues are used (Richardson et al., 1971b; Asada & Takahashi, 1971).

In previous investigations rather drastic methods have generally been used when extracting cytochrome c from plant sources, such as treatments with organic solvents (Goddard, 1944; Hagihara et al., 1958, 1959; Wassermann et al., 1963; Fridman et al., 1968) or with hot acid (Goddard, 1944). Initially it was considered that such drastic treatments did not matter as cytochrome c appeared to be very stable, even when subjected to extreme conditions. It has now been shown however, that the inhomogeneities of most preparations result from drastic treatments, particularly those during the initial extraction (Margoliash, 1954b; Nozaki, Yamanaka, Horio & Okunuki, 1957; Yamanaka, Mizushima, Nozaki, Horio and Okunuki, 1959). The use of organic solvents can cause denaturation (Kaminsky & Davison, 1969), while extremes of pH or high temperature may lead to denaturation, deamidation or polymer formation (see Margoliash & Schejter, 1966). As extraction of plant cytochrome c by blending tissue in dilute phosphate buffer had been shown to give an efficient extraction (Morita & Ida, 1968; Ida & Morita, 1969), a similarly mild technique was employed in this investigation. The tissues were extracted at pH 4.5-4.6 in the presence of ascorbic acid and EDTA. The acid pH precipitated many of the plant storage proteins still present in the germinating seed; if the pH was varied much the subsequent filtration was either very slow and/or gave cloudy filtrates. The ascorbic acid and EDTA were added to reduce oxidative destruction of the cytochrome by compounds

produced in the browning reaction (Wassermann et al., 1963). Filtrates were adjusted to pH 8 as quickly as possible, in order to avoid modification of the protein by the acid pH.

Because of the low yield of cytochrome from plants, a major problem was the initial concentration of the protein. Yamanaka, Mizushima & Okunuki (1964a) showed that the ability of cation exchangers to adsorb cytochrome c from solution gave a particularly efficient method of concentrating extracts permitting the preparation of the protein from very poor sources. With plant sources the presence of large quantities of material which precipitated in and on the resin often made the cytochrome band invisible and necessitated batchwise elution of the resin. However, this method still represented the best way of concentrating the large initial volumes from the extraction. During the elution it was important to maintain the alkaline pH since below pH 6 the cytochrome becomes very strongly bound to the resin (Boardman & Partridge, 1953, 1954, 1955). Alternative procedures used for the initial concentration of cytochrome c from plant sources have been the precipitation of the total soluble protein by ammonium sulphate (Asada & Takahashi, 1971) and the preparation by zonal rotors of mitochondria, from which the cytochrome was then extracted (Diano & Martinez, 1971). The batchwise elution of the cation exchanger, which was necessary because of the large quantities of precipitation which occurred, still gave fairly large volumes of eluate. Thus, a further concentrating step was necessary, e.g. another ion-exchanger, which lead to a logistic dialysis problem. CM-Sephadex proved the most useful next concentration resin, since the exclusion properties often gave rise to a significant

purification and the flow rates were better compared with CM-cellulose resins. Even after this concentration, care was still needed in order to avoid modifying the protein. For this reason ammonium sulphate was initially avoided since Flatmark (1966) had suggested that this led to deamidation. Repeated molecular exclusion chromatography was shown to be a suitable alternative (Flatmark, 1964a), and so this milder approach was used instead. It has been suggested however, that the inhomogeneity attributed to the ammonium sulphate fractionation, was due either to the initial extraction or poor control of pH during the ammonium sulphate step (Margoliash & Schejter, 1966). It was found that carefully controlled ammonium sulphate fractionation of plant cytochrome c did not give rise to major quantities of deamidated or modified cytochrome. The inclusion of this step gave a useful concentration and purification of the protein. Subsequent gel-filtration then provided further purification and removed the remaining ammonium sulphate from the pellet. At this stage selection of the best fraction was used; in all previous steps the total preparation was maintained. In the later preparations, 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 oxidized by potassium ferricyanide. The deamidated forms are bound less strongly and leave the oxidized protein as the last monomer eluted. Elution by linearly increasing cation concentration was used (Margoliash, 1962; Margoliash & Lustgarten, 1962), rather than the alternative method, that of constant cation concentration and increasing pH (Boardman, 1959; Margoliash, 1962; Margoliash & Lustgarten,

1962). This technique will separate deamidated forms and modified forms which are not separated by gel-filtration, but difficulties still remain with aggregate forms, e.g. histone/ cytochrome complexes (Margoliash, 1964a). This last purification step provided a useful examination of the degree of modification which had been caused by the earlier procedures. The results showed that inclusion of the ammonium sulphate precipitation step did not lead to any additional modifications occurring, and the degree of modification generally present was low.

As an alternative to gradient elution for a final purification step, isoelectric focussing was used. It was shown (Flatmark & Vesterberg, 1966) that with beef cytochrome preparations, modified forms were well separated by this technique. However, in this separation the purity of the protein loaded onto the column was already high, so that the overall loading was well within the limits of the gradient. The preparations available for loading during the plant cytochrome preparation were generally not sufficiently pure to allow a reasonable loading of cytochrome without overloading the column with impurities; if this overloading occurred, focussing could not be achieved because of the disturbances caused by precipitating proteins. Although the observed pI of plant ferrocytochrome c is lower than that of animals, (pI = 10.80 @ 4 °C for beef cytochrome c (Flatmark & Vesterberg, 1966), pI = 9.6 @ 4 °C for mungbean cytochrome c (Laycock, 1968)), it was still sufficiently high to be close to the edge of the ampholine range giving danger of losses into the electrode solution. Additional basic compounds were added to minimise this effect, but when the gradient was disturbed through precipitation

extreme caution had to be taken to avoid losses of cytochrome. However, the results using this technique, even though the cytochrome was never completely focussed, indicated a virtual complete absence of any other 'cytochrome' bands, showing that the techniques employed were mild and did not give appreciable amounts of modified product. This was verified by polyacrylamide gel-electrophoresis on various plant preparations at both pH 4.3 and pH 8.3 (Valentine & Roberts, 1971); this work also confirmed plant cytochrome c pI values of about pH 9.6.

Dialysis was avoided when possible, and particularly for the final desalting, as plant cytochrome c showed a tendency to bind to the dialysis sack at low salt concentrations (Laycock, 1968), and because at low ionic strength wheatgerm cytochrome c is reported to be unstable (Wassermann, Garver & Burris, 1963). The problem in dialysing the sucrose gradient after isoelectric focussing, was another disadvantage of this technique compared with gradient elution.

The crystallization of cytochrome c from wheat (Hagihara et al., 1958; Hagihara et al., 1959) and from rice (Morita & Ida, 1968; Ida & Morita, 1969) has been reported; this appeared easier than with animal cytochromes c (Ida & Morita, 1969). No attempts were made however, to crystallize any of the plant cytochromes purified during this investigation.

Obviously no single method for preparation is ideal when such a wide variety of sources is available. Generally, however, it is best to have a basic method established which is available to follow in principle.

The sunflower preparation contained an impurity, the nature

of which was uncertain. It was certainly very rich in glycine, and from the tryptic digest peptides probably contained in excess of 70% glycine. It did not appear to be digested by chymotrypsin and thus, the apparent lack of residues susceptible to chymotrypsin, i.e. tyrosine, phenylalanine and tryptophan, accounts for it not drastically altering the purity determined by spectral means. The amounts of impurity present varied over the four bands eluted by the gradient, suggesting it is bound in simple ratios in each. However, on gel-filtration only a single cytochrome band was observed, so either the probable molecular weight may be quite low, or, alternatively, exceeded the exclusion limit of the gel. The lack of electrophoretic mobility at pH 6.5 after the chymotryptic digest, may be due either to a large molecular weight, or carbohydrate causing binding to the paper. No analysis was made for carbohydrate content. Alternatively, if it was not bound to the cytochrome c during the preparation, the lysine groups present would make it basic and it could by chance have run with the cytochrome throughout the preparation; however, it is then difficult to understand the presence of four major bands off the gradient. Artifacts to preparations of this nature, i.e. cytochrome and basic protein, have been reported (Margoliash, 1964a), but the impurity was tentatively assigned as histone. Other examples are of cytochrome bound to globin (Margoliash, 1954a) in stoichiometric quantities, although the identification of non-cytochrome component as globin has been doubted (Henderson & Rawlinson, 1961; Henderson & Paleus, 1963; Henderson & Ada, 1963), and the non-cytochrome component suggested as a glycoprotein containing glucose and sialic acid (Henderson & Ada, 1963).

The yields and purities of the cytochrome c obtained from the plants used in this investigation, are generally comparable with those from other plant sources. In Table 20 they are compared with some of the other cytochrome c preparations which have been reported. The spectral ratio of 416 (reduced) nm/ 550 (reduced) nm obtained for plants is noticeably higher than that for animal cytochromes. The yields from seedlings are clearly better than those from mature material.

The purification methods for fungal cytochromes which have been used, are basically the same as those used for animal and other cytochromes (see Margoliash & Schejter, 1966). The major problems are involved with the initial extraction of the protein. Much of the early work on the fundamental properties of cytochromes was undertaken using delft yeast preparations (Keilin, 1930; Hill & Keilin, 1930; Dixon, Hill & Keilin, 1931); plasmolysis by sodium chloride and boiling water followed by precipitation by sulphur dioxide was used to extract the protein. The lack of sensitive purity criteria lead to erroneous conclusions, based on these and other results, that drastic extraction and purification conditions could be used without modifying the protein. However, because fungal cytochromes are more difficult to extract, recent preparative procedures have still needed drastic treatment during the extraction, usually by organic solvent autolysis (Nozaki et al., 1957; Sherman et al., 1965; Shirasaka, Nakayama, Endo, Hameishi & Okunuki, 1967).

The cells available for this investigation were obtained direct from a large industrial fermentation. Little or no information was openly available on the nature of media or conditions during the fermentation. Cells were taken at about

TABLE 20

The spectral properties and yields of various plant cytochromes c

	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
Mung bean	0.91	1.18	4.0	5.3	0.40
Sunflower	0.92	1.20	3.9	5.0	0.50
Niger	1.05	1.22	4.5	5.2	0.40
Black gram			2.5		0.10
Wheatgerm (Stevens et al., 1967)	1.16	1.20	3.8	4.4	
Soyabean (Fridman et al., 1968)	1.0	1.20	4.0	4.5	0.35
Rice (Ida & Morita, 1969)	0.96	1.30	4.0	4.7	0.34
Ginkgo biloba (Richardson et al., 1971b)	1.06	1.20	4.4	5.2	0.12
Arum maculatum (shoots) (Richardson et al., 1971b)	0.95	1.20	4.0	5.3	0.09
Brassica oleracea (inflorescences) (Richardson et al., 1971b)	1.22	1.22	4,9	5.0	0.20
Spinach (leaves) (Asada & Takahashi, 1971)			4.0	4.7	0.07
Potato (tubers) (Diano & Martinez, 1971)			4.0	5.0	0.20
Horse heart (Margoliash & Frohwirt, 1959) (Margoliash & Walasek, 1967).	1.20	1.20	4.6	4.4	250

60 h since at this time the maximum cell density occurred while the rate of fall of dissolved oxygen in medium was at a maximal level; it was assumed that the respiration would be high at this stage during the fermentation.

Direct autolysis methods (Sherman et al., 1965; Sherman et al., 1968) were attempted but were not successful, probably due to the large oil quantity present from the medium, such that insufficient organic solvent was being added.

Acetone powders of the cells were prepared which were successfully extracted. This procedure was preferred because of the commercial value of the fresh cells which left the fermenting unit.

The initial cytochrome yields appeared very large, 30-50 mg/Kg of acetone powder, but were considerably lower than those obtained from other fungal sources; 100-400 mg/Kg fresh cells for Neurospora (Heller & Smith, 1966), 400 mg/Kg wet cells for Candida, and 100 mg/Kg wet cells for Bakers yeast (Shirasaka et al., 1968). However, in these cases the conditions for growth, the media and the harvesting, were all selected so as to give maximal yields of cytochrome. The high yields obtained allowed direct elution from the amberlite, which was used to initially concentrate the extract.

The product was finally purified using gradient elution, giving multiple bands and not a single product. Only one of these bands, the last to be eluted, was considered from its spectral characteristics to represent native cytochrome c. The other bands were all considered to represent modified or denatured forms. The major band was certainly not a normal native mitochondrial c-type cytochrome, as judged by its absorption

spectrum and lack of reduction by ascorbic acid. Denaturation had probably occurred as a result of the initial acetone treatment. The other components which were found, may represent isoenzymes. In Saccharomyces, two different cytochromes c exist (Slonimski, Acher, Pere, Sels & Somlo, 1965), both of which are present in the same cell. The quantity of each is dependent on the growth conditions (Slonimski et al., 1965; Sels, Fukuhara, Pere & Slonimski, 1965), so it is possible that a similar situation may exist in another ascomycete, but not be readily observed. Only one cytochrome c has been reported for Debaryomyces (Narita & Sugeno, 1968). However, it should be borne in mind that the situation with Cephalosporium may not be typical, because of the immense amount of mutant selection work done during the past ten years on this commercial strain. It is difficult to assess the effect this may have had, if a redundant copy of the cytochrome existed within the genetic material. It would be possible to check fairly simply, if the two were different, by examining either amino acid compositions or peptide maps. These methods will only indicate differences and cannot be used to show the two to be the same, e.g. the cytochromes c from Kloechera sp. and Candida krusei have quite different properties (Yamanaka, Mizushima, Katano & Okunuki, 1964b), presumably as a result of differing sequences yet give the same amino acid compositions (Narita, Murakami & Titani, 1964).

The sunflower and Cephalosporium preparations particularly showed the inadequacies of assessing purities from spectral ratios. The variation in the spectral ratios between apparently equally pure samples of the same cytochrome has been shown (Paleus &

Neilands, 1950; Paleus, 1960). The absorbances of the soret band and the 695 nm band are susceptible to modifications in tertiary structure, the former becoming larger as the protein unfolds, while the latter disappears (Stellwagen, 1968). The absorption at 280 nm varies with amino acid composition being particularly susceptible to changes in tryptophan content, but also tyrosine and phenylalanine to a lesser extent. This means no absolute purity criterion based on spectral ratios can be applied to all species. The use of iron content as an estimate of purity has been used. Again, apparently identically pure samples did not give consistent values on iron analysis (Paleus & Neilands, 1950; Paleus, 1960; Bodo, 1955). Although this method will indicate protein purity, it cannot distinguish between native and modified cytochrome. Analytical column chromatography on weak cation exchangers is considered to be a good criteria for purity (Margoliash & Schejter, 1966), especially in conjunction with carbon monoxide combination with fractions (Margoliash, 1962). Similarly, analytical electrophoresis on media, such as starch gel or polyacrylamide gel, provides a good purity criterion. Ideally, any method used should be based on other than direct iron content (see for examples; Keilin & Hartree, 1939; Massey, 1959; van Gelder & Slater, 1962).

Large quantities of cytochrome c have been used in the sequence determinations previously described, e.g. Samia Moth, 1.30 gm (Chan & Margoliash, 1966); Rabbit, 1.71 gm (Needleman & Margoliash, 1966); Dog, 0.82 gm (McDowall & Smith, 1965); Wheat, 0.26 gm (Stevens et al., 1967). If the same quantities of plant cytochrome were to be prepared as were used in the

sequence determination of wheat cytochrome c, in which only chymotryptic peptides were sequenced, then very considerable quantities of starting material would be required. As the aim of the investigation was to examine a large number of widely distributed species, the problems of obtaining starting materials make such an approach unfeasible. The quantities of cytochrome c available from the preparations undertaken here, <50 mg, require a micro-method approach to sequencing different to the direct or subtractive Edman methods previously used in other investigations of cytochrome c sequences.

The strategy chosen was that devised by Gray & Hartley (1963a); this was paper electrophoretic separation of peptides derived from proteolytic enzyme digestion of the protein and sequence analysis by the dansyl-Edman method; in conjunction with using the homology between sequences as little as 0.5 μmol cytochrome (6 mg) was required to determine a sequence.

The proteolytic enzymes used in this investigation were chymotrypsin and trypsin. Of the enzymes available, these are the most specific and have been used successfully in the analysis of other cytochromes c (see for example; Chan & Margoliash, 1966; Needleman & Margoliash, 1966; McDowall & Smith, 1965; Stevens et al., 1967). The two sets of peptides obtained give sufficient overlaps usually to logically establish the sequence.

The plant cytochromes c were all readily digested with both enzymes with specificities generally as expected. Chymotrypsin, however, occasionally showed some rather broader specificity breaks, e.g. at asparagines. Trypsin showed very few unexpected breaks; those which did occur may be due to

pseudotrypsin in the enzyme preparation, rather than unusual specificity (Keil-Dlouha, Zylber, Imhoff, Tong & Keil, 1971). No tryptic activity was observed at either ϵ-N-trimethyllysine residue.

High-voltage paper electrophoresis was used to separate peptides. A flat-plate apparatus (Gross, 1961) was used as opposed to tanks (Michl, 1951), as there was less fire hazard and a greater length for separations was available. The use of paper techniques is much preferable to ion-exchange column methods, when only small quantities of material are available. Paper chromatography was not found to be needed however, as electrophoresis at pH 6.5 and pH 1.9 gave an adequate separation of the peptides.

The dansyl-Edman method for sequencing peptides is extremely sensitive and has been used in a micro form to determine the sequence of as little as 10 pmol of peptide (Bruton & Hartley, 1970). When the dansyl derivatives are separated by polyamide thin-layer chromatography (Woods & Wang, 1967), it also has the advantage of excellent resolution of all normally occurring amino acid derivatives, especially leucine and isoleucine, but also of unusual amino acids such as ϵ-N-trimethyllysine; a 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 residues whose derivatives are labile during the acid hydrolysis of the labelled peptide. Thus, both asparagine and glutamine are destroyed to the corresponding acids; 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. After several steps of Edman degradation lysine can also be difficult to identify. The technique is also

only qualitative and is not quantified step by step during the Edman degradation.

