The molecular evolution of plastocyanin from pteridopmrtes and gymnos perms

Green-Ogles, Pamela

THE MOLECULAR EVOLUTION OF PLASTOCYANIN

FROM PTERIDOPHYTES AND GYMNOSPERMS

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

Pamela Green-Ogles

July, 1979

Department of Botany

Frontispiece

Stereodiagrams showing the 3-dimensional structure of plastocyanin.

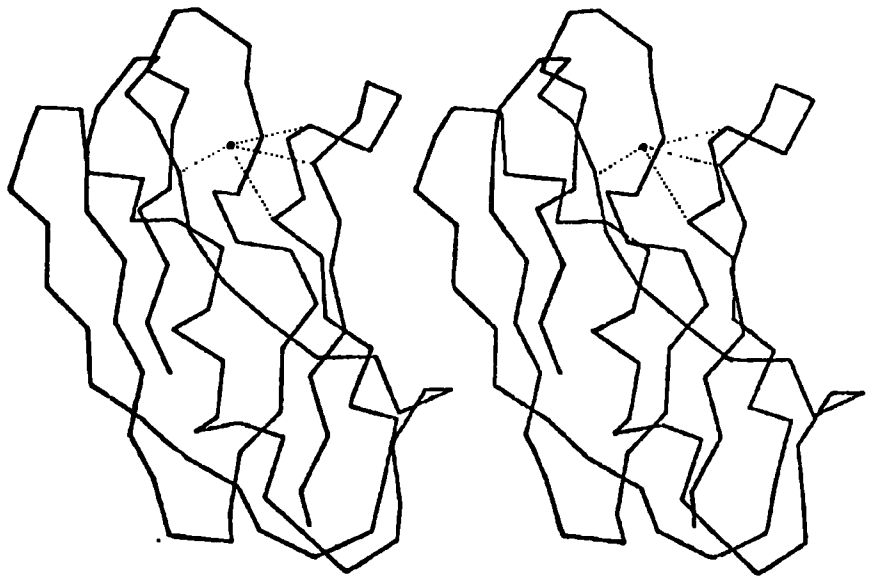
The diagrams should be viewed from the normal distance and the eyes relaxed, as if gazing to infinity. With practice, the two central images should superimpose to form a 3-D picture which can be brought into focus.

A. Configuration of peptide chain. (Dotted lines represent copper ligands).

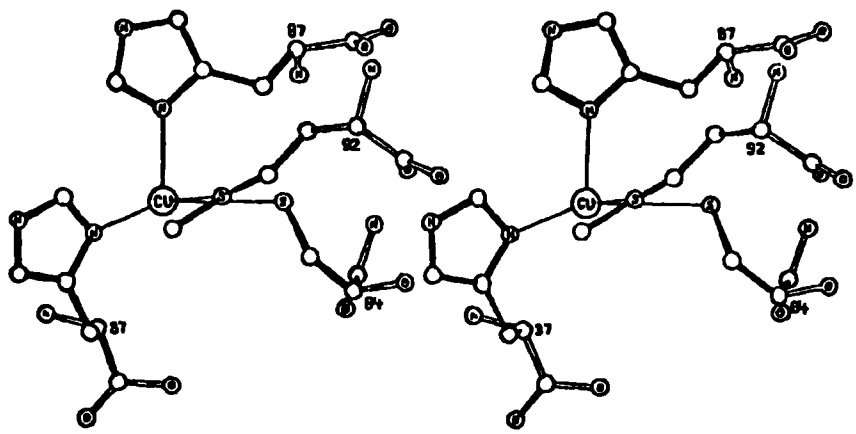
B. Arrangement of atoms around copper site.

(see Colman *et al.*, 1978)

A



B



CONTENTS

Summary 1

Acknowledgments 2

Abbreviations 3

INTRODUCTION 4

MATERIALS 17

METHODS

Extraction and Purification of Plastocyanin 21

Protein Sequence Determination

1. Manual Methods 27

2. Automatic Methods 48

Computer Analysis of Sequence Data 64

RESULTS

Extraction and Purification 66

Sequence of Plastocyanin from Pteridium aquilinium 79

N-terminal sequence determination 119

Computer analysis of N-terminal data 124

DISCUSSION

Choice of Materials 137

Extraction and Purification 141

Determination of complete sequence of plastocyanin from bracken

1. Purity of Sample 151

2. Sequence determination 152

Complete Sequence Studies 157

N-terminal sequence determination 183

Construction of phylogenetic trees 184

Time Scales 206

Validity of Methods of Molecular Phylogeny 210

BIBLIOGRAPHY 216

APPENDIX I 226

APPENDIX II 230

APPENDIX III 272

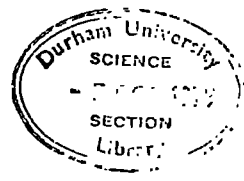
APPENDIX IV 279

SUMMARY

Plastocyanin, a small copper containing protein, was successfully extracted and purified from seven pteridophytes and three gymnosperms, the choice of species being limited by the requirement of large amounts of fresh green foliage. Many other pteridophytes and gymnosperms gave insufficient yields.

The complete sequence of plastocyanin from Pteridium aquilinum was determined by a combination of automatic and manual methods and this was compared with the primary and tertiary structures of plastocyanins from higher plants determined by other workers.

The sequences of the first forty amino acids in each of the remaining nine plastocyanins was determined by use of an automatic sequenator. These were compared using computer programmed parsimony methods and possible phylogenetic relationships of the ten species were obtained. A consideration of the validity of these results in comparison with phylogenetic relationships postulated from fossil and other evidence reached no firm conclusions.



ACKNOWLEDGMENTS

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

I am grateful to Dr. B.G. Haslett and Dr. M. Richardson; Dr. T. Gleaves and Dr. J.A.M. Ramshaw for help with the sequencing, computing and X-ray spectroscopy sections respectively and Mr. J. Gilroy and Mrs. R. McIntosh for invaluable technical assistance at all stages.

My thanks are also due to Mrs. E. Ellis for typing this thesis.

This work was carried out with the aid of a grant from the Science Research Council.

Abbreviations

The abbreviations used in this thesis are as recommended in "Instructions to Authors", Biochem. J. 131, 1-20.

INTRODUCTION

Proteins, the most versatile of macromolecules found in living organisms, are responsible for both the metabolic capabilities and the morphology of an organism. They catalyse the chemical reactions of metabolism, as specific enzymes and carrier molecules, and also catalyse the formation of structural elements (e.g. cellulose), act as structural elements themselves (e.g. cell membranes) or act as surface elements that determine the specificity of cell-cell interactions (Wood et al., 1974). Protein sequence analysis provides useful information on the structural basis of protein function, since the amino acid sequence determines its 3-D configuration and resulting functional specificity. The first protein sequence to be determined was insulin (Ryle et al., 1955), so disproving a previously held idea that proteins, like polysaccharides, had no definite molecular size or structure (Watts, 1970). Many hundreds of protein primary structures have now been determined (Dayhoff, 1972) and variations in the amino acid sequence are now known to occur when the "same" protein is extracted from different species. Comparisons of such sequences yield useful information - for example, invariant residues at certain positions along the polypeptide chains can indicate specific functions for those particular amino acids; the amino acids functioning as copper ligands in

plastocyanin had been inferred (Boulter et al., 1977) in this way before confirmation by X-ray crystallography (Colman et al., 1978). The active site of enzymes can be deduced in this manner, as can certain amino acids important for maintaining the 3-D structure of the molecule. As tertiary structures of proteins become available (Dickerson et al., 1971; Colman et al., 1978) the corresponding primary structures become even more revealing. Areas of the molecule containing predominantly hydrophobic, acidic or basic amino acids can be detected, giving insight into the characteristics and behaviour of the molecule - for example, its solubility and stability, the point of attachment to the cell membrane or even the method by which it functions as an electron transporter (Colman et al., 1978).