Of these difficulties only the identification of amide residues presents a major problem. The electrophoretic mobility of the peptides at pH 6.5 are often sufficient to assign most of the amide residues (Offord, 1966). This is an additional advantage of this separation technique. However, certain amide residues remain unplaced; if sufficient peptide is available the mobilities of degraded peptides at pH 6.5 can be used to identify these residues. Otherwise, they must be tentatively assigned by homology with other cytochromes where the amide residues have been experimentally determined.

There are several characteristic features of the plant cytochrome c sequences determined in this investigation and the others (see Appendix 3) when they are considered as a 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". Between all members of the plant group 85 residues are identical in their positions in the sequences, and of these 15 are found in these positions only in the plant sequences. All contain two residues of the unusual amino acid ϵ-N-trimethyllysine in positions 80 and 94. It is assumed that this represents a secondary modification of the protein after synthesis is complete, as has been shown for the single residue of this amino acid present in Neurospora (Scott & Mitchell, 1969). The number of proline residues in the plant sequences is higher than for the animal cytochromes, while the numbers of basic amino acids is

lower. The latter explains the lower observed pI values for plant cytochromes (Laycock, 1968; Valentine & Roberts, 1971). The sequences determined in this investigation were unique and no positive evidence of heterogeneity was observed. This has not always been the case with the other plant cytochromes c and heterogeneity was observed in castor, sesame (Thompson et al., 1970), pumpkin (Thompson et al., 1971a) and Rape (Richardson et al., 1971a). This may be due to any of four different possible causes; (1) heterozygosity of the allele specifying the polypeptide at one or more loci; (2) intergenic (inter-cistronic differences of the gene specifying the protein either within a population (polymorphism), or between populations; (3) translational ambiguity of the messenger RNA template (Carbon, Berg & Yanofsky, 1966; Rifkin, Hirsh, Rifkin & Konigsberg, 1966); or (4) as an artefact during the course of the purification or sequencing of the protein.

With the exception of the iso-cytochromes -1 and -2 of Bakers yeast (Slonimsky 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 cytochromes c which differ by a single residue (Walasek & Margoliash, 1969). During the determination of the sequence of human cytochrome c leucine was found to replace methionine in position 65 in 10% of the protein (Matsubara & Smith, 1962), but in this case a mixed population of about 70 individuals was used (see King & Jukes, 1969). The various forms of beef cytochrome c reported by Flatmark (1964b) probably represent preparation artefacts. Attempts to demonstrate polymorphism by examining cytochrome preparations from about 12 individual horses and

from 18 humans, both proved unsuccessful (Margoliash, 1969a). The cytochrome from different organs of the same species has also been shown to be identical (Stewart & Margoliash, 1965). Margoliash (1969b) estimated that the sequencing methods which he had used in determining sequences from sources derived from mixed individuals, would detect 5% heterogeneity if it was present in the form of a single change. With the dansyl-Edman method the degree of heterogeneity required for detection is considerably greater than 5%, particularly in the cases of Ginkgo and Niger, where a minimum of material was being used. Changes which result in a change in charge are more likely to be detected, as observed in pumpkin (Thompson et al., 1971a). It is probable that the examples of heterogeneity which have been found in the plant sequences are as a result of using a mixed population as the source. Leucaena ferredoxin sequence, which was also obtained using material from a mixed population, showed heterogeneity at four residues (Benson & Yasunoba, 1969). In all these cases however, one amino acid could be considered to represent the major form in the population at each position.

Apart from the general characteristics certain specific points were encountered in each individual sequence determination.

In the determination of the mungbean sequence, certain of the more sophisticated techniques were not used. This meant that the exact location of the two ϵ-N-trimethyllysine residues, which had been shown to be present, was only tentative and also that the blocking group was assumed to be an acetyl group by homology with wheat cytochrome c (Stevens et al., 1967). Attempts to demonstrate the nature of this block by

carboxypeptidase-A digestion of the N-terminal peptide followed by detection of the acetyl-alanine by the starch-iodide stain (Rydon & Smith, 1952) after paper chromatography failed through lack of material. Two 'impurity' peptides were found, neither of which could be assigned to the sequence. These peptides may reflect the fact that the preparation used for sequence determination was not completely pure.

During the sunflower sequence determination, the presence of a glycine rich impurity to the preparation was found, which prevented a complete amino acid composition determination and the use of a tryptic digest. The use of all the material for chymotryptic digestion enabled all the peptides to be obtained pure; from these peptides it was possible to deduce the complete sequence by homology with other plant cytochromes, even though the overlap peptides were not available. No additional unplaced peptides were found. The techniques for identifying trimethyl-lysine residues and acetyl groups were developed and used in this sequence determination, so that both were positively identified.

Only a very small quantity, 6 mg, of Ginkgo cytochrome was available for its sequence determination. This was the total quantity of pure protein prepared from 50 Kg of seeds (Richardson *et al.*, 1971b). Both chymotryptic and tryptic peptides were examined, but because of the low quantities of material, several residues were not identified in certain of the peptides. However, the two overlapping sets of peptides enabled each residue in the sequence to be identified. Ginkgo cytochrome *c* is the longest cytochrome *c* so far examined, consisting of 113 residues. The additional residues compared

to the other plant cytochromes c, are located at the C-terminus. The Ginkgo sequence has a glycine residue at position 59. In all mitochondrial cytochromes c previously examined, the residue in this equivalent position has been an alanine. This reduces the number of unvaried residues found between all the eukaryotic mitochondrial cytochrome c sequences to 33. Lysine-80 is presumably common to all when the protein is initially synthesised but subsequently modified in certain cases (Scott & Mitchell, 1969).

Because of lower final yields than expected very little cytochrome was available for the sequence determination of niger cytochrome c. The spare material available was used to confirm the sequence of the chymotryptic tryptophan peptide, for which some doubt had existed after analysis of the initial digest. No material was available for an amino acid analysis. The tryptic peptides were very weak and not completely analysed so the overlap available is not extensive. However, every residue in the sequence was identified from one of the digests.

The sequences of the plant cytochromes c showed remarkable similarities among themselves and with the other cytochromes c from eukaryotic mitochondrial sources. Little is known about the cytochromes which function in anaerobic energy-yielding reactions and which occur in chemosynthetic bacteria (Klein & Cronquist, 1967), nor of those which function in the photoreduction processes of photosynthetic organisms. All these contain heme c as the prosthetic group and as such may probably be closely related in origins (Dickerson, 1971); however, with the exceptions of Pseudomonas fluorescens c_{551} (Ambler, 1963), Desulphovibrio vulgaris c_3 (Ambler, 1968), and

Rhodospirillum rubrum c_2 (Dus *et al.*, 1968), sequence determinations have only been done with eukaryotic mitochondrial cytochromes *c*. This group are chemically and physically all very similar, functioning as electron carriers in the mitochondrial respiratory chain (see Margoliash & Schejter, 1966; Keilin, 1966). The reasons for the similar properties were clear when the considerable similarities in sequences became evident.

The similarities between the 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); it is possible to distinguish between the first process and the other two possibilities (see for example; Cantor & Jukes, 1966; Fitch, 1966a; Fitch, 1966b; Fitch, 1970). An example is the statistical approach of Fitch (1970), in which the frequency of occurrence of minimum mutation distances between various peptide segments of determined length, are compared with those expected from a purely random situation. However, even when a departure from randomness has been shown for the mitochondrial cytochromes *c* (Margoliash, Fitch & Dickerson, 1969), it is still necessary to distinguish between convergence and homology. It is, in fact, very difficult to rigorously rule out convergence. A semi-rigorous approach was presented by Fitch & Margoliash (1967b), for use with a set of proteins in which many invariant residues exist. They statistically determined the expected number of invariant residues, which will be found irrespective of the eventual number of sequences determined; if the number of identical

residues in two sequences exceeds this value then they conclude that the sequences are ancestrally related. Fitch & Margoliash (1967b) calculated the expected number of invariant residues for eukaryotic cytochromes to be 27-29; recently more sophisticated calculations (Fitch & Markowitz, 1970) have given a value of 32 residues. As the number of observed unvaried residues between the eukaryotic cytochromes is presently 33, it is concluded that they are all ancestrally related.

Dickerson (1971) has noted however, that the use of minimum mutational distances in all the statistical approaches has inbuilt errors arising from the degeneracy of the genetic code, and from the nature of the code in minimizing the effects of random mutations on the chemical properties of the polypeptide chain subsequently produced.

Another source of difficulty in comparing sequences arises from the cases where additions or deletions of residues may have occurred. Certain methods to show departure from randomness between sequences show the probability of either homology or convergence without the need to consider such events (Fitch, 1969, 1970; Gibbs & McIntyre, 1970). For the mitochondrial cytochromes c the numbers plus 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 essential cysteine residues common to all the cytochromes, are aligned. Particular difficulties arise when attempts are made to optimise the similarity between more distantly related cytochromes, especially those not belonging to the same group. Relationships between the sequences of Pseudomonas fluorescens c_{551} (Ambler, 1963)

and Rhodospirillum rubrum c_2 (Dus et al., 1968), and the mitochondrial cytochromes c, have been proposed (Cantor & Jukes, 1966; Dus et al., 1968; Needleman & Blair, 1969; Dickerson, 1971), but the various proposals of alignments disagree with one another.

The only certain way to obtain such comparisons is from x-ray crystallographic analysis of the 3-dimensional structure of the molecules. Such an experimental approach offers a solution to the rigorous differentiation between convergence and homology; the minimal sequence constraints required by function would not require the same tertiary structure over the entire molecule. Thus, if the structure determinations show that the 3-dimensional structure of the two molecules is essentially the same over the entire molecule, this is much more likely to have arisen by homology. For example, the closely related trypsin, chymotrypsin and elastase, all have serine residues at the catalytic site and possess similar tertiary structures indicating homology (see Shotton & Hartley, 1970). The bacterial subtilisins, also with an active centre serine residue, possess little structural relation to the other group (Smith et al., 1966; Wright, Alden & Kraut, 1969), indicating convergence. The tertiary structure of cytochrome c has only been determined for the horse and bonito proteins (Dickerson et al., 1971), although studies are currently in progress on Rhodospirillum rubrum c_2 (Kraut, Singh & Alden, 1968) and Pseudomonas fluorescens c_{551} (Dickerson, 1971).

Dickerson (1971) has used the x-ray data from the horse cytochrome to show the homology and to deduce the location of the deletions between Pseudomonas fluorescens c_{551} and horse cytochrome c. The sequence comparisons which he makes are

fairly clear despite only 25% similarity between the sequences. The plant cytochrome c sequences clearly fit with the structural constraints determined for the horse and bonito proteins; many major points of similarity exist between the sequences which are basic structural or functional requirements in the horse cytochrome. It is concluded therefore, that the tertiary structures of the animal and plant cytochromes are essentially identical, indicating a common ancestry for both.

There are many features in the plant sequences which lead to a high degree of confidence in this conclusion. However, only a complete x-ray investigation can verify identities in tertiary structure.

In horse cytochrome the chain folding is such that the molecule has an hydrophobic interior and polar exterior; this gives a good example of the "oil drop" model for a protein. The molecule shows four main features which are probably the important interaction sites for the molecule. These are the hydrophobic crevice, in which the heme group is located, two hydrophobic 'channels' and specific groupings of charged residues on the surface (Dickerson et al., 1971). The horse, bonito and plant proteins show large stretches of identical residues and where differences are found, these are generally of a conservative nature being chemically very similar. The essential residues required by the heme group are, the two cysteine residues 22 and 25*, which are covalently bound to the porphyrin

*The numbering used here is based on the plant, and not the animal sequences. The sequences have been aligned relative to the two cysteine residues.

ring, the histidine-26 residue, and the methionine-88 residue, which provide the fifth and sixth ligands for the iron atom. In horse cytochrome there are 16 hydrophobic residues which point towards the heme group. The identical residues are generally found in the plant sequences; the exceptions are still changes to other hydrophobic residues. These residues account for the majority of the nine aromatic residues present; those not directed inwards are postulated to be involved in electron transport through overlapping π-orbitals. In plants there are an extra two aromatic residues in addition to the nine others, as in horse cytochrome. In horse cytochrome all the polar residues are located on the surface of the molecule. The charge distribution is not random but distinct areas of positive or negative charge are found. Similar patches in the same areas can be deduced to be present in plant cytochromes, but not necessarily formed from the same residues in the sequences. The hydrogen bonding areas were often ill-defined in the analysis of the horse cytochrome molecule, but potentially hydrogen bonding residues are generally found in similar positions in the plant cytochromes. Nearly all the glycine residues found are invariant to all cytochromes and for most of them steric reasons for their presence are obvious from the horse structure. The addition of the sequences of the 16 plant cytochromes has led to only one change in the number of unvaried residues found between all mitochondrial cytochromes c. This is a result of a change, which is found only in the Ginkgo sequence, of alanine to glycine in position 59. The structural requirements necessitating invariant residues would suggest that statistical estimates that 32 would be a minimum

number (Fitch & Markowitz, 1970), would appear to be a reasonable result.

Despite all these striking similarities, several notable differences between plant and animal cytochromes exist.

The stability of plant cytochromes during preparation is less than for animal cytochromes; they are more readily denatured by ethanol and digested by proteolytic enzymes and the ratio of the absorbances at 416(Red)/550(Red) are higher. The latter suggests a possibly more open tertiary structure (Stellwagen, 1968), which may account for the other observations.

All the plant cytochrome sequences contain two residues of the unusual amino acid ϵ-N-trimethyllysine in position 80 and 94. Only one residue of this amino acid is found in the fungal cytochromes c in a position equivalent to residue 80 in the plant sequences (Delange et al., 1969; Delange, Glazer & Smith, 1970), and none are found in any animal sequences (Delange et al., 1970). Either one or two residues are also present in the c-type cytochromes of certain protozoans (Pettigrew & Meyer, 1971; Hill, Gutteridge & Mathewson, 1971). Trimethyllysine has been reported in two other proteins; myosin (Kuehl & Adelstein, 1969; Hardy, Harris, Perry & Stone, 1970a, b) and histone (Hempel, Lange & Birkofer, 1968).

Since all the lysine residues are located on the surface of the cytochrome molecule (Dickerson et al., 1971), it is difficult to understand why only certain lysine residues are modified subsequent to the synthesis of the protein (Scott & Mitchell, 1969). This may result from either the specificity of the methylating enzyme to the adjacent sequence or from the specific conformation of the chain or a combination of both.

The enzyme may require a Lys-Lys sequence since in the plants and fungi a Me_3Lys-Lys sequence results from methylation, but in the fungi the sequence Lys-Lys occurs in the position where the second methylated residue occurs in plants, position 94, but in this case no methylation is found (Delange et al., 1970). In the animal cytochromes a similar Lys-Lys sequence is invariably found. It is possible that the conformation of the polypeptide chain in the tertiary structure prevents access by the enzyme. Clearly such a means of protection could be important, since modification of certain lysine residues, particularly lysine-21, could be deleterious (Wada & Okunuki, 1969). It is possible that the presence of proline residues, 91 and 96 common to all the higher plant sequences, are important from this respect; these residues are not necessarily found in the fungal sequences. However, this does not explain why the methylation of residue 80, which is located in an 11 residue sequence, which is invariant to all mitochondrial cytochromes, is not common to all.

The plant sequences contain 15 residues which are only found in their position in the plant sequences. No correlation between these could be observed from studies of the 3-dimensional model of horse cytochrome c.

Compared to other cytochromes the plant sequences are longer; thus, at first sight it would appear that the cytochrome gene has shortened during evolution (Smith, 1968). However, there is now strong evidence that in animals and fungi cytochrome c is synthesised on 80S cytoplasmic ribosomes (Boulter, 1970). It has been demonstrated that the mechanism for initiation of protein synthesis on 80S ribosomes, involves

$tRNA_{F^*}^{Met}$ (Smith & Marcker, 1970). This would imply that the N-terminal amino acid of cytochrome c, when it is synthesised is methionine and that subsequently, a hydrolytic enzyme must have cleaved off residues from the N-terminus. The specificity of the postulated enzyme may thus be responsible for variations in sequence length at the N-terminal. It is possible that the presence of an acetyl group in most cytochromes is coupled with this process. Certainly it would appear that its presence is not for removing the otherwise free amino group from the heme environment (Dickerson et al., 1971), as was previously postulated (Margoliash & Smith, 1965). The variations of sequence length at the C-terminus where wheat has one and Ginkgo two extra residues compared to the other plant sequences, probably reflects the possibility of mutations in the chain termination mechanism.

The sequence positions where variability is found between the plant cytochromes, are all located on the surface of the molecule and are in different positions to those for animal cytochromes. Thus, the surface distribution of the variable residues from the two groups is completely different when examined on the 3-dimensional model, but both sets show variability centred on particular zones of the molecule's surface.

Despite the large amount of evidence, it is still necessary to assume homology between the plant 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 c gene. As the cytochrome c is an expression

of a small part of the genetic material of the species from which it was obtained, it therefore, follows automatically that the relations between the sequences relate directly to the relations between the species themselves. It is thus, possible to derive the relations between the various kingdoms and the individual members of each kingdom from the relationships which exist between their cytochromes.

When establishing gene or species phylogenies, it is important that distinction is made between 'parologous' and 'orthologous' genes (Fitch & Margoliash, 1971). Genes originally identical may, after duplication, diverge subsequently to the extent that they code for proteins with different functions. Such genes are homologous, but have been termed 'parologous' (Fitch & Margoliash, 1971), to distinguish them from the other class of homologous genes, 'orthologous' genes, which are those which have remained identical in function throughout. The mitochondrial cytochromes c apparently represent an 'orthologous' gene set. When detailed comparisons are to be made between sequences, a problem which may exist in the data is the presence of acidic residues in the sequence, which may be present in the form of amide residues but for which this information was not determined. In this investigation amide content was known for all the cytochrome sequences and when necessary the amide residues were assigned by homology, so as to give a minimum number of differences between the cytochromes (see Appendix 3).

One of the first and most interesting observations on the animal cytochrome c sequences, was that the number of interclass amino acid variations was always greater than intra-

class differences, and that the size of the interclass difference increased as the species considered became more distantly related (Margoliash, 1963; 1964; Margoliash, Needleman & Stewart, 1963; Smith & Margoliash, 1964; Margoliash & Smith, 1965). Further, the constancy which exists in groups of interclass comparisons is striking when one considers the wide range of variations on intraclass comparisons. These relationships have been interpreted to imply that elapsed time, as such, is an important parameter in determining the number of effective mutations accumulated by the cytochrome structural gene along any line of evolutionary descent (Zuckerkandl & Pauling, 1962; Margoliash, 1963). The factors determining this empirically derived linearity concept cannot be certain, but it is suggested (Margoliash & Smith, 1965; Margoliash & Schejter, 1966) that over long periods of evolutionary history - perhaps a minimum of 200 years for cytochromes - other factors relating to the rate of fixation of amino acid differences have either cancelled or averaged themselves out, leaving elapsedtime as the variable more clearly related to primary structure differences. For such an empirically derived relation however, it is unnecessary to assume any of the mechanisms through which the differences may have been achieved (Margoliash & Schejter, 1966). Accepting the relation enables a direct correlation to be made between time and the rate of change of cytochrome c (Margoliash, 1963).

In this investigation the single amino acid difference and minimum mutation distance 'unit evolutionary periods' (Margoliash & Smith, 1965; Margoliash & Fitch, 1968) have been calculated using all the cytochrome data, and these values used to examine the divergence times of the major taxonomic

groupings (see Table 19). The time periods obtained for the divergences between the major animal groups are fairly consistent and independent of whether amino acid differences or minimum mutation distances were being considered (see Table 19). This is not the case when the time for the divergence between the three major kingdoms is considered. The time for the divergence between the plant and animal kingdoms is noticeably less when minimum mutation distance is considered. This may reflect the partially arbitrary rules used in calculating this measure (Fitch & Margoliash, 1967a), or the inherent problems associated with this parameter (Dickerson, 1971). The values obtained for the times of divergence of the three major kingdoms are such that no specific order in their descent can be determined. It is probable that all three diverged at approximately the same time. The linear extrapolation from the unit evolutionary period to give the date for this point of divergence will always represent an underestimate of the true value. 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 & Margoliash (1968) estimate that of the mutations which have occurred in animal cytochromes, 20% are parallel and 1% are back mutations. A statistical correction (Margoliash & Smith, 1965) based on the number of actual events expected to produce the observed number of changes assuming a definite number of variable positions has been used. The main failure of this approach is that it assumes that all the variable sites are equally amenable to effective variations; this is certainly not the case for

cytochrome c, in which certain positions are considerably more variable than expected if a random situation existed (Fitch & Margoliash, 1967b). Despite the drawbacks and assumptions, these corrections probably give a more correct idea on times of divergence and lead to a value of about 1500 million years for the time since the divergence of these three eukaryotic kingdoms. Since the sets of sequence comparisons which are averaged, do not consist of an independent set of observations, it is not practicable to apply detailed statistical estimates as to the degree of error associated with this value. McLaughlin & Dayhoff (1970) have calculated from cytochrome and tRNA sequence data, that the divergence of eukaryotic and prokaryotic lines was 2.6 times more remote than the divergence of the eukaryotic kingdoms. A similar value was obtained by Jukes (1969), by considering only tRNA data. Further, McLaughlin & Dayhoff (1970) calculated that the differentiation of the various tRNA types was 1.2 times more remote than the divergence of the eukaryotic and prokaryotic lines, and showed that since this point the rate of change in the differing tRNA species was independent of the line of descent to which they belonged. These ratios would suggest times of 3,900 and 4,700 million years since these major divergences. These values are noticeably in excess of previously predicted values (see Klein & Cronquist, 1967; Dayhoff, 1969).

The unit evolutionary periods derived for cytochrome c show that it is accepting mutations more slowly than almost all other proteins so far examined (Dayhoff, 1969). The large value of the unit evolutionary period suggests that severe structural constraints exist for the cytochrome c molecule.

The reasons for these are evident from the tertiary structure (Dickerson et al., 1971) and the minimum of three interacting functions, with oxidase, reductase and membrane.

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 difficulties in constructing such trees is the enormous number of alternative trees which are possible and ideally have to be evaluated. For sixteen species this number is about 10^{14}, so that even with a computer it is impossible to assess all the possible trees. It is therefore, necessary to use methods for constructing phylogenetic trees which drastically reduce the number of trees to be evaluated, while still finding the optimum solution.

When constructing a tree from all the cytochrome c data, a simple reduction can be made if the limited biological information is assumed to split the data into three groups by kingdoms. The tree for each kingdom may then be separately constructed and then all three combined to give the final tree. Even so, the number of alternative trees which can be constructed from the sequences in a single kingdom, is still too large and further reductions in the number of trees evaluated are necessary.

Two types of approach to this problem have been described; these are the numerical matrix method (see for example; Fitch & Margoliash, 1967a; Gibbs & McIntyre, 1970), and the 'ancestral sequence' method (Dayhoff & Eck, 1966). Both types of method have been used here to construct phylogenetic trees from the

plant sequence data.

The numerical matrix method can be divided into two parts; the choice and calculation of the similarity or dissimilarity measure matrix, and the choice of grouping strategy used.

The simplest measure relating two sequences is the number of amino acid differences between them. The other measure used in this investigation has been 'minimal mutation distance' (Fitch & Margoliash, 1967a). In scoring each amino acid difference as a single mutation event, one assumes that any number from 1 to 3 base changes could be accomplished by a single mutation event. Fitch (1966a) introduced the use of the minimum number of DNA mutations required to interconvert homologous sequences as a measure of similarity. Since the genetic code is degenerate, Fitch (1966a) found the minimum number by choosing for each amino acid the optimum pair of triplets. This assumption means that possible intermediate stages may still be missed.

However, with both these measures the distances along the branches of the resulting phylogenetic trees have a numerical meaning clearly related to the sequences. Other proposed similarity measures, such as 'diagonal index', 'runs index' (Gibbs & McIntyre, 1970), or 'cross association' (Sackin, 1971), lack such a clear meaning and hence the tree produced does not have dimensions identifiable with the sequences, but only has shape. Whereas, either amino acid difference or minimum mutation distance measures, although containing hidden assumptions, give phylogenetic trees for animals which are in very close agreement with trees constructed from the biological and fossil data (Fitch & Margoliash, 1967), the other measures do not

necessarily give such compatible results (see Gibbs & McIntyre, 1971). The numerical meaning of these other measures is not clear, particularly when compared to time for example, so it cannot be certain what the resulting trees really represent.

It would be possible to derive other measures relating sequences. For example, those which could be related to amino acid differences or minimum mutation distance by application of additional weighting factors. However, any such additional measures would lack a clear numerical meaning on the resulting tree. There can be no justification in using any essentially arbitrating weighting system, since by suitable adjustments any desired tree could be obtained. If the aim is to obtain a best fit between a sequence data set and an existing tree, this approach could be useful, but when sequence data is being used to look for new phylogenetic and taxonomic insights, the direct measures used in this investigation should be employed.

Having constructed a similarity or dissimilarity matrix relating the species, it is possible to use this to construct a phylogenetic tree. A wide number of approaches are available; Lance & Williams (1966a; 1967a) have demonstrated however, that most of the normally used agglomerative sorting strategies (see Lance & Williams, 1967a) could be generalized as a single strategy defined by variable coefficients.

Agglomerative strategies, which are always polythetic, may be divided into 'clustering' strategies, in which a property of a group of elements is optimised, and 'hierarchical' strategies, in which the route by which groups are obtained is optimised. For the purposes of constructing a phylogenetic tree, a 'hierarchical' strategy is used. The phylogenetic

tree which is produced is a graphical representation of the order in which sequences and groups of sequences are fused. Once a sequence has been incorporated in the tree only the group to which it belongs is considered at further fusion stages; this drastically reduces the amount of calculation needed. A 'hierachical' strategy may be 'combinatorial', in which new matrix elements can be directly calculated from the old at successive fusions or 'non-combinatorial' when additional information is required after each fusion; the former has distinct computational advantages. It may be 'compatible' or 'incompatible', depending on whether or not the measures calculated at latter fusion stages are of the same kind as those initially supplied. Finally, it may have variable 'space' properties. When groups are formed it does not follow that the new inter-group measures define a 'space' with the original properties; if they do the strategy is 'space-conserving'. With 'space-distorted' strategies groups may, when formed, move nearer some or all the remaining elements so that the chance that an individual will add to a pre-existing group is increased giving the tendency to 'chain' (Williams, Lambert & Lance, 1966); this is a 'space-contracting' system. Alternatively, 'space-dilating' systems may exist in which groups recede from other elements on formation and tend to give 'non-conformist' groupings (Watson, Williams & Lance, 1966).

The fusion strategies used in this investigation were those of Fitch & Margoliash (1967a), and the 'flexible' strategy of Lance & Williams (1967a).

The flexible strategy is derived from the generalized formula of Lance & Williams (1966a). It is combinatorial by

definition and can be considered compatible with the two measures used. Provided that certain constraints are observed, it will also be 'monotonic'. The flexible nature of the strategy lies in its complete range of space-distorting properties, which may be used to advantage to search for optimum groupings.

The method of Fitch & Margoliash (1967a) was used without the statistical evaluation. By a comparison with the generalized formula of Lance & Williams (1966a), the method is expected to show space-contracting properties which may lead to 'chaining' occurring, and it will not necessarily be monotonic. Because of these suspected limitations to this method, the flexible strategy of Lance & Williams (1967a) was used as the major numerical approach. Of the other strategies available from the same computer programme, for example, nearest-neighbour, furthest-neighbour (see Lance & Williams, 1967a), it was felt that the variability of this approach would allow better correlations to be made with the trees derived by the 'ancestral sequence' method (Dayhoff & Eck, 1966).

How meaningful and useful are the trees derived by numerical methods? The use of the method assumes that evolution has occurred with a minimum number of changes. This assumption, common to all tree constructions, that parallel and back-mutations have not or have only rarely occurred, is clearly wrong (for example, see Fitch & Margoliash, 1969). The degree to which such events may effect the final tree will depend on the methods being used in constructing the tree. In the numerical methods of comparison used in this investigation, no account can be taken in the initial dissimilarity matrix of any parallel or back mutations which have occurred. This

may mean that the final tree is distorted. Secondly, as differences between sequences from present-day organisms are used, the resulting groupings will only be consonant with evolutionary groups if the rates of change are reasonably constant along all lines of descent (Jardine, van Rijsbergen & Jardine, 1969). Over periods in excess of 200 million years this is probably correct for mitochondrial cytochrome c. This means that numerical methods are of value when relating more distant species, but care has to be taken when using the results obtained on species that have diverged more recently (Jardine et al., 1969), such as with the higher plants.

Fitch & Margoliash (1967a) suggested that a 'percentage standard deviation' can be calculated, between the input sequence differences and sequence differences reconstructed from a final tree, as an estimate of the relative value of different trees. This was not used with the numerical method trees calculated using their method, because the 'chaining' observed clearly indicated inadequacies in the fusion strategy.

The alternative approach to the numerical methods for constructing phylogenetic trees is the 'ancestral sequence method' of Dayhoff & Eck (1966). In the 'ancestral sequence method', the ancestral sequences at all nodes of a topology are inferred and the topology then evaluated by counting the minimum number of amino acid changes which must have occurred. Initially, three species are arbitrarily chosen and the best topology formed when a fourth species is added determined and used for subsequent construction. Additional species are added in turn to the previous best topologies until a complete tree is obtained. Because an error may have occurred in any

early fusion step which altered the location of subsequent sequences, the final topology is checked by systematically relocating branches, to see if a better tree exists, i.e. one in which fewer mutations will have occurred. This method gives a closed topology for the final tree and gives no time scale or earliest point of time information. If this is required, it must be obtained either as a result of numerical calculations or from biological evidence. Since not all the possible trees are evaluated the minimum value obtained may be a 'local' minimum as opposed to the absolute minimum and therefore, it cannot be absolutely certain that the best tree has been found. In assessing the trees integral values are obtained as opposed to continuous values when numerical trees are assessed (Fitch & Margoliash, 1967a), and it is therefore, possible that no unique minimum may exist and that several alternative best trees exist. When this situation occurs, it cannot be certain that all of the equally probable trees have been found.

A minimum mutation hypothesis is again used as a basis for this method; this cannot rigorously be shown to be correct (Jardine et al., 1969). It is quite possible that additional mutations above the minimum have occurred in the natural course of events and that the best tree calculated does not represent the correct tree. This can be assessed to a limited extent when irrefutable biological evidence can be correlated with the tree.

The numerical methods are not considered to represent the best method compared with the 'ancestral sequence method' for constructing phylogenetic trees. This is because the 'ancestral sequence method' is not directly effected by varying rates of

evolution. If the rate of change along a certain line of descent is, for example, faster than average, this will not mean that the resemblance of a species to its common ancestor is lost and that it will fit elsewhere on the topology but only that the branch length will be larger. The numerical trees, on the other hand, may be considerably distorted by differences in rates of evolution. Also, because the ancestral sequences are specifically determined for each topology, parallel and back mutations will generally be detected and taken into account in the evaluation of the topology, whereas they are not evaluated in the numerical methods. However, it has been shown that, with the 'ancestral sequence method', a single change in a sequence may be sufficient in certain cases to cause a change in topology. This method, therefore, may be susceptible to sequencing errors or to misinterpretations of heterogeneity should either occur, whereas the numerical methods may well be affected less by such errors.

Generally, the results of numerical methods may be amenable to several interpretations, particularly when the results are of borderline statistical significance. However, this type of output may provide useful supporting evidence for other evolutionary hypotheses (Sackin, 1971).

The results from the numerical method calculations confirm that the method for Fitch & Margoliash (1967a) has a tendency to 'chain' (see Figs. 20 and 21). When the 'chaining' did occur, the order in which the species added was not random but related groups were added in sequence, e.g. niger and sunflower, arum and maize. The major difference between the results from the two dissimilarity measures used is the

location of castor which the minimum mutation distance gives as being of much earlier descent. Compared with the arum and maize positions, the location of wheat is anomalous, since the biological evidence strongly suggests that the monocotyledonous plants represent a monophyletic group (Cronquist, 1968; Takhtajan, 1969). Accepting that the position of wheat is in error, the point of earliest time divides Ginkgo, the single gymnosperm from the remaining plants which are all angiosperms. This is in full agreement with the fossil record (Walton, 1953).

With the flexible strategy it was found that over the range of space-distortion examined, at any given level, little difference existed in the topologies obtained from the two different measures considered; in no case however, were they the same (see Figs. 24 and 25). The species which led to the differences were tomato and sesame. When the minimum mutation distance measure was considered, tomato was strongly associated with the arum and maize group, but when amino acid differences were considered it was associated with castor next to the cotton and abutilon group. The differences for sesame only occurred in 'space-dilating' (β negative) topologies. When minimum mutation distance was considered sesame could be associated with castor; no such association was evident when amino acid differences were considered, when sesame was always more remote than the niger and sunflower group. Very few major changes exist over the range of 'space-distortion' considered for either of the series of topologies obtained from each measure considered. Most of the changes observed involve the last fusion steps. All agglomerative hierarchical strategies suffer from what are called 'migration problems'

(Lance & Williams, 1967b); fusions rightly made in the early stages of the process may later prove unprofitable. Fusions once made are irrevocable, so that eccentricities and errors in the data may direct the subsequent analysis along an unprofitable path (Lance & Williams, 1965). Spurious divisions are thus likely to be encountered at the later stages of the analysis (Lance & Williams, 1965; 1966b), and so the results obtained from the last fusions should be treated with caution or ignored. Chaining occurs in the most 'space-contracting' topologies, particularly when the amino acid differences measure is used, but when this chaining does occur, the order in which fusion occurs still places similar species in adjacent positions, e.g. maize next to arum, sunflower next to niger. The most noticeable change occurs when minimum mutation distance is considered and the topologies are becoming increasingly 'space-contracting'. Sesame loses its direct association with the group containing abutilon, cotton, castor, arum, maize and tomato. The arum, maize and tomato group then associates with the group containing mungbean, pumpkin, cauliflower and rape, instead of with cotton, abutilon and castor. This latter group then splits leaving castor in the same relative position, but places abutilon and cotton in closer association with the mungbean, pumpkin, cauliflower and rape group than that of the arum, maize and tomato group, which had previously moved. As these changes are occurring when the system is space-contracting, they may need treating with caution when the results are examined from a phylogenetic point of view.

The position of wheat on all topologies is separate from the other two monocotyledonous plants, arum and maize. Since the

biological evidence strongly suggests that monocotyledonous plants all belong to a single monophyletic group (Cronquist, 1968; Takhtajan, 1969) and because wheat is on occasions more remote on the topologies even than Ginkgo, a gymnosperm, it is believed that it is its position that is anomalous and not those of arum and maize. This means that the position of wheat should be ignored when the trees are examined with relation to plant phylogeny. The anomalous positioning of wheat may have affected the ordering of the last fusions in certain cases and may be responsible for Ginkgo not being the single most primitive line of descent as is expected from fossil and biological evidence (Seward, 1937; Walton, 1953). Otherwise, the sometimes anomalous position of Ginkgo may result from either the nature of the fusion strategy, particularly in the space-dilating fusions which tend to produce non-conformist groups (Watson, Williams & Lance, 1966; Lance & Williams, 1967a), or from the tendency of later fusions to have errors.