Another role of protein sequences dealt with in this thesis which has become increasingly important over the past decade is in the information they contain regarding the composition of that piece of DNA by which they are specified. An organism has been described as being analogous to an informostat (Zuckerkandl & Pauling, 1965a) because the information it contains is kept virtually constant - by nature of the 'base-plate' of information held in the DNA molecules. Following the demonstration of colinearity between gene structure and protein structure by Yanofsky (1967), the

potential of protein sequences was recognised. Due to degeneracy of the genetic code, much but not all of the DNA sequence could be inferred from the amino acid sequence whereas until recently it was not possible to sequence DNA directly. In addition, proteins have specialised and many varying functions, therefore a single protein, which may be the product of only one gene, can be relatively easily recognised, separated and analysed. Nucleic acids, however, have only a small range of functions which barely vary from cell to cell, and specified parts of the molecule are not yet easily obtained although this situation is changing with the use of restriction endonucleases. It has been forecast (Malcolm,1978) that as DNA sequencing techniques improve, protein sequencing will become redundant. However, this view does seem rather extreme. Although DNA sequences provide the information from which protein primary structures can be deduced, interest in the structure and consequent behaviour of proteins will continue and since more insight can be gained from a knowledge of a combination of protein primary and tertiary structures, it would seem to be a more practical approach to the problem to extract sufficient material for crystallographic and, at the same time, sequence studies rather than design a separate experiment to isolate the particular piece of DNA involved. Therefore,

depending upon the eventual use of the data obtained, protein sequencing still has many advantages.

When sequence data are obtained from several different organisms for a single protein, i.e. the same piece of DNA, they can be used to draw conclusions as to the possible evolutionary paths of the organisms concerned, and a great deal of work has now been done in this field (Dayhoff, 1972). There are two extremes of thought regarding such work as to whether a comparison of two organisms at their macromolecular level may give a similar or different result to that obtained from a comparison at a higher level of biological integration. Simpson (1961) and Cronquist (1976) for example maintain that in any attempt to construct a phylogenetic classification it is the organisms that should be classified, not their molecules. However, no taxonomic scheme has yet been proposed (or could it ever be) which takes into account every characteristic of the organism, in which case the morphological characters which are proposed as being taxonomically significant may have been determined by the interaction of several genes which differ in many respects but coincidentally show the same result at the level of the organism. Conversely, although the structural genes which code for a protein in two different individuals may be almost identical, the regulator genes which control their

reactions may be different. In such a case, the polypeptides produced from each structural gene, although almost identical in amino acid sequence, will react at different rates, due to the action of the regulator gene, hence producing different phenotypes. If the regulator gene fails to produce a metabolically functional polypeptide product, the net result will be two organisms, almost identical at the macromolecular level, but totally different at the organismal level. Even at a high level of biological integration therefore, changes in a particular characteristic of an organism may be a reflection of no more than a single mutation in a single gene amongst the many whose actions contribute to that characteristic (e.g. melanism in moths). In contrast, Zuckerkandl and Pauling (1965a) argue that no level of integration is more informative than that of the macromolecule, and if this does not show every change in the nucleotide sequence it can only mean that this level is still not close enough to the gene, maintaining that evolutionary studies are most revealing when DNA sequences are compared, so eliminating any artefacts due to gene-gene interactions, inaccuracies due to degeneracy of the genetic code and gaps in knowledge as a result of incomplete translation of the DNA sequence into a polypeptide product. In order to appreciate the significance of these arguments, some mention must be made of relevant taxonomic methods.

It is convenient to classify all living organisms. Two hundred years ago, purely phenetic classifications (i.e. systems in which the organisms are grouped to give the maximum similarity) were commonly used. Such systems can best be described as logical and at times can be extremely useful, for example the dichotomous keys in floras, but are purely artificial, being based on characters such as leaf shape and flower colour, the choice of relative importance of the characters depending on the taxonomist. Linnaeus' system falls into this category, and from it was derived the logical framework of nomenclature on which more elaborate present-day classifications are based. By the mid-nineteenth century a supposedly 'natural' (i.e. phylogenetic; based on closeness of evolutionary descent) system of classification had been devised. Ideally, in such a system the organisms should fit unambiguously together, in a way in which evolutionary relationships could be fairly easily recognised. In practice, of course, this is not the case, which is why modern approaches to classification are so exciting. Phylogenetic systems have the advantage of giving a better insight into the relationships of the organisms. Although the majority of classification systems in wide useage are now phylogenetic, advocates of phenetic systems still exist (Sokal and Sneath, 1963), indeed groups of organisms for which no fossil or other

reliable evolutionary record exist, such as the algae, are still regarded as being classified in this way (Meatyard, 1974).

The classification systems adopted in this thesis are shown in Appendix I. Although there is little dissent as to the main groupings of the plant species studied, no firm evidence is available regarding their true phylogenetic relationships and very little new fossil or biochemical evidence has recently appeared to significantly alter this situation.

The traditional methods of deriving phylogenies from fossil evidence together with morphological, embryological and ontological comparisons of existing species are well-known. Advances in the information so obtained can only come with subjective re-examinations of past fossil evidence, the discovery of new significant fossils, or application of new more revealing morphological techniques such as electron microscopy (Raven, 1977), all of which offer only relatively limited prospects of improvement. Consequently, other techniques have been examined in the search for a new, objective and more revealing taxonomic aid. Although attempts have been made to trace phylogenies and rates of evolution by cytological methods (Anderson, 1934; Levin and Wilson, 1976), by far the greatest amount

of data has been obtained from biochemical techniques. In establishing phylogenies based on biochemical evidence, three groups of compounds are recognised (Zuckerkandl and Pauling, 1965b). Semantides - DNA (primary semantides), RNA (secondary semantides) and proteins (tertiary semantides) - contain, or are obtained directly from, the base-plate of genetic information. Episemantides - such as glucose - are usually polygenic, being produced as a direct result of the metabolic activity of several tertiary semantides. Asemantides are molecules present in, but not produced by, an organism. Using the argument put forward by Zuckerkandl and Pauling, the values of these molecules as documents of evolutionary history decreases the higher the level of biological integration at which they were produced is from the gene. Thus, asemantides are of little use except in coding for their presence or absence. Episemantides can be of help, indeed Swain (1974) believes that the value of such 'secondary products' is under-rated. However, it must be recognised that the same molecule can be produced independently by two different metabolic pathways, as is lysine (Bartnicki-Garcia, 1970) - a problem of evolutionary convergence which, together with the polygenic nature of such molecules is reminiscent of the difficulties encountered when morphological characters are considered. Examples of molecules which fit into this category are fatty acids,

terpenoid compounds, carotenoids (including spore and pollen wall chemistry), flavonoids and related compounds, alkaloids and cyanogenic compounds (Swain, 1974). Semantides are regarded by many as the most valuable documents of the evolutionary history. The practicality of studying tertiary as opposed to primary semantides has already been mentioned, and in terms of evolutionary studies, it is the tertiary semantides, proteins, that have been most extensively studied. Vaughan produced good results which correspond with classically established phylogenies from serological studies on proteins from Solanum and Brassica spp. (Vaughan, 1968a) and Brassica and Sinapis spp. (Vaughan, 1968b). In the field of comparative enzymology, Yamanaka and Okunuki (1963) used differences in reaction rates in the reaction of cytochrome oxidase with cytochrome c to establish relationships between many organisms. Gel electrophoresis has been useful, not only in contributing to knowledge of protein polymorphism (Vaughan, 1968a), but in evolutionary comparisons, for example, of plant histones (Spiker, 1975). Other characteristics of proteins used (Lyddiatt, 1975) are comparisons of amino acid compositions, electrophoretic and chromatographic properties and most extensively applied, amino acid sequence studies.

In addition to comparisons of individual molecules, metabolic pathways or their control mechanisms can be utilised

to yield clues to possible evolutionary paths and similarly, comparisons of photosynthetic pigments, the photosynthetic pathway and the structure of cell walls have all been considered (Swain, 1974).