All the trees produced using both measures show that mung-bean, pumpkin, cauliflower and rape represent a very closely related group. They also show that certain pairs of species, niger and sunflower, arum and maize, and abutilon and cotton, are all closely linked into groups. All four of these groups would be expected from a consideration of biological evidence (Cronquist, 1968; Takhtajan, 1969). The trees indicate that the monocotyledonous plants were descended from a dicotyledonous line of descent and that the divergence of the line of descent leading to buckwheat was phylogenetically primitive. Lance & Williams (1967a) suggest that a value of -0.25 for the variable, β, used in the flexible method, would represent the best system

(see Figs. 22 and 23). The results here however, would indicate that better trees, i.e. those with fewer errors in the later stages, are obtained when a value for ß of nearer 0 to 0.10 is used.

Because it was not possible to construct a unique topology relating all the plant sequences using the ancestral sequence method, a strategy involving the construction of a 'basic tree' was attempted. The aim of the 'basic tree' was to have the largest possible unique tree using any combination of the sequences available which could then be considered as a fixed core to which the remaining species could be related individually.

The largest 'basic tree' (see Results section) which could be found contained ten species (see Fig. 16). When compared with the best trees obtained when all sixteen species were considered (see Fig. 30), the basic tree is reasonably compatible. The exceptions which exist are the alteration of the maize position and the placing of cotton and abutilon on a single line of descent. Both of these alternatives require an extra mutation than is required by the basic tree (see Fig. 27). Using the 10 species basic tree, the addition singly of each remaining species did not lead to a better 11 species basic tree. Only wheat, when added, gave a unique topology as the new best tree. This tree was disregarded as wheat was not adjacent to maize, the other monocotyledonous plant. With the other species many alternative trees were found; two with castor, arum and maize, three with niger and nine with sesame (see Figs. 28 and 29). This high variability found with sesame suggests that possibly large errors may exist in the location of this species on the trees constructed when all species are considered.

Although the species input order was random when constructing the basic tree, it cannot be definitely certain that this solution represents an absolute minimum for the species used. Also, it cannot be certain that because of the limited number of permutations of the species which can be examined in the construction of the basic tree, that an alternative basic tree, which is either larger or contains a different ten species, does not exist.

Various alternative trees were obtained when all sixteen species were considered. Without any additional evidence, such as biological evidence, it is not possible to say which of the possibilities represent the true evolution of cytochrome c. The variations between the trees are often localised among small groups of species, so that few major topological differences exist.

From all the trees it is again seen that wheat occupies an anomalous position away from the other monocotyledonous plants, as was observed with the numerical trees and its position should be disregarded. Certain noticeably constant features emerge. Cauliflower, rape, pumpkin and mungbean, always form a close group. Similarly, castor, cotton and abutilon, are always in a single group. Niger and sunflower are always closely linked and the two monocotyledonous plants, arum and maize, are always associated with this group. The topology relating Ginkgo and buckwheat with tomato is normally the same, although the tomato may be one node removed. In many of the alternative trees, tomato, sesame, abutilon, cotton and castor are closely linked.

No point of earliest time is obtained for a tree calculated

by the ancestral sequence method. This may be achieved either by comparison with numerical methods or by additional biological information. Both these methods indicate that the point of earliest time lies along the branch leading to Ginkgo. This conclusion was confirmed when a tree relating all the mitochondrial cytochrome c sequences and the cytochrome c_2 sequence of Rhodospirillum rubrum was constructed (see Fig. 31). This tree showed the early divergences of Enteromorpha and Ginkgo in the plant kingdom. It also confirmed the conclusions drawn from the numerical comparisons, that the origin of the three separate kingdoms probably occurred at approximately the same time, and that no definite order could be assigned; only two mutations separate the three alternative topologies (see Fig. 32), and in this case, this is not considered significant. This conclusion on the origins of the kingdoms, is in accord with the biological evidence (Whittaker, 1969).

From these trees relating all the cytochrome sequences available, the ancestral sequences at the origin of the plant kingdom, at the point of divergence of Enteromorpha and at the point of divergence of Ginkgo were determined (see Fig. 33). It was impossible to assign a residue to every position in each case and errors may exist in the earliest ancestral sequence, as a result of using cytochrome c_2 in the comparisons. However, if the ancestry of the c-type cytochrome can be extended back to include this and other bacterial cytochromes, further and better ancestral sequences may be derived. The potential which now exists for synthesising cytochrome c (Sano & Tanaka, 1964) would enable a direct examination of the properties of these extinct proteins. From an examination of these ancestral

sequences an increasing degree of basicity is observed along the animal line of descent (Smith, 1968), but this is not so evident along the plant line.

None of the trees derived by the numerical matrix methods has the same topology as any of the alternative best trees found by the ancestral sequence method. This may be because not all the alternative best trees have been found, and that the numerical topologies are equivalent to an undetected best solution. However, when the topologies equivalent to the two approximately 'space-conserving' solutions, calculated by the flexible method (Figs. 22 and 23), were evaluated by the ancestral sequence method, neither represented a minimum value solution.

Despite the fact that no single phylogeny can be derived using both methods, several important features appear common to both.

Species which are morphologically very similar, belonging to a genus or genera of the same family, have sequences which are more similar to each other than to sequences belonging to other species. Thus, cauliflower and rape sequences (same genus) are identical, niger is most similar to sunflower (Compositae), and cotton is most similar to abutilon (Malvaceae). Families closely related on general biological grounds are often grouped together; thus, cauliflower and rape are closely related to pumpkin, and cotton and abutilon are often related to castor. Buckwheat, a member of the Polygonaceae, is always found on a phylogenetically primitive line of descent.

The positions of arum, maize and wheat, the only monocotyledonous plants so far examined, show certain anomalies.

Although the biological evidence strongly supports their having a monophyletic origin as a group, on all the trees wheat is separated from the others and on the ancestral sequence trees arum and maize are not on a single line of descent. Since the linking of wheat is totally different between the two types of method, it is probable that its position is the major error and that arum and maize are grouped fairly close to their true position. This means that the monocotyledonous plants were derived from a dicotyledonous plant line of descent, probably at a time fairly close to the divergence of the Compositae line of descent.

Different possible causes for this anomaly exist. The grasses, and wheat in particular, have been subjected to intensive plant breeding, in which the selection pressures on the species have been totally changed. Furthermore, wheat is inbred, whereas natural populations normally have a considerable measure of outbreeding. It is possible that this may have increased the rate of fixation of mutations in the wheat cytochrome. Certainly, from the numerical trees, it would appear that the rate of acceptance of mutations has been much faster along the wheat line of descent than along any other. Another possible cause is the occurrence of polyploidy. The large potential which a polyploid will have over the diploid for accepting mutations by a redundant gene copy could give rise to a more rapid rate of acceptance. Thus, for wheat, a hexaploid, this may have been an important process. However, only one cytochrome c is obtained from wheat, suggesting that the additional copies of the gene have been lost or are all still identical.

As mentioned previously, the ancestral sequence method is to be preferred, and the phylogenies obtained using this method have been compared with the 'phylogenetic classifications' obtained from classical comparative methods. Turril (1942) has pointed out however, that none of these present-day so-called 'phylogenetic classifications' is truely phylogenetic, but are only presumed so as they have been built using phylogenetic principles. An important distinction has been made between trees which show geneological relationships (cladograms) and those which, based upon phenotypic evidence, show the relative similarities among the organisms being considered (phenograms) (Sokal & Sneath, 1963; Fitch & Margoliash, 1969). The trees constructed here are not strictly cladograms (Sokal & Sneath, 1963), but as they represent considerably more information than a phenogram classification, it is suggested (Fitch & Margoliash, 1969) that referring to them as phylogenetic trees is acceptable. If a less rigorous definition of a cladogram is accepted, then phylogenetic trees may be referred to as cladograms (Sneath, 1969).

The idea common to most plant classifications, is that species with simple morphological characters are primitive and have given rise to more complex characters which have subsequently given rise to even more complex characters. This meant that families with simple characters were regarded as belonging to phylogenetically primitive lines of descent, and those with complex characters as belonging to phylogenetically advanced lines of descent (see Lawrence, 1951; Cronquist, 1968; Takhtajan, 1969). Unfortunately, there is no fossil evidence to support such a contention; the earliest certain angiosperm remains

date from Middle Cretaceous deposits, when forty-nine modern families are represented (Sporne, 1971). The cytochrome c tree however, indicates that the morphologically primitive group of mungbean, pumpkin, cauliflower and rape group are phylogenetically advanced. Since there is no direct correlation between morphological and biochemical evolution, it is not possible to distinguish whether this represents a single primitive line of descent from which several different lines of descent have evolved, each of which has independently acquired advanced morphological characters or whether a gradual simplification of morphological characteristics has occurred during evolution. The evidence at present suggests that the first flowering plant was simple in nature and probably similar to members of the Magnoliaceae or their allies (Sporne, 1971), which would indicate that the former course of evolution occurred. Should this prove to be the case, a major revision of the existing 'phylogenetic classifications' will be required in order to give a truly phylogenetic tree relating flowering plants.

Before such a revision could occur, it is necessary to consider the degree of error which may exist in the protein phylogeny. Currently, there is a lot of variation in the trees obtained, both by different and the same methods. Since only a fraction of all the possible trees are evaluated, the methods used to reduce the numbers calculated may miss the optimum solution. However, since the two types of method both give similar results, this possibility is minimal. The numbers of best 'ancestral sequence method' trees should decrease with the addition of more sequences and could well eventually give a unique topology. The addition of more

sequences helps since on the present tree the distances between the nodes are often small compared to the node-species distances, and this leads to instability in the topology. At present certain errors can be identified, e.g. the position of wheat. It cannot be certain that others do not exist which cannot be so readily identified. The addition of extra sequences should however, minimize this possibility.

If a unique topology can be achieved or else the situation when very few closely related alternatives exist, it is fairly probable that these will represent a true phylogeny. This assumption is however, based primarily on the results of protein phylogenies derived using animal proteins. In these cases, the protein phylogeny agrees, with a few minor exceptions, with the fossil record (Fitch & Margoliash, 1967a; Dayhoff, 1969). This meant that no new taxonomic insights were obtained, but this is a strength, not weakness, in the method. There are however, differences in the mechanism of plant and animal evolution, which may effect the validity of the protein phylogeny. Plants differ from animals in that they commonly show polyploidy, which is often stabilized by the widespread occurrence of vegetative reproduction. Hermaphroditism can simplify problems associated with sex determination (Stebbins, 1950). A further difference is that with the vertebrates there is a good correlation between major taxonomic groups and ecological niches where the evolutionary pressure has been towards adaptation to a niche (Cronquist, 1968). Such a situation is not as apparent within the angiosperms, where a single family may occupy many different ecological niches. Woolhouse (1970) considers it possible that similarities in sequences could be determined by

the demands inherent in these life forms and environments, rather than by the families to which they belong.

To be important, any of these differences must directly effect the selection pressure that exists on the cytochrome c molecule. However, the extent and nature of this selection pressure are difficult to define. Despite the wide range of primary structures determined, no cytochrome appears to be significantly better when chemical comparisons are made. For example, all show similar reactivities with mammalian cytochrome oxidase preparations (Jacobs & Sanadi, 1960; Morrison, Hollocher & Stotz, 1961; Davis, Smith & Wasserman, 1964), and removal of the C-terminal amino acid residues from the protein does not affect activity in this system (Titani, Ishikura & Minakami, 1959). The C-terminal region is a very variable region in the protein; however, other positions exist where even greater variability of fixed amino acid residues is found. This situation is usually considered to have arisen by a positive selection pressure for any character (amino acid substitution) which therefore, becomes fixed in a population. However, the apparent chemical and enzymatic similarities between different cytochromes are difficult to reconcile with the uniqueness of sequences observed, and Margoliash, Barlow & Byers (1970) have suggested that such a selection pressure may arise from additional functions of cytochrome c, other than that in the electron transport chain. Alternatively, all the fixed mutations may be a result of neutral mutations (Kimura, 1968; King & Jukes, 1969). Since cytochrome c has existed in its present role for such a considerable period of time, the probability of advantageous mutations occurring is now remote

while lethal mutations will not survive by definition. Possibly the strongest evidence in support of this theory of neutral mutations is the remarkable uniformity in the rate of change of each different protein in the course of evolution (Kimura, 1968; 1969; King & Jukes, 1969; Crow, 1969; Kimura & Ohta, 1971a). If neutral mutations are occurring the occurrence of polymorphism would be expected to be detected (Kimura, 1968; Kimura & Ohta, 1971a). The observed polymorphism in Drosophila and man have been cited as evidence for the theory (Kimura, 1968), although this has been criticised (Prakash, Lewontin & Hubby, 1969; Petras, Reimer, Biddle, Martin & Linton, 1969; Maynard Smith, 1970). Polymorphism was not observed in cytochrome c when several individual human and horse preparations were examined (Margoliash, 1969a). However, Margoliash et al. (1969) explained this by calculating that there was only a 1% chance that any given individual would carry an unusual cytochrome.

Mathematical models for fixation of neutral mutations have been derived and from these, approximate values for average nucleotide substitution rates have been calculated (Kimura, 1968; 1969; Kimura & Ohta, 1971a; 1971b). The high values obtained for these rates are regarded as further support for this mechanism (Kimura, 1968), and indicate the probable importance of random genetic drift due to finite population numbers (Wright, 1931) in forming the genetic structure of biological populations. Although the random nucleotide substitution rates in all proteins should be reasonably similar, a wide variation is found in the rate at which various proteins accept substitutions. This variation is greater than can be accounted for directly by differences in structural and

functional requirements. However, Fitch & Markowitz (1970) have demonstrated that at any one time the number of positions in a given protein from a single source, which are capable of accepting mutations, may be very small compared to the total number of residues which may vary between many sources over large periods of time; this group of positions they termed the 'covarions'. Mutations at any other position would be lethal. They showed that the fixation rate per 'covarion' was essentially constant for proteins with widely differing amino acid substitution rates, and concluded that this was strong evidence in favour of neutral mutations and that if quantified, this rate could lead to an excellent evolutionary time scale. For cytochrome c they calculated that at any one time the number of 'covarions' is between 4 and 10 (Fitch & Markowitz, 1970; Fitch, 1971a), and that a non-identity of these positions exists in the cytochromes of different species (Fitch, 1971b). Fitch (1971a) showed that the difference between the large 'local' invariance, about 90%, and the still substantial 'global' invariance, about 30%, resulted of a continual changing of the positions comprising the 'covarion' set with the occurrence of one or more changes within the set.

Consequently, although certain assumptions are made, it is felt that a correct phylogenetic tree relating plants may be obtained by studying protein phylogeny. However, anomalies do exist at the present time with such a tree constructed using cytochrome c data only. These anomalies may be removed by extending the tree, but use should also be made of correlations with phylogenies derived using other proteins, for example, ferredoxin or plastocyanin, before a definitive tree can be

constructed. Also, it is an important principle of taxonomy that all available information is utilised; therefore, when a phylogenetic tree is available from protein phylogeny, this result, based on the information of only a single gene, must be integrated with morphological and biochemical data, both products of many genes. When a final protein phylogeny has been obtained relating the plant and other cytochromes, this will enable the existing statistical calculations to be re-evaluated. Also, by determining additional parameters, such as leg lengths on the trees, it will be possible to examine the assumptions made about constant rates of change, numbers of parallel and back mutations. It will also be possible to establish a better time scale relating to events on the tree.

REFERENCES

ALSTON, R. E. and TURNER, B. L. (1963). Biochemical Systematics. Englewood Cliffs, N.J.: Prentice-Hall.

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

AMBLER, R. P. (1968). Biochem. J. 109, 47P.

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

BACHMANN, E., ALLMAN, D. W. and GREEN, D. E. (1966). Arch. Biochem. Biophys. 115, 153.

BENSON, A. M. and YASUNOBA, K. T. (1969). J. Biol. Chem. 244, 955.

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.

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

BOULTER, D., THURMAN, D. A. and TURNER, B. L. (1966). Taxon 15, 135.

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

BRYSON, V. and VOGEL, H. J. (1965). Evolving Genes and Proteins. New York: Academic Press.

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

CARBON, J., BERG, P. and YANOFSKY, C. (1966). Cold Spring Harbor Symp. 31, 487.

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

CRICK, F. H. C. (1958). In The Biological Replication of Macromolecules, Soc. Exp. Biol. Symp. 12, 138. Cambridge: University Press.

CRONQUIST, A. (1968). The Evolution and Classification of Flowering Plants. London and Edinburgh: Nelson.

CROW, J. F. (1969). Proc. Twelfth Intern. Cong. Genet. 3, 105.

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

DAVIS, H. C., SMITH, L. and WASSERMANN, A. R. (1964). Biochim. Biophys. Acta 85, 238.

DAYHOFF, M. O. (1969). Atlas of Protein Sequence and Structure, vol. 4. Silver Spring, Md.: National Biomedical Research Foundation.

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

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

DELANGE, R. J., GLAZER, A. N. and SMITH, E. L. (1970). J. Biol. Chem. 245, 3325.

DIANO, M. and MARTINEZ, G. (1971). C. R. Acad. Sc. Paris D272, 2692.

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

DICKERSON, R. E., TAKANO, T., EISENBERG, D., KALLAI, O. B., SAMSON, 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.

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

EASLEY, C. W. (1965). Biochim. 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 ed., vol. 1, p.71. New York: Wiley.

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. (1971a). J. Molec. Evolution 1, 84.

FITCH, W. M. (1971b). Biochem. Genet. 5, 231.

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. (1971). Evolutionary Biology 4, in the press.

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

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

FLATMARK, T. (1964b). Acta Chem. Scand. 18, 1656.

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

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

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

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

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

GRAY, W. R. (1967). In Methods in Enzymology, vol. 11, (Colowick, S. P. and Kaplan, N. O. eds.), p.469. New York and London: Academic Press.

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.

GROSS, D. (1961). J. Chromatog. 5, 194.

GURTLER, L. and HORSTMANN, H. J. (1970). Eur. J. Biochem. 12, 48.

HAGIHARA, B., TAGAWA, K., MORIKAWA, I., SHIN, M. and OKUNUKI, K. (1958). J. Biochem. (Tokyo) 45, 725.

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

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

HARDY, M., HARRIS, I., PERRY, S. V. and STONE, D. (1970a). Biochem. J. 117, 44P.

HARDY, M. F., HARRIS, C. I., PERRY, S. V. and STONE, D. (1970b). Biochem. J. 120, 653.

HAWKES, J. G. (1968). Chemotaxonomy and Serotaxonomy. New York: Academic Press.

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

HEMPEL, K., LANGE, H. W. and BIRKOFER, L. (1968). Naturwissenschaften 55, 37.

HENDERSON, R. W. and RAWLINSON, W. A. (1961). In Haematin Enzymes (Falk, J. E., Lemberg, R. and Morton, R. K., eds.), p.370. Oxford: Pergamon Press.

HENDERSON, R. W. and ADA, G. L. (1963). Biochim. Biophys. Acta 77, 513.

HENDERSON, R. W. and PALEUS, S. (1963). Acta Chem. Scand. 17, S110.

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. B107, 286.

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

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

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

JARDINE, N., van RIJSBERGEN, C. J. and JARDINE, C. J. (1969). Nature 224, 185.

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

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

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

KEILIN, D. (1925). Proc. Roy. Soc. B98, 312.

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

KEILIN, D. (1930). Proc. Roy. Soc. B106, 418.

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

KEILIN, D. and HARTREE, E. F. (1937). Proc. Roy. Soc. B122, 298.

KEILIN, D. and HARTREE, E. F. (1939). Proc. Roy. Soc. B127, 167.

KIMURA, M. (1968). Nature 217, 624.

KIMURA, M. (1969). Proc. U.S. Natl. Acad. Sci. 63, 1181.

KIMURA, M. and OHTA, T. (1971a). Nature 229, 467.

KIMURA, M. and OHTA, T. (1971b). J. Molec. Evolution 1, 1.

KING, J. L. and JUKES, T. H. (1969). Science 164, 788.

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

KOHNE, D. E. (1968). In Chemotaxonomy and Serotaxonomy, (Hawkes, J. G., ed.) Systematics Association Special Vol. 2, p.117.

KRAUT, J., SINGH, S. and ALDEN, R. A. (1968). In Structure and Function of Cytochromes, (Okunuki, K., Kamen, M. D. and Sekuzu, I., eds.), p.252. Tokyo: University Press.

KUBY, S. A., PALEUS, S., PAUL, K.-G. and THEORELL, H. (1956). Acta Chem. Scand. 10, 148.

KUEHL, W. M. and ADELSTEIN, R. S. (1969). Biochem. Biophys. Res. Commun. 37, 59.

LANCE, G. N. and WILLIAMS, W. T. (1965). Comp. J. 8, 246.

LANCE, G. N. and WILLIAMS, W. T. (1966a). Nature 212, 218.

LANCE, G. N. and WILLIAMS, W. T. (1966b). Comp. J. 9, 60.

LANCE, G. N. and WILLIAMS, W. T. (1967a). Comp. J. 9, 373.

LANCE, G. N. and WILLIAMS, W. T. (1967b). Comp. J. 10, 271.

LAWRENCE, G. H. M. (1951). Taxonomy of Vascular Plants. New York: Macmillan.

LAYCOCK, M. V. (1968). Unpublished experiments.

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

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

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. (1954a). Biochem. J. 56, 529.

MARGOLIASH, E. (1954b). Biochem. J. 56, 535.

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

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

MARGOLIASH, E. (1964a). Unpublished experiments; cited by Margoliash and Schejter (1966).

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

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

MARGOLIASH, E. (1969b). Discussion to paper by 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, (Bryson, V. and Vogel, H. J., eds.), p.221. New York: Academic Press.

MARGOLIASH, E. and WALASEK, O. F. (1967). In Methods in Enzymology, vol. 10, (Colowick, S. P. and Kaplan, N. O., eds.) p.339. New York: Academic Press.

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. R. (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. (1970). Nature 228, 723.

MASSEY, V. (1959). Biochim. Biophys. Acta 34, 255.

MATSUBARA, H. and SASAKI, R. M. (1969). Biochem. Biophys. Res. Commun. 35, 175.

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

MAYNARD SMITH, J. (1970). Amer. Nat. 104, 231.

MEATYARD, B. T. (1971). Unpublished experiments.

MICHL, H. (1951). Mh. Chem. 82, 489.

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

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

MORRISON, M., HOLLOCHER, T. and STOTZ, E. (1961). Arch. Biochem. Biophys. 92, 338.

NAKAYAMA, T., TITANI, K. and NARITA, K. (1971). Unpublished experiments cited by Dickerson et al. (1971).

NARITA, K. and SUGENO, K. (1968). In Structure and Function of Cytochromes (Okunuki, K., Kamen, M. D. and Sekuzu, I., eds.), 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.

NIELANDS, J. B. (1952). J. Biol. Chem. 197, 701.

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

NOLAN, C. and MARGOLIASH, E. (1968). Ann. Rev. Biochem. 37, 727.

NOZAKI, M., YAMANAKA, T., HORIO, T. and OKUNUKI, K. (1957). J. Biochem. (Tokyo) 44, 453.

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

OKUNUKI, K. (1960). In A Laboratory Manual of Analytical Methods of Protein Chemistry, vol. 1, (Alexander, P. and Block, R. J. eds.), p.32. Oxford: Pergamon Press.

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

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

PAN, S. C. and DUTCHER, J. D. (1956). Analyt. Chem. 28, 836.

PATTERSON, E. L., VAN METER, J. C. and BOHONOS, N. (1964). J. Med. Chem. 7, 689.

PETRAS, M. L., REIMER, J. D., BIDDLE, F. G., MARTIN, J. E. and LINTON, R. S. (1969). Canad. J. Genet. Cytol. 11, 497.

PETTIGREW, G. and MEYER, T. (1971). Biochem. J. 125, 46P.

POTGIETER, D. J. J. (1964). Personal communication cited by Margoliash and Schejter (1966).

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

PRAKASH, S., LEWONTIN, R. C. and HUBBY, J. L. (1969). Genetics 61, 841.

RICHARDSON, D. L. (1971). Unpublished experiments.

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

RICHARDSON, M., RICHARDSON, D. L., RAMSHAW, J. A. M., THOMPSON, E. W. and BOULTER, D. (1971b). J. Biochem. (Tokyo) 69, 811.

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

ROMER, A. S. (1966). Vertebrate Paleontology, 3rd Edition. Chicago: The University of Chicago Press.

RYDON, H. N. and SMITH, P. W. G. (1952). Nature 169, 922.

SACKIN, M. J. (1971). Biochem. Genet. 5, 287.

SANO, S. and TANAKA, K. (1964). J. Biol. Chem. 239, PC3109.

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

SCOGIN, R. (1971). Unpublished experiments.

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

SELS, A. A., FUKUHARA, H., PERE, G. and SLONIMSKI, P. P. (1965). Biochim. Biophys. Acta 95, 486.

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, G. J., GARDINSKY, R. L. and MARGOLIASH, E. (1968). J. Biol. Chem. 243, 5446.

SHOTTON, D. M. and HARTLEY, B. S. (1970). Nature 225, 802.

SIMPSON, G. G. (1964). Personal communication cited by Smith and Margoliash (1964).

SLONIMSKI, P. P., ACHER, R., PERE, G., SELS, A. and SOMLO, M. (1965). In Mechanismes de Regulation des Activites Cellulaires chez les Microorganismes, p.435. Paris: Editions du Centre National de la Recherche Scientifique.

SMITH, A. E. and MARCKER, K. (1970). Nature 226, 607.

SMITH, E. L. (1968). Harvey Lectures 62, 231.

SMITH, E. L. and KIMMEL, J. R. (1960). In The Enzymes, vol. 4, (Boyer, P. D., Lardy, H. and Myrbäck, K. eds.), p.133. New York: Academic Press.

SMITH, E. L. and MARGOLIASH, E. (1964). Fed. Proc. 23, 1243.

SMITH, E. L., MARKLAND, F. S., KASPER, C. B., DELANGE, R. J., LANDON, M. and EVANS, W. H. (1966). J. Biol. Chem. 241, 5974.

SMITH, I. (1958). Chromatographic and Electrophoretic Techniques, vol. 1, p.7. London: Heinemann.

SMYTH, D. G. (1967). In Methods in Enzymology, vol. 11, (Colowick, S. P. and Kaplan, N. O. eds.), p.214. New York: Academic Press.

SNEATH, P. H. A. (1969). Personal communication cited by Fitch and Margoliash (1969).

SOKAL, R. R. and MICHENER, C. D. (1958). Univ. Kansas Sci. Bull. 38, 1409.

SOKAL, R. R. and SNEATH, P. H. A. (1963). Principles of Numerical Taxonomy. San Francisco: Freeman and Co.

SPORNE, K. R. (1971). The Mysterious Origin of Flowering Plants. Oxford: University Press.

STEBBINS, G. L. (1950). Variation and Evolution in Plants. New York: Columbia University Press.

STELLWAGEN, E. (1968). Biochemistry 7, 2893.

STEVENS, F. C., GLAZER, A. N. and SMITH, E. L. (1967). J. Biol. Chem. 242, 2764.

SEWARD, A. C. (1937). Sci. Progress 32, 420.

STEWART, I. (1963). J. Chromatog. 10, 404.

STEWART, J. W. and MARGOLIASH, E. (1965). Can. J. Biochem. 43, 1187.

TAKHTAJAN, A. (1969). Flowering Plants - Origin and Dispersal. Edinburgh: Oliver and Boyd.

TECHNICON (1963). Instructions for Assembly and Operation of Amino Acid Analyser. Chertsey, Surrey: Technicon Instruments Company Ltd.

TECHNICON (1968). 6th Colloquium in Amino Acid Analysis, Monograph 3, p.162. Geneva: Technicon International Division S.A.

TERTZAKIAN, G., HASKINS, R. H., SLATER, G. P. and NESBITT, L. R. (1964). Proc. Chem. Soc. 1964, 195.

THEORELL, H. (1936). Biochem. Z. 285, 207.

THEORELL, H. and ÅKESSON, Å. (1939). Science 90, 67.

THOMPSON, E. W., RICHARDSON, M. and BOULTER, D. (1970). Biochem. J. 121, 439.

THOMPSON, E. W., RICHARDSON, M. and BOULTER, D. (1971a). Biochem. J. 124, 779.

THOMPSON, E. W., RICHARDSON, M. and BOULTER, D. (1971b). Biochem. J. 124, 783.

THOMPSON, E. W., NOTTON, B. A., RICHARDSON, M. and BOULTER, D. (1971c). Biochem. J. 124, 787.

TITANI, K., ISHIKURA, H. and MINAKAMI, S. (1959). J. Biochem. (Tokyo) 46, 151.

TURRIL, W. B. (1942). Bot. Rev. 8, 655.

UPHAUS, R. A., GROSSWEINER, L. I., KATZ, J. J. and KOPPLE, K. D. (1959). Science 129, 641.

VALENTINE, J. and ROBERTS, S. J. (1971). Unpublished experiments.

VAN GELDER, B. F. and SLATER, E. C. (1962). Biochim. Biophys. Acta 58, 593.

VAUGHAN, J. G. (1968). Sci. Progress 56, 205.

WADA, K. and OKUNUKI, K. (1969). J. Biochem. (Tokyo) 66, 249.

WALASEK, O. F. and MARGOLIASH, E. (1969). Unpublished experiments cited by Margoliash, Fitch and Dickerson (1969).

WALTON, J. (1953). An Introduction to the Study of Fossil Plants, 2nd edition. London: A. and C. Black.

WASSERMANN, A. R., GARVER, J. C. and BURRIS, R. H. (1963). Phytochem. 2, 7.

WATSON, L., WILLIAMS, W. T. and LANCE, G. N. (1966). J. Linn. Soc. 59, 491.

WHITTAKER, R. H. (1969). Science 163, 150.

WILLIAMS, W. T., LAMBERT, J. M. and LANCE, G. N. (1966). J. Ecol. 54, 427.

WOODS, K. R. and WANG, K. T. (1967). Biochim. Biophys. Acta 133, 369.

WOOLHOUSE, H. W. (1970). In Phytochemical Phylogeny, (Harborne, J. B. ed.), p.207. London and New York: Academic Press.

WRIGHT, C. S., ALDEN, R. A. and KRAUT, J. (1969). Nature 221, 235.

WRIGHT, S. (1931). Genetics 16, 97.

YAMANAKA, T. (1966). Ann. Report Biol. Works, Faculty of Sci., Osaka Univ. 14, 1.

YAMANAKA, T., MIZUSHIMA, H., NOZAKI, M., HORIO, T. and OKUNUKI, K. (1959). J. Biochem. (Tokyo) 46, 121.

YAMANAKA, T., MIZUSHIMA, H. and OKUNUKI, K. (1964a). Biochem. Biophys. Acta 81, 223.

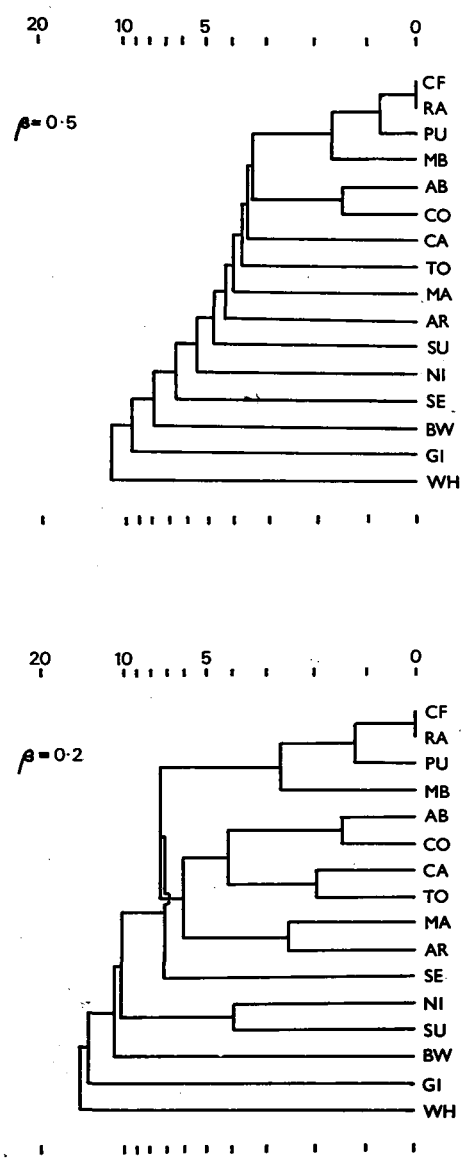
YAMANAKA, T., MIZUSHIMA, H., KATANO, H. and OKUNUKI, K. (1964b). Biochim. Biophys. Acta 85, 11.

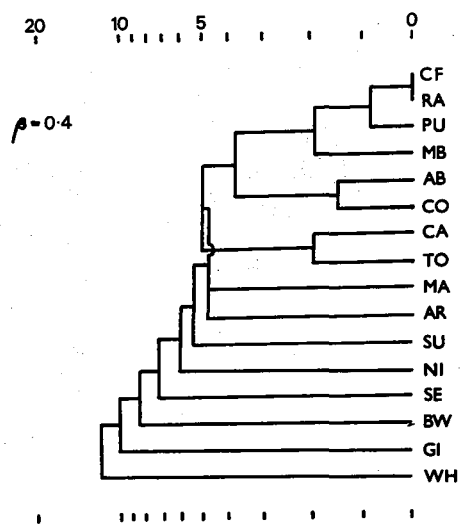
ZUCKERKANDL, E. and PAULING, L. (1962). In Horizons in Biochemistry, (Kasha, M. and Pullman, B. eds.), p.189. New York: Academic Press.

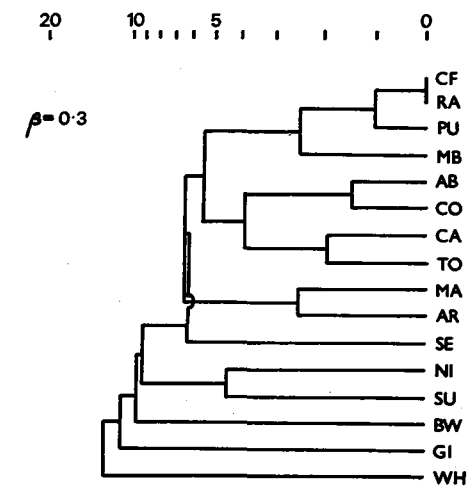
ZUCKERKANDL, E. and PAULING, L. (1965). J. Theoret. Biol. 8, 357.

FIGURE 25. A series of phylogenetic trees relating sixteen plant species.