In summarising the tremendous range of biochemical taxonomic evidence, studies on asemantides, episemantides and multimolecular systems in general have been inconclusive or contradictory. Although some trends towards evolutionary advantage can be discerned in for example, the utilisation of flavonoids as antifungal agents by higher plants, or the development of advanced energy-trapping systems in photosynthesis or development towards greater flexibility, strength and impenetrability of cell walls, on the whole no pattern can be seen and the best argument which can be applied is one of increasing biochemical complexity (Swain, 1974). In contrast, studies on semantides have proved more successful, and one of these methods, protein sequence studies, has been applied to the results in this thesis.

Plastocyanin, the subject of this thesis is a blue copper protein composed of a single polypeptide chain, of approximate molecular weight 10,500, containing one atom of copper per molecule (Aitken, 1977; Ramshaw *et al.*, 1973). The protein is found within the chloroplast membranes, being involved in photosynthetic electron transfer, linking

photosystem 2 and photosystem 1 at some point near cytochrome f (Bishop, 1971; Gorman and Levine, 1966 ; Trebst, 1974). With the exception of plastocyanin extracted from Anabaena variabilis (Lightbody and Krogmann, 1967; Aitken, 1975), the protein from all other species has been found to be acidic.

A consideration of the suitability of plastocyanin for sequence and phylogenetic studies reveals that it is widely distributed throughout the plant kingdom, previously having been extracted from higher plants (Boulter et al., 1977), green algae (Katoh, 1960; Kelly and Ambler, 1974) and cyanobacteria (Aitken, 1975; Aitken, 1976). Its presence has now been detected, contrary to previous evidence, in strains of the red alga Porphyridium and Euglenoids (Aitken, 1977). To date, plastocyanin has not been detected in yellow-green algae or bacteria (Boulter et al., 1977). Comparisons with other proteins that have been used for sequence studies, such as cytochrome c (Meatyard, 1974), show that plastocyanin can be found in relatively high yields, at least in some species; also plastocyanins from higher plants have been shown to be an analogous group (Ramshaw et al., 1973), unlike the structurally diverse ferredoxins which have undergone gene duplication at several points during their evolution (Matsubara et al., 1977).

Plastocyanin does not possess a blocked N-terminus and is thus amenable to automatic and rapid sequencing techniques which have enabled an enormous amount of data to be established from a wide range of species. The copper atom is in a type 1 ligand environment within the molecule (Malkin and Malmstrom, 1970) which therefore has a high extinction coefficient. The consequent intense blue colour of the oxidised protein is advantageous during its extraction and purification and has also led to a great deal of spectrophotometrical and chemical interest in the molecule. The nature of the unusual co-ordination of the copper atom has now been elucidated (Colman et al., 1978).

Other small blue type 1 copper proteins have been isolated (azurins from bacteria, stellacyanin and umecyanin from higher plants), but none have been so intensively studied with respect to their primary structures and consequent phylogenetic relationships. There has been speculation (Ryden and Lundgren, 1976; Colman et al., 1978) as to the possibility of a common ancestral protein for these proteins, as they all possess the unusual copper co-ordination complex. Plastocyanin and azurin are the only proteins with known comparable tertiary structures, especially around the active site and some homology between the two proteins has been demonstrated (Ambler, 1971). Stellacyanin has been

partially sequenced (Wang and Young, 1977), but no tertiary structure is yet available. However, since methionine (an essential copper ligand in plastocyanin) is totally absent from the molecule (Peisach et al., 1967) the copper binding sites in the two molecules must differ.

Primary structure information from several green plant plastocyanins has been determined using automatic and manual sequencing methods. The sequence data have been subjected to computer handling methods in an attempt to establish phylogenetic relationships as an alternative to the more classical comparative morphological approach, particularly in view of the inadequate fossil record.

MATERIALS

Biological Materials

With the exceptions listed below, all biological materials were collected from local wild habitats or the Botanical Gardens Durham.

Cycas revoluta was obtained as a complete plant from the Botanical Gardens, Edinburgh.

Abies grandis and *Picea abies* were collected from Hamsterley forest.

Taxus baccata was collected from the gardens of Ushaw College, Durham.

Ephedra spp. was collected from the Botanical Gardens, Cambridge.

Selaginella spp. was obtained locally and from the Royal Botanic Gardens, Kew.

Angiopteris palmiformis, *Dicksonia antarctica* and *Thelypteris erubescens* were obtained from the Royal Botanic Gardens, Kew.

Osmunda regalis was obtained from the Botanical Gardens, Durham; Saville Gardens, Windsor Great Park and the Royal Botanic Gardens, Kew.

Chemicals and Reagents

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

Chymotrypsin	E.C.3.4.4.5. (Three times recrystallized)
Papain	E.C.3.4.4.10. (Twice recrystallized)
Trypsin	E.C.3.4.4.4. (Twice recrystallized, salt free TPCK (L-(1-tosylamido-2-phenyl) ethyl chloromethyl ketone treated).

were obtained from the Worthington Biochemical Corporation, Freehold, New Jersey, U.S.A.

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

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

Thermolysin	(crystalline)

was obtained through Dawa Karkei K.K. Osaka, Japan.

Amberlite M.B.1

was obtained from Rohn and Haas Co., Philadelphia, U.S.A.

DE-23 Sephadex
Sephadex G 75
DEAE Sephadex
Sephadex G 50 fine
Blue Dextran 2000

were obtained from Pharmacia Ltd., Uppsala, Sweden.

Biogel P-10

was obtained from Biokad Laboratories Ltd., London.

Ninhydrin (Indantrione hydrate) Puriss grade, 9-10 - Phenanthraquinone

was obtained through Koch-Light Laboratories Limited, Colnbrook, Bucks., England.

Pyridine

was obtained through Rathburn Chemicals, Walkerburn, Peebleshire.

Arginylarginine

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

Triethylamine

was obtained from Pierce Chemical Co., Rockford, Illinois, U.S.A.

Anhydrous n-heptafluorobutyric acid
Qyadrol-TFA buffer
Ethyl acetate
1-Chlorobutane
Benzene
5% (v/v) Phenylisothiocyanate in heptane

were obtained as sequencer grade reagents from Pierce Chemical Co., Rockford, Illinois, U.S.A.

All chemicals were used as supplied except for phenylisothiocyanate, which was vacuum distilled once before use.

Other Materials

Polyamide sheets were obtained through BDH from the Chen Chin Trading Co. Limited, Taipei, Taiwan.

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

20 cm x 20 cm Kieselgel 60T254 (D.C. Alufolien) TLC plates were obtained from Merck, Darmstadt, Germany.

Preparation of Solutions

The compositions of solutions are as given individually in Methods or otherwise as described in Lyddiatt (1975).

METHODS

Extraction and Purification of Plastocyanin

1. Plastocyanin Assay

a) Qualitatively

The presence of plastocyanin in solution was detected by the appearance of a blue colour on the addition of a trace of potassium ferricyanide.

b) Quantitatively

The amount of oxidised plastocyanin present in a solution was estimated with a Perkin Elmer 402 recording spectrophotometer using silica cells of a 1 cm light path. From the absorption at 597 nm, the protein concentration was calculated assuming an extinction coefficient of 4.5×10^3 and a molecular weight of 10,500.

2. Criteria of Purity

The purity of plastocyanin was estimated spectrophotometrically, by determining the absorbance of the reduced and oxidised forms at 278 and 597 nm respectively. Fractions with A_{278}/A_{597} nm less than 3 were pooled for use. The reduced form of plastocyanin was obtained by addition of a trace of ascorbic acid, and the oxidised form by addition of a trace of potassium ferricyanide (see Fig. 10).

3. Extraction of Crude Plastocyanin

Washed plant material was homogenised in 1 kg batches for 3 min in a 1 gallon Waring blender with 3 L of one of the following buffers:-

30, 40 or 50% Acetone buffer (at -20°C)

1,000, 1,200 or 1,500 ml acetone

200 ml 500 mM Tris-HCl buffer (pH 7.2)

2 g potassium metabisulphite

distilled water to 3,000 ml

Aqueous buffer containing detergent (at 4°C)

2.5 g sodium metabisulphite

10 g Tween 80

500 ml 500 mM Tris-HCl buffer (pH 7.2)

distilled water to 5,000 ml

If the material was fresh, 2 kg of ice were substituted for 2 kg of distilled water, to keep the temperature of the blend below 4°C. If the plant material was tough, the blend was left to stand at 4°C for 20-30 min and re-blended. Waste plant material was removed from the pulp by squeezing through muslin, and the resulting liquid centrifuged at 2,500 r.p.m. in an MSE 4L centrifuge for 1 h at 0°C, to remove cell debris.