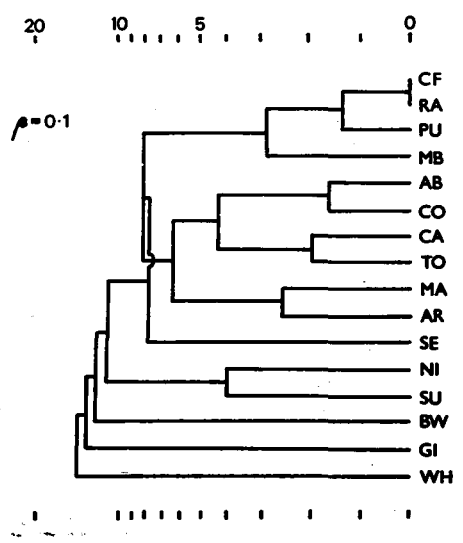
The trees were constructed by the 'flexible' method of Lance & Williams (1967a) using amino acid differences between species. The values of ß which were used, are indicated.

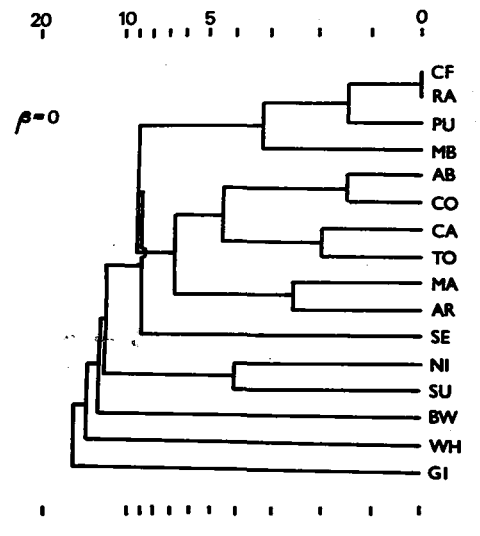

β=0·5
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
TO
MA
AR
SU
NI
SE
BW
GI
WH

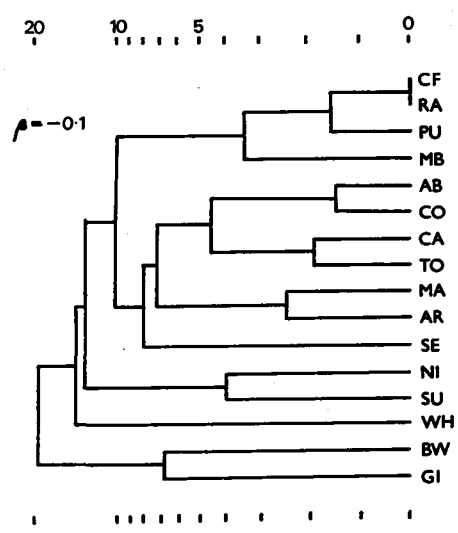

β=0·4
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
TO
MA
AR
SU
NI
SE
BW
GI
WH

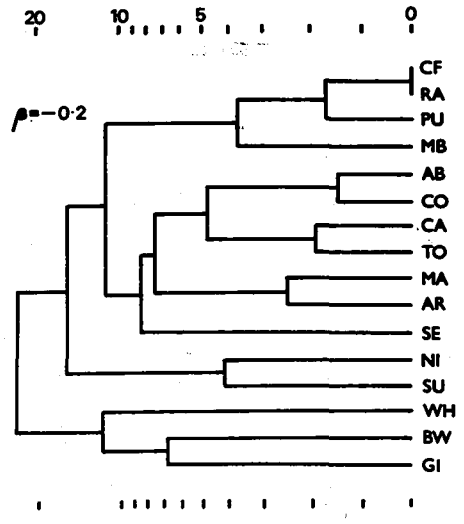

β=0·3
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
TO
MA
AR
SE
NI
SU
BW
GI
WH

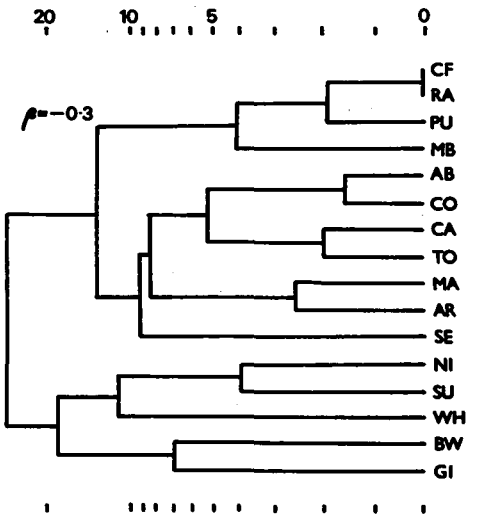
β=0·2
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
TO
MA
AR
SE
NI
SU
BW
GI
WH

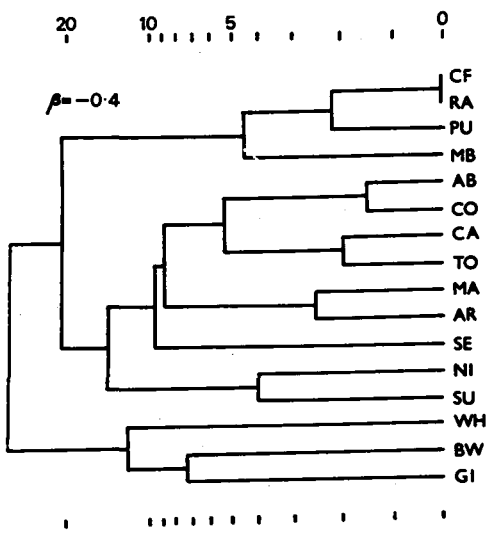

β=0·1
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
TO
MA
AR
SE
NI
SU
BW
GI
WH

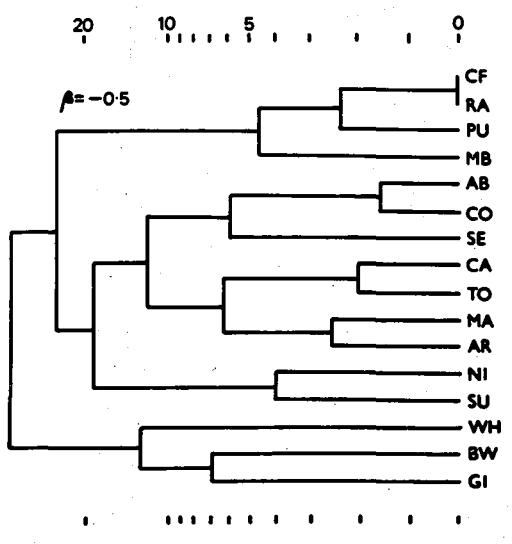

β=0
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
TO
MA
AR
SE
NI
SU
BW
WH
GI

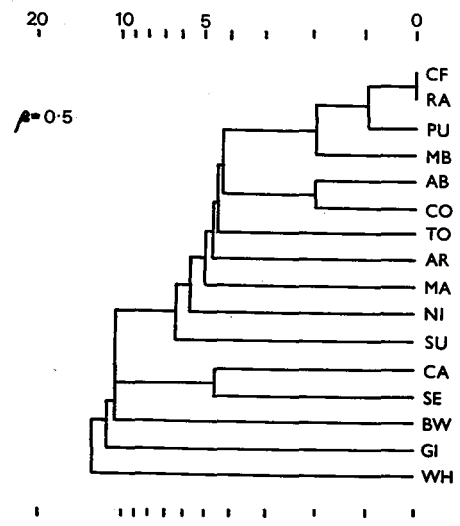

β=−0·1
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
TO
MA
AR
SE
NI
SU
WH
BW
GI

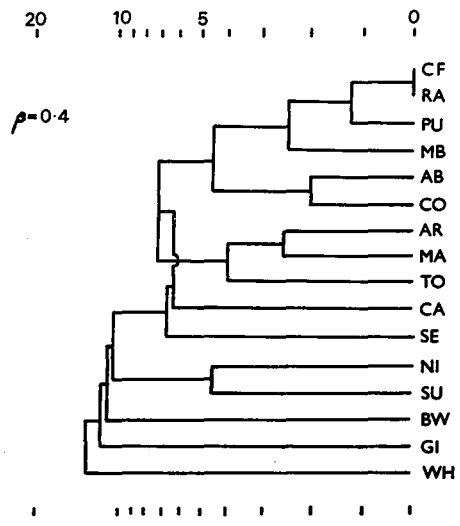

β=−0·2
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
TO
MA
AR
SE
NI
SU
WH
BW
GI

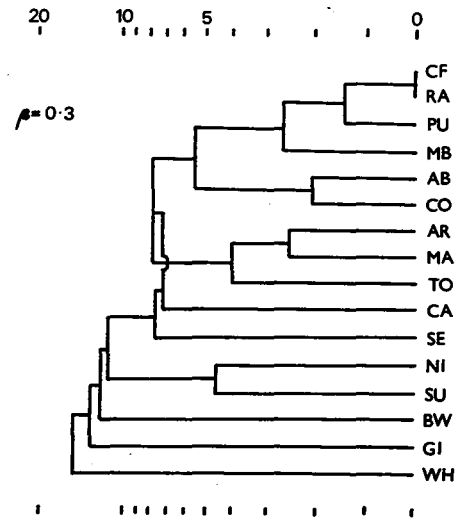

β=−0·3
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
TO
MA
AR
SE
NI
SU
WH
BW
GI

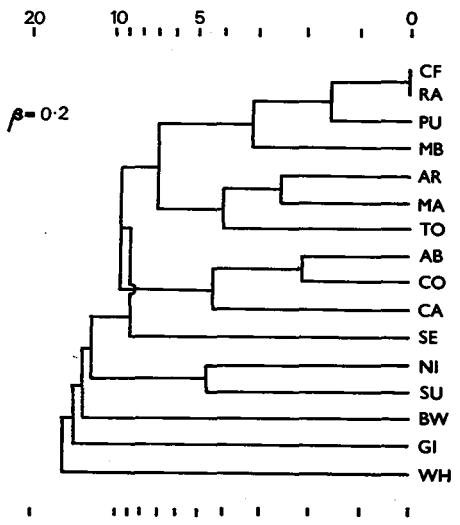

β=−0·4
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
TO
MA
AR
SE
NI
SU
WH
BW
GI

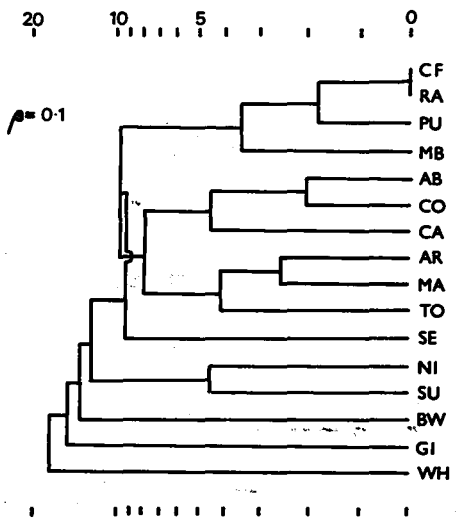

β=−0·5
20 10 5 0
CF
RA
PU
MB
AB
CO
SE
CA
TO
MA
AR
NI
SU
WH
BW
GI

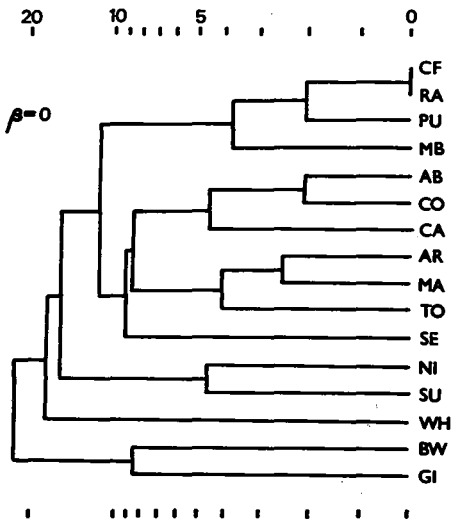

β=0·5
20 10 5 0
CF
RA
PU
MB
AB
CO
TO
AR
MA
NI
SU
CA
SE
BW
GI
WH

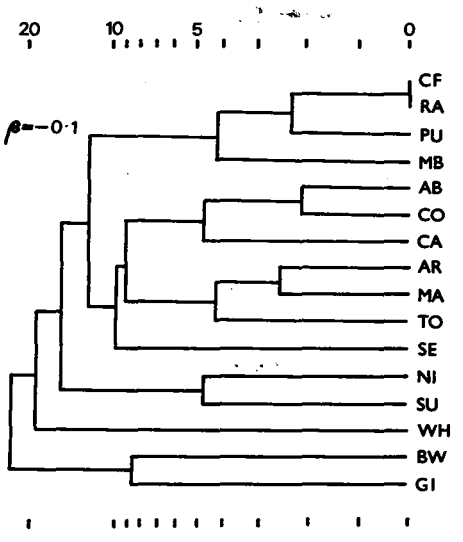

β=0·4
20 10 5 0
CF
RA
PU
MB
AB
CO
AR
MA
TO
CA
SE
NI
SU
BW
GI
WH

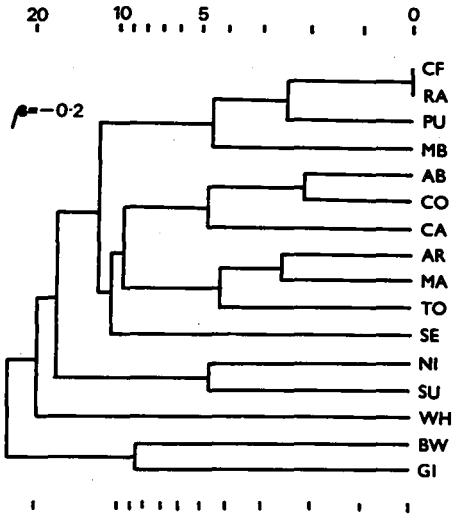

β=0·3
20 10 5 0
CF
RA
PU
MB
AB
CO
AR
MA
TO
CA
SE
NI
SU
BW
GI
WH

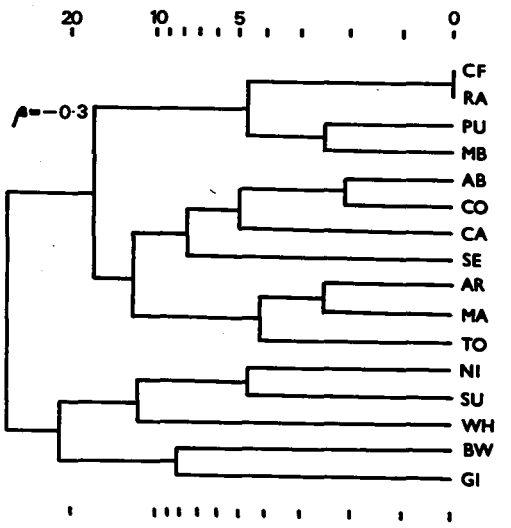

β=0·2
20 10 5 0
CF
RA
PU
MB
AR
MA
TO
AB
CO
CA
SE
NI
SU
BW
GI
WH

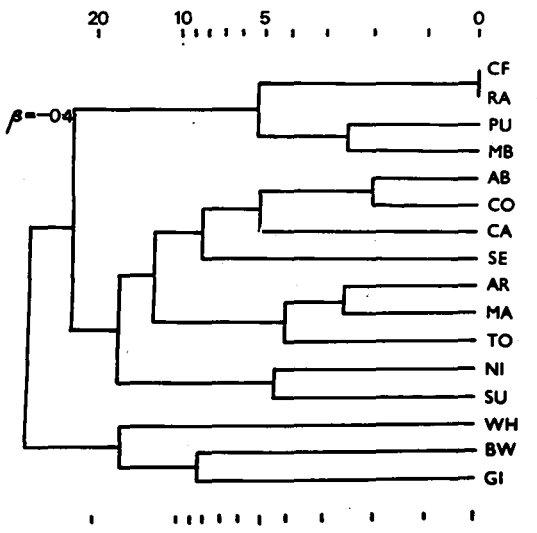

β=0·1
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
AR
MA
TO
SE
NI
SU
BW
GI
WH

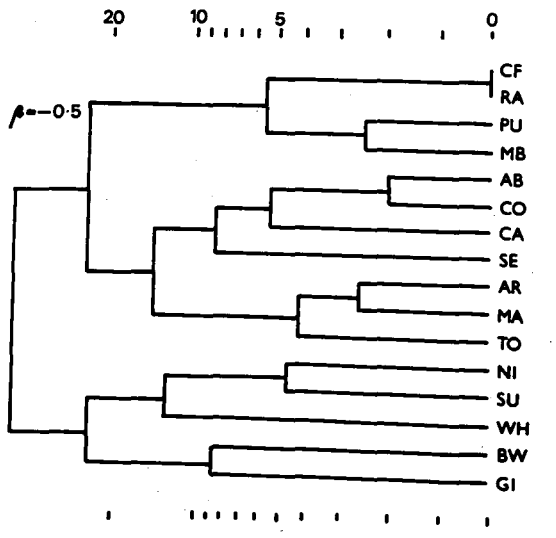

β=0
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
AR
MA
TO
SE
NI
SU
WH
BW
GI


β=−0·1
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
AR
MA
TO
SE
NI
SU
WH
BW
GI


β=−0·2
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
AR
MA
TO
SE
NI
SU
WH
BW
GI


β=−0·3
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
SE
AR
MA
TO
NI
SU
WH
BW
GI


β=−04
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
SE
AR
MA
TO
NI
SU
WH
BW
GI