After centrifuging, acetone at -20°C was added to the supernatant, to give a final concentration of 80% (v/v) acetone. The suspension was allowed to stand in the cold for 10-60 min. so that the precipitate settled sufficiently to allow most of the liquid to be sucked off. The loose sediment was centrifuged at 2,500 r.p.m. in an MSE 4L

centrifuge for 3 min. at 0°C. The precipitate was redissolved in a 1% solution of Tween in 20 mM Tris-HCl buffer (pH 7.2). Insoluble material was removed by centrifuging at 2,500 r.p.m. in an MSE centrifuge for 10 min. at 0°C.

4. Ion-exchange Chromatography on DE 23 Sephadex resin

a) Preparation and Regeneration

New and used resin was treated in the following manner:-

i) Precycling

The resin was stirred in approximately 15 vol. 0.5 N hydrochloric acid, and allowed to settle for at least 30 min. The slurry was then filtered and washed with distilled water in a Buchner funnel containing Whatman No. 54 filter paper, attached to a water pump until the effluent had reached pH 4. This procedure was repeated, using 0.5 N sodium hydroxide, until the effluent had reached pH 8.

ii) Equilibration

The resin was stirred into approximately 15 vol 500 mM Tris-HCl buffer (pH 7.2), and titrated with conc. hydrochloric acid until the pH remained constant at pH 7.2. After settling, the supernatant was decanted off, and the resin stored in 300 mM Tris-HCl buffer (pH 7.2) at 4°C.

When required, the slurry was poured into a column of convenient size and washed with 20 mM Tris-HCl buffer (pH 7.2).

b) Adsorption

The crude protein solution was applied to the resin in one of two ways, either on columns or batchwise, depending on the condition of the solution.

i) Column method

The filtrate was passed through a column of DE 23 Sephadex at room temperature at a flow rate of 500-1500 ml/hr. and the charged column washed with 150-500 ml. 20 mM Tris-HCl buffer (pH 7.2). The size of the column varied from 4.5 cm x 35 cm to 6.5 cm x 40 cm. depending on the quantity and quality of the filtrate to be applied.

ii) Batch Method

1 vol of DE 23 Sephadex resin was stirred for 1 h with 2 vol of filtrate and then allowed to settle for 30 min. before the supernatant was decanted off. The resin was then washed exhaustively with 20 mM Tris-HCl buffer (pH 7.2) in a Buchner funnel containing Whatman No. 54 filter paper attached to a filter pump, and finally poured into a column for elution, as in the column method.

c) Elution

The charged and washed resin was eluted with 50 mM Tris-HCl buffer (pH 7.2) containing 250 mM NaCl, at room temperature. The eluate was collected in 100 ml fractions, and all fractions containing plastocyanin were pooled.

5. Storage

At this point in the extraction the crude extract was stored until enough plastocyanin had been collected (normally approximately 40 mg) before continuing with the purification.

6. Purification

a) Ion-exchange chromatography on DE 23 Sephadex

All the crude plastocyanin extracts were bulked and dialysed overnight against 20 mM Tris-HCl buffer (pH 7.2). The solution (1-5 L) was then applied to a further DE 23 Sephadex column 3.5 cm x 15 cm at room temperature. The charged column was washed with 200 ml. 20 mM Tris-HCl buffer (pH 7.2) followed by 200-400 ml 50 mM Tris-HCl buffer (pH 7.2) containing 100 mM NaCl. Elution was effected with 50 mM Tris-HCl buffer (pH 7.2) containing 200 mM NaCl. All fractions containing plastocyanin were pooled.

b) Concentration

Following dialysis against distilled water, the sample was concentrated in an Amicon Diaflo model 402 containing a

PM 10 membrane followed by an Amicon Diaflo model 52 containing a UM 2 membrane, until the sample size was less than 5 ml.

c) Gel filtration on Sephadex G 75

The gel filtration step was carried out at 4°C using Sephadex G 75, equilibrated in 50 mM Tris-HCl buffer (pH 7.2) containing 50 mM KCl. A maximum of 5 ml of plastocyanin solution was applied to a 3 cm x 100 cm column which was developed in an ascending direction with a flow rate of 30 ml/hr.

d) Ion-exchange chromatography on DEAE-Sephadex

After repeating step 7, final purification was performed at 4°C by gradient elution of reduced plastocyanin from DEAE-Sephadex in a 1.5 x 12 cm. column using 50 mM Tris-HCl buffers (pH 7.2) containing 50-250 mM KCl, with a hydrostatic head of 20 cm.

e) De-salting on Amberlite Monobed MB-1 resin

Pure plastocyanin samples were dialysed overnight at 4°C against distilled water and concentrated, as in step 7, to a final concentration of 2 mg/ml. Up to 5 ml of solution was then applied to a 1 cm x 15 cm column of Amberlite Monobed MB-1 resin in distilled water with a flow rate of 20 ml/min. Finally, the column was washed with distilled water.

7. Storage

The pure plastocyanin was lyophilized and stored at -20°C.

Protein Sequence Determination

1. Manual Methods

A. Reduction and Carboxymethylation (Crestfield et al., 1963)

20 mg of plastocyanin was dissolved in 3 ml 6 M guanidine 0.6 M Tris-HCl pH 8.6, and the solution flushed with nitrogen. Reduction was effected by addition of 30 µl of 2-mercaptoethanol, and the mixture allowed to stand under nitrogen for 3 hours at room temperature. S-carboxymethylation was brought about by addition of freshly prepared iodoacetic acid (0.3 ml of 0.268 g/ml) in 0.1 M NaOH, and the solution kept in the dark for 15 min. The protein was separated from the reaction products by dialysis for 24 h against distilled water, and lyophilised.

B. Cyanogen bromide cleavage (Kasper 1970)

20 mg of reduced and S-carboxymethylated protein were dissolved in 3 ml 70% (v/v) formic acid. Sufficient cyanogen bromide was added to provide a 100-fold molar excess of the reagent, with respect to the methionine content of the protein for 24 h at room temperature in the dark. Excess reagent and volatile by-products were removed by lyophilisation. The resultant peptides and non-volatile by-products were separated by passage through a 1 cm x 200 cm column of

Sephadex G 50 fine or Bio-Gel P 10 equilibrated and eluted with 70% (v/v) formic acid, at a flow rate of 20 ml/h. 2 ml fractions were collected in an LKB Ultrarack fraction collector, fitted with a Uvicord III absorbance monitor, which recorded the absorbance of the samples at 280 nm and 206 nm. The peptide peaks were pooled and lyophilised.

C. Proteolytic digestion of cyanogen bromide peptides

5-15 mg of peptide were dissolved in 0.25-1.0 ml of appropriate buffer (see Table 1). The enzyme, dissolved in the same buffer was added to give 2% (w/w) enzyme to peptide. Incubation took place at 37°C for 1-3 h and digestion was halted by the addition of excess glacial acetic acid, followed by lyophilization. To test whether digestion was complete, an aliquot of the resulting mixture of peptides (0.1% by volume) was removed and the N-terminal amino acids identified by dansylation by the method of Gray and Hartley (1963a)

D. Peptide Purification

I. Electrophoretic separation

Separation of peptides was achieved by high voltage paper electrophoresis, in a flat plate apparatus (107 cm x 15 cm, the Locarte Co., London) on Whatman 3 MM Paper (15 cm).

Table 1 Digestion with Proteolytic Enzymes

Enzyme	Substrate and Enzyme Buffer	Digestion Time	Order of Methods of Peptide separation
Chymotrypsin	200 mM N-ethylmorpholine/glacial acetic acid pH 8.6	10 min repeated for 1 h	1. electrophoresis, pH 6.5 2. electrophoresis, pH 1.9 3. BAWP
Trypsin	"	1 h	1. electrophoresis, pH 6.5 2. electrophoresis, pH 1.9 3. BAWP
Thermolysin	" + 1% (v/v) 5 mM $CaCl_2$	2 h	1. electrophoresis, pH 1.9 2. electrophoresis, pH 6.5 3. BAWP
Papain	pyridine/acetic acid/water (25/1/225 by vol) pH 6.5 0.1% (v/v) mercaptopropanol	1 h repeated for 1½ h	1. electrophoresis, pH 6.5 2. electrophoresis, pH 1.9 3. BAWP

Buffers at pH 6.5 or pH 1.9 were used, and a voltage of 9 Kv was applied at 7 p.s.i. pressure (see Table 2).

Electrophoresis Standard solution

0.1 M arginyl arginine in 1 M $NaHCO_3$ was added to 0.2 M dansyl chloride in acetone. After incubating for 1 h at 37°C, the solution was diluted 1,000 fold and dansyl arginine in ethanol was added to give a final concentration of 0.1 M.

II Paper Chromatography

Separation of peptides was achieved by using freshly prepared BAWP solvent as developer on Whatman 3 MM chromatography paper (55 cm x 46 cm). Development of the chromatograms took 18 hours at room temperature and the paper was dried for 4 hours at 45°C.

BAWP Solvent

Butan-1-ol	150 ml
Acetic acid (glacial)	30 ml
Pyridine	100 ml
Water	120 ml

Chromatography Marker Solution

0.1 mM dansyl arginine in 95% (v/v) ethanol.

III Location of Peptides (Easley, 1965)

a) Paper Strips

Paper strips, 1" wide were cut from the edge of

Table 2 Separation of Peptides by Electrophoresis

Buffer	Vol (ml)	pH	Current (at 9 Kv and 7 p.s.i.)	Time
pyridine	250			
glacial acetic acid	10	6.5	30 - 50 mA	90 - 120 min
water	2250			
glacial acetic acid	200			
formic acid	50	1.9	50 - 60 mA	60 - 90 min
water	2250			

the electrophoretograms or chromatograms. This removed approximately 10% of the peptides.

The strips were dipped in freshly prepared ninhydrin reagent and allowed to dry, first at room temperature, then at 60-80°C for 5-10 min. Paper chromatograms in which the BAWP solvent was used were not heated, due to the development of a high background colour, which made detection of ninhydrin positive regions difficult. These strips were air dried at room temperature, and the positive ninhydrin colour developed over a number of hours.

If a tryptophan-containing peptide was suspected, the relevant strip was dipped in freshly prepared Ehrlich reagent after completion of the ninhydrin staining procedure. The strip was air dried at room temperature for 30-60 min. In this test, the pink ninhydrin spots should turn colourless, and a purple colouration indicate a tryptophan containing peptide.

Cadmium-ninhydrin reagent (Heilmann et al., 1957)

Cadmium acetate	100 mg
Water	10 ml
Glacial acetic acid	5 ml
Acetone	100 ml
Ninhydrin	1% (w/v)

Ehrlich reagent

2% 4-dimethylaminobenzaldehyde in 20% (v/v) HCl.

b) Whole Electrophoretogram or Chromatogram

The method used was a modification of that of Mendez & Lai, 1975.

The whole electrophoretogram or chromatogram was washed in acetone, with a final wash in acetone containing 1% triethylamine to adjust the pH, and air dried at room temperature for 5 mins. The paper was wet (by either dipping or spraying) with a solution of 0.001% fluorescamine in acetone, air dried for 5 mins. at room temperature, washed in acetone and again air dried at room temperature.

In this test, the peptides are identifiable as fluorescent bands when viewed at 336 nm (Udenfriend et al., 1972, Bohlen et al., 1974, Vandekerckhove and Montagu, 1974).

IV Peptide Elutions

The electrophoretograms and chromatograms were cut into the relevant strips, and the peptides eluted for 4 hours into "Pyrex" screwcap tubes (1 cm x 5 cm) using 20% pyridine. The eluted samples were lyophilised and stored at -20°C.

V Peptide Mobilities

a) Electrophoresis

At pH 6.5, peptide mobilities were measured from

a true neutral point (determined as 4/11 of the distance between the standard dansyl-arginylarginine and the 1-dimethylaminonaphthalene-S-sulphonic acid) and calculated relative to the distance from the true neutral point to the dansyl-arginylarginine.

At pH 1.9, the mobilities were measured from the 1-dimethylaminonaphthalene-S-sulphonic acid and expressed relative to the distance from the neutral point to the dansyl-arginine.

b) Chromatography

The mobilities of the peptides were measured from the origin and expressed relative to the distance moved by the dansyl-arginine standard.

E) Quantitative Amino Acid Composition of Proteins and Peptides

Amino acid compositions were determined using a Locarte amino acid analyser.

Approximately 0.05 μm of protein or peptide were hydrolysed with 0.5 ml constant boiling 5.7 m HCl in evacuated Pyrex tubes (Moore and Stein, 1963). Protein samples were hydrolysed for 24, 48 and 72 hours to obtain zero time values for serine and threonine, and maximum values for valine, isoleucine and leucine. Peptide samples were hydrolysed for 24 h. No special precautions were made to ensue the recovery of tryptophan (Ramachandran and

Witkop, 1967 ; Matsurbara and Sasaki, 1969; Liu and Cheng, 1971).

Following hydrolysis, samples were dried in vacuo over NaOH and stored at -20°C until analysed.

F. Semi-quantitative amino acid Composition of Peptides

An aliquot of peptide (10 n mol) was dried in vacuo over NaOH in a Durham tube (6 mm x 30 mm. A. Gallenkamp Limited, London). After addition of 50 µl of constant boiling 5.7 M HCl the tube was sealed and heated at 105°C for 18 h. If tryptophan was suspected, the method was repeated using 50 µl of constant boiling 5.7 M HCl containing 4% thioglycollic acid, and heated at 105°C for one hour.

The acid was removed in vacuo over NaOH and the free peptide amino acids labelled by the dansyl method of Gray and Hartley (1963a), omitting the final hydrolysis. The dansyl derivatives of the amino acids were identified by chromatography on polyamide sheets. (See Methods G IV)

G. Peptide Sequencing

Peptides were sequenced using the N-terminal dansyl-Edman procedure of Gray and Hartley (1963b) 20-500 nmol of peptide were degraded sequentially and 5-20% removed at the end of each cycle, to identify the N-terminal amino acid by the dansyl method.

Any contaminating free amino acids and the free acid after the final Edman degradation were identified using the dansyl method without acid hydrolysis.

C-terminal amino acid sequences were determined using digestions with carboxypeptidase A. The liberated amino acids were identified as their dansyl derivatives, without prior acid hydrolysis.

I. Edman Degradation Procedure

This procedure was based on the methods of Edman (1956) and Blomback et al., (1966).

The peptide was dissolved in up to 150 µl of 20% (v/v) pyridine. 150 µl of 5% (v/v) redistilled phenylisothiocyanate (PITC) in pyridine was added and the solutions mixed. The sample was immediately flushed with oxygen-free nitrogen (Isle and Edman , 1963) and capped quickly. The samples were reacted at 45°C for 1 h and the excess PITC and volatile by-products were removed by drying in vacuo over NaOH and P_2O_5 at 60°C. When the sample was completely dry, the tube was flushed with nitrogen (Percy and Buchwald, 1972) and 200 µl of anhydrous trifluoroacetic acid (TFA) was added (Elmore and Toselan, 1956). The tube was sealed with "parafilm" and incubated 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 200µl of distilled water

and extracted twice with 1.5 ml of butyl acetate (Gray, 1967) and then dried in vacuo over concentrated H_2SO_4 and NaOH.

II. Determination of C terminal amino acid sequences

10 µl of a carboxypeptidase A-disopropyl phosphorofluoridate (CBA-DFP) suspension were washed three times with distilled water. The enzyme was suspended in 150 µl 0.2 M $NaHCO_3$ at 0-2°C and dissolved using 100-150 µl 0.1 M NaOH. The solution was neutralised with 100-150 µl of 0.1 M HCl, and made up to 1.5 ml. with 0.2 M N-ethylmorpholine acetic acid buffer, pH 8.5.