β=−0·5
20 10 5 0
CF
RA
PU
MB
AB
CO
CA
SE
AR
MA
TO
NI
SU
WH
BW
GI

APPENDIX 1. MATRIX OF AMINO ACID DIFFERENCES

	Human	Monkey	Horse	Donkey	Bovine	Dog	Whale	Rabbit	Kangaroo	Turkey	Penguin	Duck	Pigeon	Turtle	Rattlesnake	Bullfrog	Tuna Fish	Dogfish	Lamprey	Bonito	Carp	Fruitfly	Screwfly	Samia Moth	Tobacco Moth	Neurospora	Candida	Saccharomyces	Debaryomyces	Abutilon	Arum	Buckwheat	Castor	Cauliflower	Cotton	Ginkgo	Maize	Mungbean	Niger	Pumpkin	Sesame	Sunflower	Tomato	Wheat	Enteromorpha
Human	0	1	12	11	10	11	10	9	10	13	13	11	12	15	14	18	21	24	20	21	19	29	27	31	31	49	52	46	47	45	45	42	44	46	47	44	44	48	45	45	41	44	44	45	46
Monkey	1	0	11	10	9	10	9	8	11	12	12	10	11	14	15	17	21	23	20	21	19	28	26	30	30	48	52	46	47	45	46	41	44	46	47	44	44	49	45	46	42	45	44	45	46
Horse	12	11	0	1	3	6	5	6	7	11	12	10	11	11	22	14	19	17	16	18	41	24	22	29	28	47	52	47	46	49	49	46	48	49	51	47	47	51	48	49	46	48	48	48	46
Donkey	11	10	1	0	2	5	4	5	8	10	11	9	10	10	21	13	18	16	15	17	13	24	22	28	27	47	51	46	45	48	48	45	47	48	50	46	46	50	47	48	45	47	47	47	45
Bovine	10	9	3	2	0	3	2	4	6	9	10	8	9	9	20	11	17	16	14	61	12	24	22	27	27	47	51	46	45	47	48	44	46	48	49	45	46	50	47	48	44	47	46	47	45
Dog	11	10	6	5	3	0	3	5	7	10	10	8	9	9	21	12	18	17	14	17	12	23	21	25	25	47	50	46	44	46	47	43	45	47	48	44	45	49	46	47	43	46	45	46	44
Whale	10	9	5	4	2	3	0	2	6	9	9	7	8	8	19	11	17	16	15	16	21	24	22	27	27	47	51	46	45	47	47	43	45	47	48	44	45	49	46	48	44	46	45	46	44
Rabbit	9	8	6	5	4	5	2	0	6	8	8	6	7	9	18	11	17	17	17	17	14	23	21	26	26	47	51	46	45	47	47	42	45	47	48	44	45	49	46	48	44	46	45	46	44
Kangaroo	10	11	7	8	6	7	6	6	0	12	10	10	11	11	21	13	18	20	17	18	14	26	24	28	28	50	52	47	47	47	47	44	46	48	49	47	46	49	47	47	43	46	46	49	44
Turkey	13	12	11	10	9	10	9	8	12	0	2	3	4	8	19	11	17	19	18	17	16	25	23	28	28	48	51	47	45	49	49	42	48	49	50	45	47	51	48	50	47	48	47	48	45
Penguin	13	12	12	11	10	10	9	8	10	2	0	3	4	8	20	12	18	20	19	18	16	26	24	27	27	49	50	46	45	49	49	42	48	49	50	45	47	51	48	50	47	48	47	48	44
Duck	11	10	10	9	8	8	7	6	10	3	3	0	3	7	17	11	17	17	18	17	15	24	22	27	27	47	51	47	45	47	47	42	45	47	48	45	45	49	46	48	44	46	45	48	43
Pigeon	12	11	11	10	9	9	8	7	11	4	4	3	0	8	18	12	18	19	19	18	16	25	23	27	26	47	51	47	45	47	47	42	45	47	48	44	45	49	46	48	45	46	45	48	44
Turtle	15	14	11	10	9	9	8	9	11	8	8	7	8	0	22	10	18	19	19	17	14	24	24	28	29	50	53	50	47	47	47	42	45	47	48	45	45	49	46	48	44	46	45	48	43
Rattlesnake	14	15	22	21	20	21	19	18	21	19	20	17	18	22	0	24	26	26	27	27	26	31	29	31	33	49	53	49	50	45	44	43	45	46	46	46	44	47	45	46	42	44	44	48	48
Bullfrog	18	17	14	13	11	12	11	11	13	11	12	11	12	10	24	0	15	20	21	15	14	22	22	29	30	50	52	48	48	51	50	45	48	50	51	47	48	52	49	52	48	49	48	50	47
Tuna Fish	21	21	19	18	17	18	17	17	18	17	18	17	18	18	26	15	0	20	19	2	8	25	24	32	30	49	48	48	47	52	51	48	50	51	53	50	48	53	50	52	49	50	50	51	47
Dogfish	24	23	17	16	16	17	16	17	20	19	20	17	19	19	26	20	20	0	17	20	15	26	25	32	31	50	53	49	48	52	51	50	51	52	53	53	50	53	50	53	50	50	51	51	48
Lamprey	20	20	16	15	14	14	15	17	17	18	19	18	19	19	27	21	19	17	0	19	12	29	28	32	33	52	56	50	49	52	52	51	53	54	54	53	51	54	51	54	51	51	53	53	50
Bonito	21	21	18	17	16	17	16	17	18	17	18	17	18	17	27	15	2	20	19	0	8	26	25	33	31	47	47	46	46	50	50	48	48	50	51	49	46	52	48	51	47	48	48	49	45
Carp	19	19	14	13	12	12	12	14	14	16	16	15	16	14	26	14	8	15	12	8	0	23	22	28	27	49	51	48	46	50	50	48	49	50	52	49	47	52	49	50	47	49	49	50	45
Fruitfly	29	28	24	24	24	23	24	23	26	25	26	24	25	24	31	22	25	26	29	26	23	0	2	15	14	42	48	46	43	52	50	46	49	50	52	49	48	52	48	52	49	48	49	49	45
Screwfly	27	26	22	22	22	21	22	21	24	23	24	22	23	24	29	22	24	25	28	25	22	2	0	14	12	42	48	46	44	50	48	44	47	48	50	47	46	50	46	50	47	46	47	47	43
Samia Moth	31	30	29	28	27	25	27	26	28	28	27	27	27	28	31	29	32	32	33	28	15	14	0	5	48	48	48	45	43	45	41	47	47	45	48	45	47	45	47	46	46	45	47	45	
Tobacco Moth	31	30	28	27	27	25	27	26	28	28	27	27	26	29	33	30	30	31	33	31	27	14	12	5	0	47	47	46	44	45	46	42	45	46	47	46	44	48	45	46	46	45	44	43	
Neurospora	49	48	47	47	47	47	47	47	50	48	49	47	47	50	49	50	49	50	52	47	49	42	42	48	47	0	42	41	46	55	57	56	55	53	55	57	54	55	55	54	54	55	55	55	53
Candida	52	52	52	51	51	50	51	51	52	51	50	51	51	53	53	52	48	53	56	47	51	48	48	48	47	42	0	28	25	51	52	55	52	51	52	56	52	53	53	52	53	53	52	51	52
Saccharomyces	46	46	47	46	46	46	46	46	47	47	46	47	47	50	49	48	48	49	50	46	48	46	46	48	46	41	28	0	29	48	51	49	48	49	49	50	48	51	47	50	50	49	49	48	51
Debaryomyces	47	47	46	45	45	44	45	45	47	45	45	45	45	47	50	48	47	48	49	46	46	43	44	45	44	46	25	29	0	53	53	54	52	51	54	54	52	54	52	53	54	53	53	50	49
Abutilon	45	45	49	48	47	46	47	47	47	49	49	47	47	47	45	51	52	52	52	50	50	52	50	43	45	55	51	48	53	0	7	12	6	9	3	18	8	9	10	8	8	12	5	15	27
Arum	45	46	49	48	48	47	47	47	47	49	49	47	47	47	44	50	51	51	52	50	50	50	48	45	46	57	52	51	53	7	0	11	8	8	7	17	5	7	8	8	9	7	5	14	26
Buckwheat	42	41	46	45	44	43	43	42	44	42	42	42	42	42	43	45	48	50	51	48	48	46	44	41	42	56	55	49	54	12	11	0	12	13	12	14	10	13	13	15	13	15	11	17	27
Castor	44	44	48	47	46	45	45	45	46	48	48	45	45	45	45	48	50	51	53	48	49	49	47	47	45	55	52	48	52	6	8	12	0	7	5	16	7	9	11	8	8	12	4	13	25
Cauliflower	46	46	49	48	48	47	47	47	48	49	49	47	47	47	46	50	51	52	54	50	50	50	48	47	46	53	51	49	51	9	8	13	7	0	7	18	6	5	11	3	10	14	7	10	24
Cotton	47	47	51	50	49	48	48	48	49	50	50	48	48	48	46	51	53	53	54	51	52	52	50	45	47	55	52	49	54	3	7	12	5	7	0	19	8	7	10	8	8	12	6	13	25
Ginkgo	44	44	47	46	45	44	44	44	47	45	45	45	44	45	46	47	50	53	53	49	49	49	47	48	46	57	56	50	54	18	17	14	16	18	19	0	14	20	18	20	18	20	14	17	31
Maize	44	44	47	46	46	45	45	45	46	47	47	45	45	45	44	48	48	50	51	46	47	48	46	45	44	54	52	48	52	8	5	10	7	6	8	14	0	10	8	8	9	10	4	10	24
Mungbean	48	49	51	50	50	49	49	49	49	51	51	49	49	49	47	52	53	53	54	52	52	52	50	47	48	55	53	51	54	9	7	13	9	5	7	20	10	0	13	5	9	12	9	15	27
Niger	45	45	48	47	47	46	46	46	47	48	48	46	46	46	45	49	50	50	51	48	49	48	46	45	45	55	53	47	52	10	8	13	11	11	10	18	8	13	0	13	12	8	10	16	26
Pumpkin	45	46	49	48	48	47	48	48	47	50	50	48	48	48	46	52	52	53	54	51	50	52	50	47	46	54	52	50	53	8	8	15	8	3	8	20	8	5	13	0	8	14	8	13	26
Sesame	41	42	46	45	44	43	44	44	43	47	47	44	45	44	42	48	49	50	51	47	47	49	47	46	46	54	53	50	54	8	9	13	8	10	8	18	9	9	12	8	0	11	7	16	25
Sunflower	44	45	48	47	47	46	46	46	46	48	48	46	46	46	44	49	50	50	51	48	49	48	46	46	46	55	53	49	53	12	7	15	12	14	12	20	10	12	8	14	11	0	10	15	24
Tomato	44	44	48	47	46	45	45	45	46	47	47	45	45	45	44	48	50	51	53	48	49	49	47	45	45	55	52	49	53	5	5	11	4	7	6	14	4	9	10	8	7	10	0	13	25
Wheat	45	45	48	47	47	46	46	46	49	48	48	48	48	48	48	50	51	51	53	49	50	49	47	47	44	55	51	48	50	15	14	17	13	10	13	17	10	15	16	13	16	15	13	0	23
Enteromorpha	46	46	46	45	45	44	44	44	44	45	44	43	44	43	48	47	47	48	50	45	45	45	43	45	43	53	52	51	49	27	26	27	25	24	25	31	24	27	26	26	25	24	25	23	0

APPENDIX 2. MATRIX OF MINIMUM MUTATION DISTANCES

	Human	Monkey	Horse	Donkey	Bovine	Dog	Whale	Rabbit	Kangaroo	Turkey	Penguin	Duck	Pigeon	Turtle	Rattlesnake	Bullfrog	Tuna Fish	Dogfish	Lamprey	Bonito	Carp	Fruitfly	Screwfly	Samia Moth	Tobacco Moth	Neurospora	Candida	Saccharomyces	Debaryomyces	Abutilon	Arum	Buckwheat	Castor	Cauliflower	Cotton	Ginkgo	Maize	Mungbean	Niger	Pumpkin	Sesame	Sunflower	Tomato	Wheat	Enteromorpha
Human	0	1	16	15	12	13	13	12	12	18	18	17	16	20	21	26	32	35	26	31	26	37	35	37	38	64	66	57	59	48	50	46	48	53	52	48	48	54	50	50	44	49	48	53	59
Monkey	1	0	15	14	11	12	12	11	13	17	17	16	15	18	22	25	33	34	27	32	27	36	34	36	37	63	66	58	59	48	52	45	48	53	52	48	49	55	49	52	45	50	48	53	59
Horse	16	15	0	1	5	9	9	10	10	15	16	15	15	15	33	22	29	23	21	27	21	27	25	33	31	66	69	63	57	57	57	54	57	57	60	54	56	59	56	56	54	57	57	60	61
Donkey	15	14	1	0	4	8	8	9	11	14	15	14	14	14	32	21	27	22	20	26	20	27	25	32	30	66	68	61	56	56	56	53	56	56	59	53	55	58	55	55	53	56	56	59	60
Bovine	12	11	5	4	0	3	3	5	7	12	13	12	12	12	30	16	26	23	18	25	19	30	27	31	32	66	68	61	58	54	56	50	54	56	57	51	55	58	55	55	50	56	54	59	61
Dog	13	12	9	8	3	0	4	7	8	14	14	12	12	13	30	20	30	25	22	29	20	27	25	27	29	66	66	61	56	52	54	47	52	54	55	50	53	56	53	53	48	54	52	58	59
Whale	13	12	9	8	3	4	0	2	8	11	11	10	10	10	27	17	26	25	22	25	20	30	27	32	33	65	67	63	57	55	55	49	53	55	56	50	54	57	54	56	52	55	53	58	60
Rabbit	12	11	10	9	5	7	2	0	8	11	11	10	9	11	26	17	26	27	24	25	22	27	25	30	31	64	67	61	58	54	55	48	52	56	55	51	53	58	53	57	50	54	52	56	59
Kangaroo	12	13	10	11	7	8	8	8	0	15	13	14	14	14	30	20	27	27	23	26	20	32	30	32	33	68	69	60	62	53	55	50	53	57	56	54	53	58	55	55	48	54	53	60	59
Turkey	18	17	15	14	12	14	11	11	15	0	2	3	5	9	28	17	26	27	26	25	24	31	29	34	36	63	69	65	63	56	56	47	55	56	57	53	54	58	54	57	53	56	53	60	58
Penguin	18	17	16	15	13	14	11	11	13	2	0	3	5	9	28	18	27	28	27	26	24	32	30	33	35	64	68	64	63	56	56	47	55	56	57	53	54	58	54	57	53	56	53	60	57
Duck	17	16	15	41	12	12	10	10	14	3	3	0	4	8	25	17	26	25	26	25	23	30	27	32	34	63	68	63	62	53	54	47	52	55	54	53	52	57	50	56	49	54	50	60	56
Pigeon	16	15	15	14	12	12	10	9	14	5	5	4	0	10	26	18	27	26	26	26	24	32	30	34	34	61	69	63	62	54	54	48	50	54	55	51	52	56	53	55	50	54	50	59	58
Turtle	20	18	15	14	12	13	10	11	14	9	9	8	10	0	30	16	27	26	28	26	22	32	32	36	38	67	70	67	62	54	54	47	53	54	55	52	53	56	52	55	50	55	52	61	57
Rattlesnake	21	22	33	32	30	30	27	26	30	28	28	25	26	30	0	38	39	39	39	41	41	44	42	42	45	63	68	61	64	50	50	48	50	54	52	53	49	54	50	54	47	52	49	58	61
Bullfrog	26	25	22	21	16	20	17	17	20	17	18	17	18	16	38	0	22	30	32	21	21	30	30	38	39	70	70	64	65	60	61	55	58	63	61	59	59	65	57	64	57	59	58	65	65
Tuna Fish	32	33	29	27	26	30	26	26	27	26	27	26	27	27	39	22	0	29	27	3	13	35	33	44	42	74	70	69	68	68	67	64	65	67	68	66	63	70	63	69	64	63	66	66	65
Dogfish	35	34	23	22	23	25	25	27	27	27	28	25	26	26	39	30	29	0	24	29	22	35	34	41	38	71	71	67	65	60	60	60	61	61	64	65	60	63	59	60	58	60	61	65	66
Lamprey	26	27	21	20	18	22	22	24	23	26	27	26	26	28	39	32	27	24	0	27	18	37	36	38	39	68	70	65	60	60	60	60	61	61	63	61	59	63	59	61	58	60	63	64	64
Bonito	31	32	27	26	25	29	25	25	26	25	26	25	26	26	41	21	3	29	27	0	13	36	34	45	43	70	68	66	66	65	64	63	61	64	65	64	59	67	59	66	60	59	63	63	61
Carp	26	27	21	20	19	20	20	22	20	24	24	23	24	22	41	21	13	22	18	13	0	32	31	38	37	72	71	67	64	63	63	60	61	61	65	60	59	65	59	61	58	61	63	64	61
Fruitfly	37	36	27	27	30	27	30	27	32	31	32	30	32	32	44	30	35	35	37	36	32	0	2	19	16	60	67	67	62	63	62	56	61	63	64	59	60	63	57	64	60	59	61	60	61
Screwfly	35	34	25	25	27	25	27	25	30	29	30	27	30	32	42	30	33	34	36	34	31	2	0	17	14	60	67	67	63	61	60	54	59	61	62	57	58	61	55	62	58	57	59	58	59
Samia Moth	37	36	33	32	31	27	32	30	32	34	33	32	34	36	42	38	44	41	38	45	38	19	17	0	6	62	63	62	62	52	55	49	57	58	55	55	55	55	54	57	54	57	54	56	59
Tobacco Moth	38	37	31	30	32	29	33	31	33	36	35	34	34	38	45	39	42	38	39	43	37	16	14	6	0	63	64	62	62	56	57	53	57	58	59	55	55	57	56	57	56	57	56	53	58
Neurospora	64	63	66	66	66	66	65	64	68	63	64	63	61	67	63	70	74	71	68	70	72	60	60	62	63	0	59	58	63	70	71	71	71	68	70	71	69	70	69	69	70	69	69	69	69
Candida	66	66	69	68	68	66	67	67	69	69	68	68	69	70	68	70	70	71	70	68	71	67	67	63	64	59	0	39	29	70	68	73	68	67	69	73	68	72	72	70	74	73	69	66	73
Saccharomyces	57	58	63	61	61	61	63	61	60	65	64	63	63	67	61	64	69	67	65	66	67	67	67	62	62	58	39	0	37	61	63	64	61	63	62	65	60	65	62	62	62	61	63	64	68
Debaryomyces	59	59	57	56	58	56	57	58	62	63	63	62	62	62	64	65	68	65	60	66	64	62	63	62	62	63	29	37	0	72	71	72	68	67	71	71	69	72	70	71	74	71	71	68	68
Abutilon	48	48	57	56	54	52	55	54	53	56	56	63	65	65	60	60	68	60	60	65	63	63	61	52	56	70	70	61	72	0	8	15	8	11	4	18	10	10	12	9	9	14	6	18	34
Arum	50	52	57	56	56	54	55	55	55	56	56	54	54	54	50	61	67	60	60	64	63	62	60	55	57	71	68	63	71	8	0	14	10	10	9	16	5	8	9	9	11	8	5	16	33
Buckwheat	46	45	54	53	50	47	49	48	50	47	47	47	48	47	48	55	64	60	60	63	60	56	54	49	53	71	73	64	72	15	14	0	16	17	16	16	13	16	15	18	14	18	14	22	36
Castor	48	48	57	56	54	52	53	52	53	55	55	52	50	53	50	58	65	61	61	61	61	61	59	57	57	71	68	61	68	8	10	16	0	8	6	18	10	11	13	10	9	15	6	16	34
Cauliflower	53	53	57	56	56	54	55	56	57	56	56	55	54	54	54	63	67	61	61	64	61	63	61	58	58	68	67	63	67	11	10	17	8	0	8	18	7	6	11	4	12	16	8	12	30
Cotton	52	52	60	59	57	55	56	55	56	57	57	54	55	55	52	61	68	64	63	65	65	64	62	55	59	70	69	62	71	4	9	16	6	8	0	20	10	9	12	10	10	15	8	15	33
Ginkgo	48	48	54	53	51	50	50	51	54	53	53	53	51	52	53	59	66	65	61	64	60	59	57	55	55	71	73	65	71	18	16	16	18	18	20	0	14	19	19	19	18	20	14	21	37
Maize	48	49	56	55	55	53	54	53	53	54	54	52	52	53	49	59	63	60	59	59	59	60	58	55	55	69	68	60	69	10	5	13	10	7	10	14	0	11	8	9	10	11	6	11	30
Mungbean	54	55	59	58	58	56	57	58	58	58	58	57	56	56	54	65	70	63	63	67	65	63	61	55	57	70	72	65	72	10	8	16	11	6	9	19	11	0	14	5	10	13	9	18	34
Niger	50	49	56	55	55	53	54	53	55	54	54	50	53	52	50	57	63	59	59	59	59	57	55	54	56	69	72	62	70	12	9	15	13	11	12	19	8	14	0	14	14	9	11	18	32
Pumpkin	50	52	56	55	55	53	56	57	55	57	57	56	55	55	54	64	69	60	61	66	61	64	62	57	57	69	70	62	71	9	9	18	10	4	10	19	9	5	14	0	9	15	8	16	33
Sesame	44	45	54	53	50	48	52	50	48	53	53	49	50	50	47	57	64	58	58	60	58	60	58	54	56	70	74	62	74	9	11	14	9	12	10	18	10	10	14	9	0	14	8	19	34
Sunflower	49	50	57	56	56	54	55	54	54	56	56	54	54	55	52	59	63	60	60	59	61	59	57	57	57	69	73	61	71	14	8	18	15	16	15	20	11	13	9	15	14	0	11	17	30
Tomato	48	48	57	56	54	52	53	52	53	53	53	50	50	52	49	58	66	61	63	63	63	61	59	54	56	69	69	63	71	6	5	14	6	8	8	14	6	9	11	8	8	11	0	16	31
Wheat	53	53	60	59	59	58	58	56	60	60	60	60	59	61	58	65	66	65	64	63	64	60	58	56	53	69	66	63	68	18	16	22	16	12	15	21	11	18	18	16	19	17	16	0	26
Enteromorpha	59	59	61	60	61	59	60	59	59	58	57	56	58	57	61	65	65	66	64	61	61	61	59	59	58	69	73	68	68	34	33	36	34	30	33	37	30	34	32	33	34	30	31	26	0