Samples of the protein or peptide to be digested were dried in a Durham tube in vacuo over NaOH. 20 µl of the CBA solution were added and the tube sealed with parafilm. The incubations took place for varying times, 15 secs. to 24 hours, at 37°C. The reaction was terminated by addition of excess glacial acetic acid followed by drying in vacuo over NaOH.

III. Dansyl Method for identification of N-terminal amino acids

This procedure was based on the method of Gray and Hartley (1963b).

5-50 nmol of peptide were dried in vacuo over NaOH in a Durham tube. 10 µl of 0.2 M $NaHCO_3$ were added and the sample

again dried in vacuo over NaOH. Equal volumes of a solution containing dansyl chloride (5 mg/ml) in acetone and distilled water were mixed and 10 µl added to the dried sample. The tube was immediately sealed with Parafilm and incubated at 45°C for 45 min. The reaction was terminated by drying in vacuo over NaOH.

50 µl of constant boiling 5.7 M HCl was added, the tube sealed and then heated for 5-18 hours at 105°C. The sample was finally dried in vacuo over NaOH.

IV Chromatography of dansyl derivatives

Dansyl derivatives were identified by chromatography on polyamide sheets (Woods and Wang, 1967). The sample was dissolved in 5-10 µl of 50% (v/v) pyridine and spotted onto double sided polyamide sheets. Samples were applied to both sides of the sheet, on a common origin, in a 4:1 ratio, and dried under a hot air draught. 1 µl of a standard marker solution was applied to the origin to co-chromatograph with 20% of the sample.

The chromatograms were developed by running in solvent A (Woods and Wang, 1967) for 45-60 min, drying and then running at right angles in solvent B for 45-60 min. After drying the sheets were examined under u.v. light (350 nm) and the results recorded (Figure 1). The chromatograms were then developed in solvent C, (Ramshaw et al., 1970) in the same

direction as solvent B, for 45-60 min and after drying were again examined at 350 nm (Figure 2).

The polyamide sheets after washing were used up to 60 times

Polyamide Sheet Chromatography solvents

Solvent A: formic acid, 1.5% (v/v)

Solvent B: toluene: acetone, 9:1 (v/v)

Solvent C: butyl acetate: methanol, acetic acid, 30:20:1 by vol.

Wash: acetone: 1 M ammonia, 1:1 (v/v)

Standard chromatography marker solution

1 µg/ml of dansyl derivatives of arginine, glutamic acid, glycine, isoleucine, lysine, phenylalanine, proline and serine in 95% (v/v) ethanol.

V. Determination of amide residues

Amide residues were determined, where possible, from peptide mobilities at pH 6.5 using the method of Offord (1966). (see Figures 3 and 4).

FIGURE 1.

Chromatography of dansyl-amino acids on polyamide thin layers.

Development was by solvent A (1.5% (v/v) formic acid) in the first dimension, and solvent B (Toluene: Acetic acid, 9:1, v/v) in the second dimension.

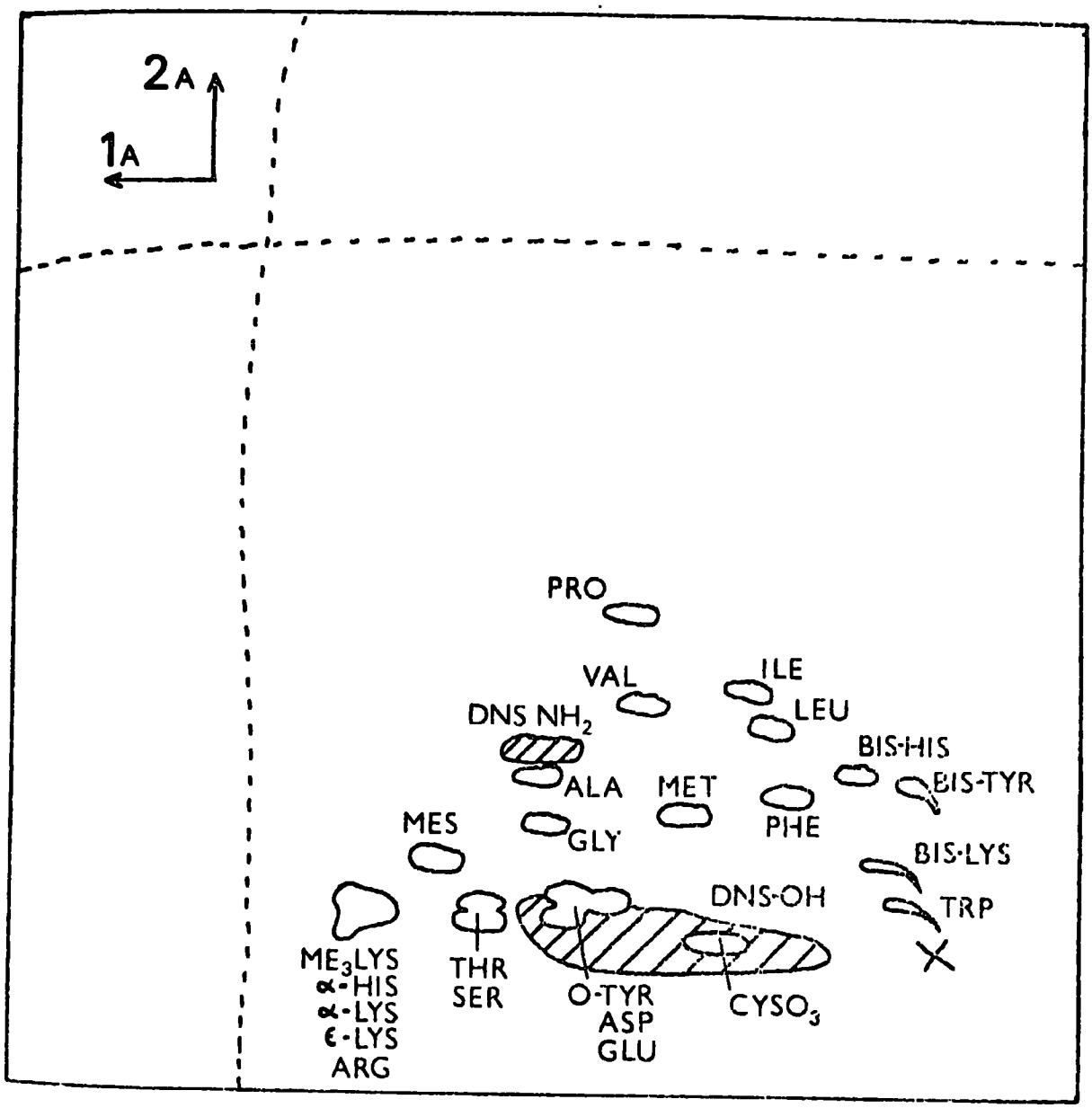

2A
1A
PRO
VAL
ILE
LEU
DNS NH2
ALA
MET
BIS-HIS
BIS-TYR
PHE
MES
GLY
BIS-LYS
DNS-OH
TRP
ME3LYS
α-HIS
α-LYS
ε-LYS
ARG
THR
SER
O-TYR
ASP
GLU
CYSO3

FIGURE 2.

Chromatography of dansyl-amino acids on polyamide thin layers.

Development was by solvent A (1.5% (v/v) formic acid) in the first dimension, and solvent B (Toluene: Acetic acid, 9:1, v/v) followed by solvent C (Butyl acetate: methanol: acetic acid, 30:20:1, by vol.) in the second dimension.

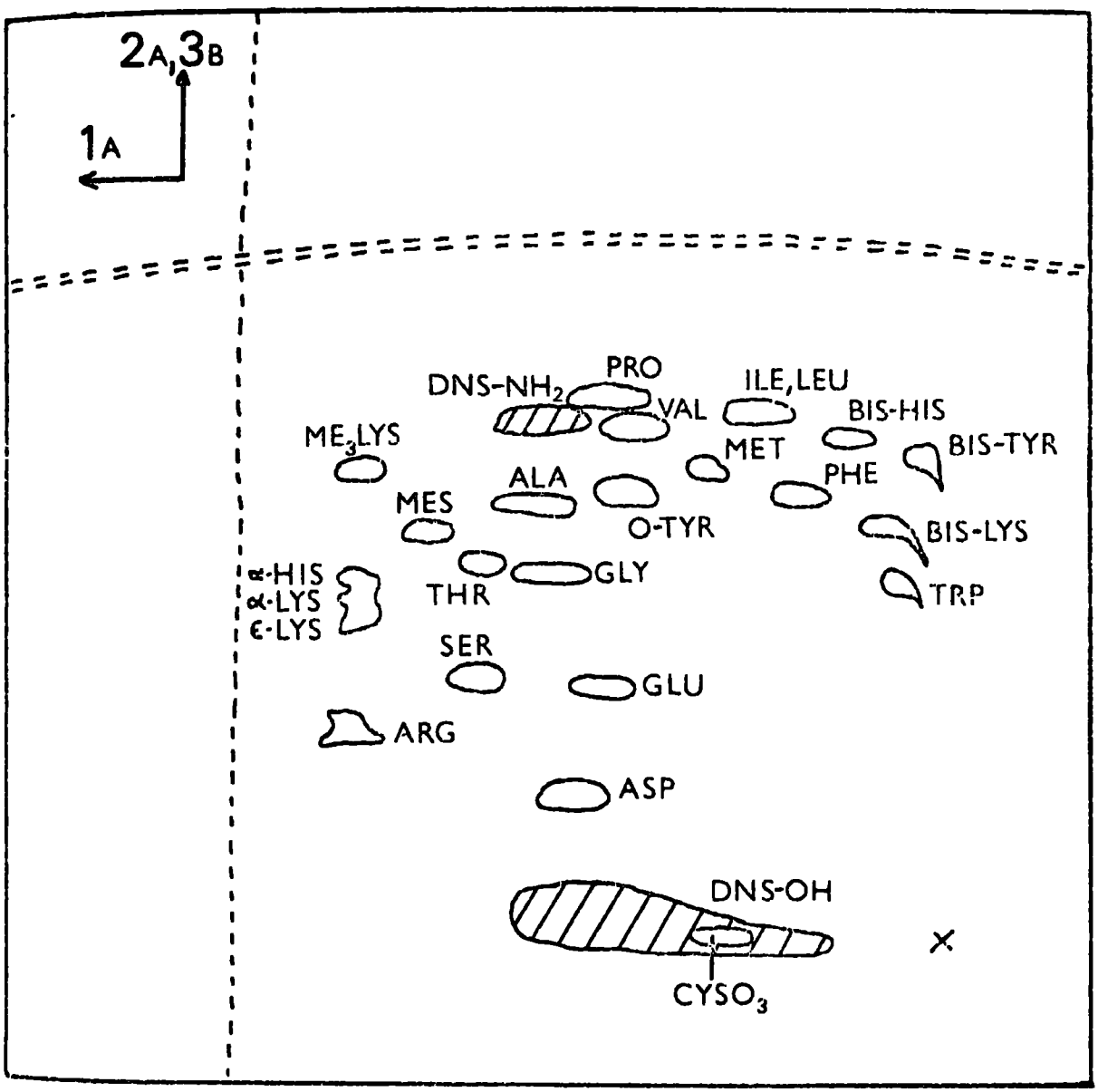

2A,3B
1A
DNS-NH2
PRO
VAL
ILE,LEU
BIS-HIS
ME3LYS
MET
BIS-TYR
ALA
PHE
MES
O-TYR
BIS-LYS
α-HIS
α-LYS
ε-LYS
THR
GLY
TRP
SER
GLU
ARG
ASP
DNS-OH
CYSO3

FIGURE 3.

The mobility of peptides on pH 6.5 electrophoresis

The electrophoretic mobility of peptides relative to dansyl-arginylarginine at pH 6.5 is plotted against their molecular weight for charges (E) of ±1 to ±3 at pH 6.5. Peptides containing histidine or cysteic acid do not conform directly to this diagram.

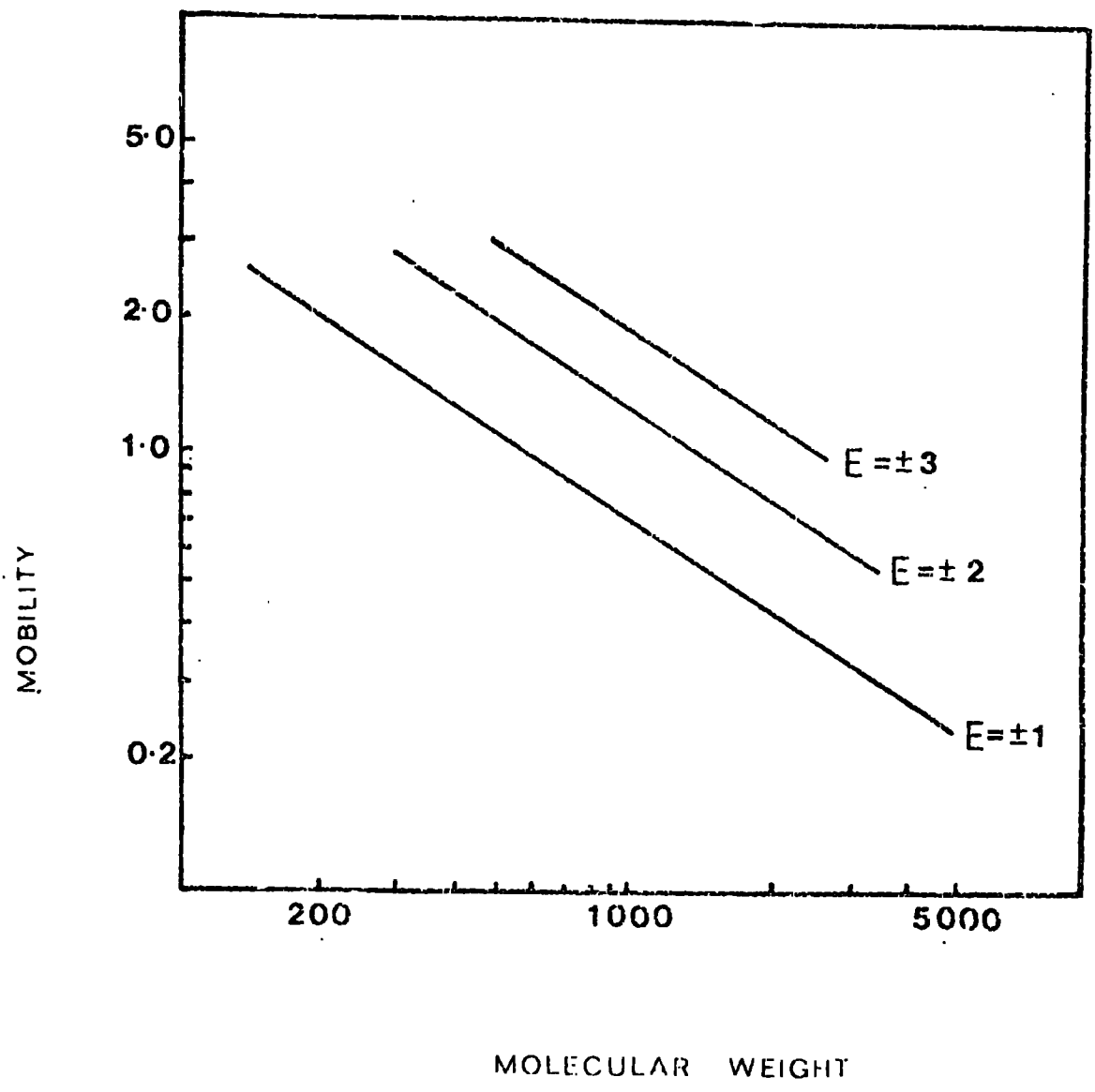
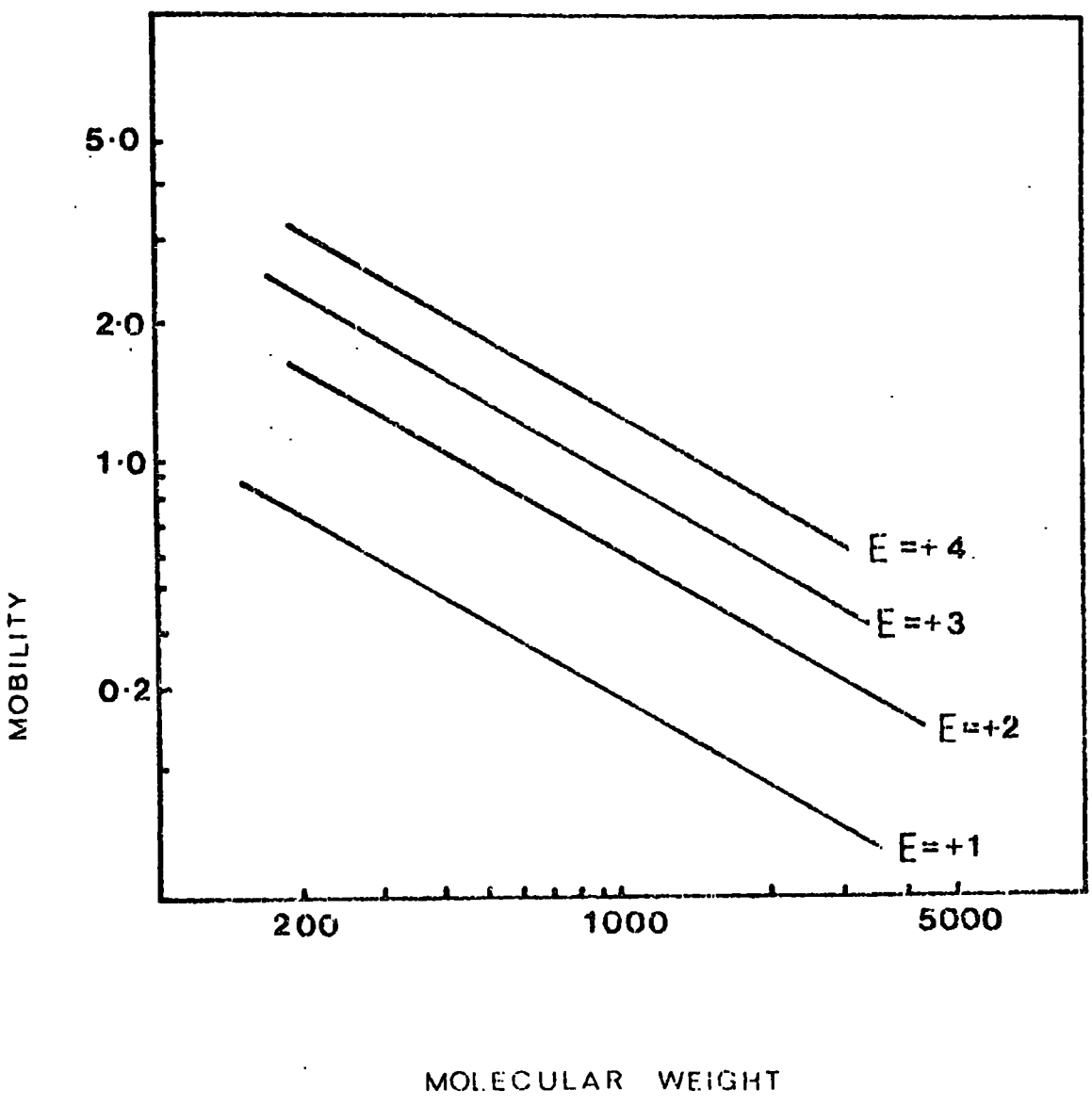