APPENDIX 3. CYTOCHROME c SEQUENCES

Species	1 0	2 0	3 0	4 0	5 0	6 0	7 0	8 0	9 0	1 0 0	1 1 0
Human	Ac	GDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGYSYTAANKNKGIIWGEDTLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKKATNE									
(Chimpanzee appears to be identical to Human).											
Rhesus Monkey	Ac	GDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGYSYTAANKNKGITWGEDTLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKKATNE									
Horse	Ac	GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFTYTDANKNKGITWKEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE									
Donkey	Ac	GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFSYTDANKNKGITWKEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE									
Bovine	Ac	GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFSYTDANKNKGITWGEETLMEYLENPKKYIPGTKMIFAGIKKKGEREDLIAYLKKATNE									
(Pig, Sheep) appear to be identical to Bovine.											
Dog	Ac	GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFSYTDANKNKGITWGEETLMEYLENPKKYIPGTKMIFAGIKKTGERADLIAYLKKATKE									
Whale	Ac	GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAVGFSYTDANKNKGITWGEETLMEYLENPKKYIPGTKMIFAGIKKKGERADLIAYLKKATNE									
Rabbit	Ac	GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAVGFSYTDANKNKGITWGEDTLMEYLENPKKYIPGTKMIFAGIKKKDERADLIAYLKKATNE									
Kangaroo	Ac	GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLNGIFGRKTGQAPGFTYTDANKNKGIIWGEDTLMEYLENPKKYIPGTKMIFAGIKKKGERADLIAYLKKATNE									
Chicken	Ac	GDIEKGKKIFVQKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAEGFSYTDANKNKGITWGEDTLMEYLENPKKYIPGTKMIFAGIKKKSERVDLIAYLKDATSK									
(Turkey appears to be identical to Chicken)											
Penguin	Ac	GDIEKGKKIFVQKCSQCHTVEKGGKHKTGPNLHGIFGRKTGQAEGFSYTDANKNKGITWGEDTLMEYLENPKKYIPGTKMIFAGIKKKSERADLIAYLKDATSK									
Pekin Duck	Ac	GDVEKGKKIFVQKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAEGFSYTDANKNKGITWGEDTLMEYLENPKKYIPGTKMIFAGIKKKSERADLIAYLKDATAK									
Snapping Turtle	Ac	GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLNGLIGRKTGQAEGFSYTEANKNKGITWGEETLMEYLENPKKYIPGTKMIFAGIKKKAERADLIAYLKDATSK									
Rattlesnake	Ac	GDVEKGKKIFTMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAVGYSYTAANKNKGIIWGDDTLMEYLENPKKYIPGTKMVFTGLSKKKERTNLIAYLKEKTAA									
Bullfrog	Ac	GDVEKGKKIFVQKCAQCHTCEKGGKHKVGPNLYGLIGRKTGQAAGFSYTDANKNKGITWGEDTLMEYLENPKKYIPGTKMIFAGIKKKGERQDLIAYLKSACSK									
Tuna Fish	Ac	GDVAKGKKTFVQKCAQCHTVENGGKHKVGPNLWGLFGRKTGQAEGYSYTDANKSKGIVWNNDTLMEYLENPKKYIPGTKMIFAGIKKKGERQDLVAYLKSATS									
Dogfish	Ac	GDVEKGKKVFVQKCAQCHTVENGGKHKTGPNLSGLFGRKTGQAQGFSYTDANKSKGITWGQETLRIYLENPKKYIPGTKMIFAGLKKKSERQDLIAYLKKTAAS									
Lamprey	Ac	GDVEKGKKVFVQKCSQCHTVEKAGKHKTGPNLQGLFGRKTGQAPGFSYTDANKSKGIVWNQETLFVYLENPKKYIPGTKMIFAGIKKEGERKDLIAYLKKSTSE									
Bonito	Ac	GDVAKGKKTFVQKCAQCHTVENGGKHKVGPNLWGLFGRKTGQAEGYSYTDANKSKGIVWNENTLMEYLENPKKYIPGTKMIFAGIKKKGERQDLVAYLKSATS									
Carp	Ac	GDVEKGKKVFVQKCAQCHTVENGGKHKVGPNLWGLFGRKTGQAPGFSYTDANKSKGIVWNQETLMEYLENPKKYIPGTKMIFAGIKKKGERADLIAYLKSATS									
Fruitfly	GVPAGDVEKGKKLFVQRCAQCHTVEAGGKHKVGPNLHGLIGRKTGQAAGFAYTNANKAKGITWQDDTLFEYLENPKKYIPGTKMIFAGLKKPNERGDLIAYLKSATK										
Screw Worm Fly	GVPAGDVEKGKKIFVQRCAQCHTVEAGGKHKVGPNLHGLFGRKTGQAAGFAYTNANKAKGITWQDDTLFEYLENPKKYIPGTKMIFAGLKKPNERGDLIAYLKSATK										
Samia cynthia	GVPAGNAENGKKIFVQRCAQCHTVEAGGKHKVGPNLHGFYGRKTGQAPGFSYSNANKAKGITWGDDTLFEYLENPKKYIPGTKMVFAGLKKANERADLIAYLKESTK										
Tobacco Moth	GVPAGNADNGKKIFVQRCAQCHTVEAGGKHKVGPNLHGFFGRKTGQAPGFSYSNANKAKGITWQDDTLFEYLENPKKYIPGTKMVFAGLKKANERADLIAYLKQATK										
Neurospora crassa	GFSAGDSKKGANLFKTRCAECHGEGGNLTQKIGPALHGLFGRKTGSVDGYAYTDANKQKGITWDENTLFEYLENPJKYIPGTKMAFGGLKKDKDRNDIITFMKEATA										
Candida krusei	PAPFEQGSAKKGATLFKTRCAECHTIEAGGPHKVGPNLHGIFSRHSGQAEGYSYTDANKRAGVEWAEPTMSDYLENPJKYIPGTKMAFGGLKKAKDRNDLVTYMLEASK										
Saccharomyces oviformis iso-1, Saccharomyces cerevisiae	TEFKAGSAKKGATLFKTRCELCHTVEKGGPHKVGPNLHGIFGRHSGQAQGYSYTDANIKKNVLWDENNMSEYLTNPJKYIPGTKMAFGGLKKEKDRNDLITYLKKACE										
Debaryomyces kloeckeri	PAPYEKGSEKKGANLFKTRCELCHTVEQGGPHKVGPNLHGV--RTSGQAEGFSYTDANKKKGVEWTEQDLSDYLENPKKYIPGTKMAFGGLKKAKDRNDLITYLVKATK										
Abutilon	Ac	ASFQEAPPGNAKAGEKIFKTKCAQCHTVEKGAGHKQGPNLNGLFGRQSGTTPGYSYSAANKNMAVNWGENTLYDYLLNPJKYIPGTKMVFPGLJKPQDRADLIAYLKESTA									
Arum	Ac	ASFAEAPPGNPKAGEKIFKTKCAQCHTVEKGAGHKQGPNLNGLFGRQSGTTAGYSYSAANKNMAVIWEESTLYDYLLNPJKYIPGTKMVFPGLJKPQERADLIAYLKESTA									
Buckwheat	Ac	ATFSEAPPGNIKSGEKIFKTKCAQCHTVEKGAGHKQGPNLNGLFGRQSGTTAGYSYSAANKNKAVTWGEDTLYEYLLNPJKYIPGTKMVFPGLJKPQERADLIAYLKDSTQ									
Castor	Ac	ASFDEAPPGNVKAGEKIFKTKCAQCHTVEKGAGHKQGPNLNGLFGRQSGTTAGYSYSAANKNMAVQWGENTLYDYLLNPJKYIPGTKMVFPGLJKPQDRADLIAYLKQATA									
Rape, Cauliflower	Ac	ASFDEAPPGNSKAGEKIFKTKCAQCHTVDKGAGHKQGPNLNGLFGRQSGTTAGYSYSAANKNKAVEWEEKTLYDYLLNPJKYIPGTKMVFPGLJKPQDRADLIAYLKEATA									
Cotton	Ac	ASFQEAPPGNAKAGEKIFKTKCAQCHTVDKGAGHKQGPNLNGLFGRQSGTTAGYSYSAANKNMAVQWGENTLYDYLLNPJKYIPGTKMVFPGLJKPQDRADLIAYLKESTA									
Ginkgo	Ac	ATFSEAPPGDPKAGEKIFKTKCAQCHTVEKGAGHKQGPNLHGLFGRQSGTTAGYSYSTGNKNKAVNWGEQTLYEYLLNPJKYIPGTKMVFPGLJKPQERADLIAYLKQATSQE									
Maize	Ac	ASFSEAPPGNPKAGEKIFKTKCAQCHTVEKGAGHKQGPNLNGLFGRQSGTTAGYSYSAANKNKAVVWEENTLYDYLLNPJKYIPGTKMVFPGLJKPQERADLIAYLKEATA									
Mungbean	Ac	ASFDEAPPGNSKSGEKIFKTKCAQCHTVDKGAGHKQGPNLNGLFGRQSGTTAGYSYSTANKNMAVIWEEKTLYDYLLNPJKYIPGTKMVFPGLJKPQDRADLIAYLKESTA									
Niger	Ac	ASFAEAPAGDAKAGEKIFKTKCAQCHTVEKGAGHKQGPNLNGLFGRQSGTTAGYSYSAANKNKAVAWEENSLYDYLLNPJKYIPGTKMVFPGLJKPQERADLIAYLKASTA									
Pumpkin	Ac	ASFDEAPPGNSKAGEKIFKTKCAQCHTVDKGAGHKQGPNLNGLFGRQSGTTPGYSYSAANKNRAVIWEEKTLYDYLLNPJKYIPGTKMVFPGLJKPQDRADLIAYLKEATA									
Sesame	Ac	ASFNEAPPGDVKSGEKIFKTKCAQCHTVDKGAGHKQGPNLNGLFGRQSGTTPGYSYSAANKNMAVIWGENTLYDYLLNPJKYIPGTKMVFPGLJKPQERADLIAYLKEATA									
Sunflower	Ac	ASFAEAPAGDPTTGAKIFKTKCAQCHTVEKGAGHKQGPNLNGLFGRQSGTTAGYSYSAANKNMAVIWEENTLYDYLLNPJKYIPGTKMVFPGLJKPQERADLIAYLKTSTA									
Tomato	Ac	ASFDEAPPGNPKAGEKIFKTKCAQCHTVEKGAGHKQGPNLNGLFGRQSGTTAGYSYSAANKNMAVNWGENTLYDYLLNPJKYIPGTKMVFPGLJKPQERADLIAYLKEATA									
Wheat	Ac	ASFSEAPPGNPDAGAKIFKTKCAQCHTVDAGAGHKQGPNLHGLFGRQSGTTAGYSYSAANKNKAVEWEENTLYDYLLNPJKYIPGTKMVFPGLJKPQDRADLIAYLKKATSS									
Enteromorpha	Ac	STFADAPPGDPAKGAKIFKAKCAQCHTVDAGAGHKQGPNLNGAFGRTSGTAAGFSYSAANKNKTADWNENTLYDYLLNPJKYIPGTKMVFAGLKKPQDRADLIAFLKDATA									
Rhodospirillum rubrum Cytochrome c_2	EGDAAAGEK--VSKCLACHTPDQGGANKVGPNLPGVFQNTAAHKDNYAYSESYKNKGITWTEANLAAYVKNPKAFVLESKMTF-KLTKDDEIENVIAYLK--TLK										
	K					TEM			KSGDPKAK		

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.

I also wish to thank Dr. B. Boothroyd for his guidance and the facilities of his laboratory during my work at the Glaxo Company, Barnard Castle.

My thanks are due to Drs. E. W. Thompson, M. Richardson and C. J. Bailey, for much helpful advice, to Mr. A. A. Young for writing the ancestral sequence method tree construction program, and to Mr. D. L. Richardson and Mr. B. T. Meatyard for the use of their unpublished sequences in tree constructions.

I am grateful to Mr. G. Bainbridge and Mrs. E. Innes, for assistance with the extraction of plant cytochrome *c*, to Dr. M. E. Boulter, Mr. D. Caird and Mrs. L. Mayo, for assistance with the dansyl/Edman analysis, and to Miss M. Berryman for typing this thesis.

I should like to acknowledge the financial support received from the Science Research Council and the Glaxo Company.

APPENDIX 3.

The single letter code used is that given in the tentative rules of the IUPAC-IUB Commission on Biochemical Nomenclature. J represents ϵ-N-trimethyllysine.

Acidic and amide residues which have been placed by homology, are underlined.

All sequences are cited in Dayhoff (1969) except:

Bonito (Nakayama, Titani & Narita, 1971).
Carp (Gurtler & Horstmann, 1970).
Debaryomyces (Narita & Sugeno, 1968).
Castor, Sesame (Thompson, Richardson & Boulter, 1970).
Pumpkin (Thompson, Richardson & Boulter, 1971a).
Buckwheat, Cauliflower (Thompson, Richardson & Boulter, 1971b).
Abutilon, Cotton (Thompson, Notton, Richardson & Boulter, 1971c).
Rape (Richardson, Ramshaw & Boulter, 1971).
Arum, Maize (Richardson, 1971).
Tomato (Scogin, 1971).
Enteromorpha (Meatyard, 1971).

The sequence of Rhodospirillum Rubrum cytochrome c_2 has been aligned by assigning deletions or insertions, so as to obtain maximum homology with the other cytochromes c. Insertions were not considered in constructing phylogenetic trees.