5·0
2·0
1·0
0·2
MOBILITY
E =±3
E=± 2
E=±1
200
1000
5000
MOLECULAR WEIGHT

FIGURE 4.

The mobility of peptides on pH 1.9 electrophoresis

The electrophoretic mobility of peptides relative to dansyl-arginine at pH 1.9 is plotted against their molecular weight for charges (E) of ±1 to ±4 at pH 1.9.


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

2. Automatic Sequencing Methods

a) Beckman 890C sequencer - for analysis of N-terminal amino acids of proteins

Edman degradation was carried out with a Beckman 890C automatic sequencer. 5 mg of protein (450 nmol) was applied to the sequencer in 0.4 ml of 70% (v/v) formic acid. Before degradation, a cycle was carried out in the absence of phenylisothiocyanate in order to stabilise the protein film. A single degradation was performed with the 'fast protein' programme as recommended in the Beckman (1972) operation manual (Beckman programme no. 072172C). The amino acid released at each degradation step was collected, as the phenylthiazolinone derivative, in a refrigerated fraction collector.

b) Solid Phase Peptide Sequencing

Some tryptic peptides were sequenced using an Anachem APS 2400 solid phase sequencer.
The peptides were attached to aminopolystyrene resin by C-terminal lysine using the diisothiocyanate method, or by C-terminal homoserine using the homoserine lactone method (Laursen, 1975).

c) Conversion of Sequencer Residues

50 n moles of PTH norleucine were added to each of the samples which were dried under a stream of nitrogen and converted to the more stable phenylthiohydantoin (PTH) derivative by reacting the dry fraction with 0.3 ml 1.0 M HCl for 10 minutes at 80°C under nitrogen.

The PTH derivatives were then extracted twice using 0.7 ml portions of ethyl acetate. Both the organic phase and the aqueous phase formed during this extraction were dried down separately and retained.

d) Identification of PTH derivatives (see Table 3)

Identification of derivatives was achieved by a logical combination of the following three methods (Haslett and Boulter, 1976).

i) Thin layer chromatography (Figures 5 and 6)

All PTH amino acids in the organic phase were identifiable by thin layer chromatography on pre-coated silica-gel sheets impregnated with a fluorescent indicator. Derivatives were seen under u.v. light of wavelength 260 nm as fluorescence-quenched spots and identified by their position relative to those of standard mixtures.

Samples were applied to the sheets 1 cm apart, on a base line 1" from the bottom. Approximately 25 nmole

of each PTH derivative was spotted and dried under a hot air draught. 5 µl each of standard A and standard B were also applied at suitable intervals along the base line.

Chromatograms were developed with the solvent systems of Jeppsson and Sjöquist (1967). The sheets were run in solvent 1 for 45-60 min and dried in a hot air draught. After recording the results, they were re-chromatographed in solvent 2 for 45-60 minutes, dried, and any newly resolved derivatives were recorded.

Solvent 1	Heptane	58 ml
	propionic acid	17 ml
	dichloroethane	25 ml
Solvent 2	Heptane	50 ml
	n-butanol	30 ml
	75% formic acid	9 ml

Standard A. 5 n mole/ml of PTH derivatives of glutamic acid, glycine, phenylalanine and leucine in ethyl acetate.

Standard B. 5 n mole/ml of PTH derivatives of asparagine lysine, alanine and valine in ethyl acetate.

ii) Gas Chromatography (Figures 7, 8 and 9)

Most PTH amino acid derivatives in the organic phase were identified by this method, using a Varian 1400 gas chromatograph, and the methods described by Pisano et al., (1972) and Beckman (1973). The column contained

Chromasorb W AW-DCMS (100-120 mesh) coated with SP400. Silylation of samples was performed by the 'on-column' method (Beckman, 1972) with N, O-bis (trimethyl-silyl) acetamide as the silylating agent.

iii) Regeneration by hydriodic acid (Inglis et al., 1971)

All PTH derivatives could be identified by regeneration to the parent amino acid. Samples were transferred to Durham tubes using ethyl acetate and dried in vacuo over NaOH. 50 µl of constant boiling HI were added to each of the dried samples, the tubes sealed and then heated for 6 hours at 140°C. The hydrolysate was dried in vacuo over NaOH and the resultant amino acids identified by the dansyl method and chromatography on polyamide sheets (see Methods G IV)

Calculation of Yields

Yields of PTH derivatives were estimated from gas chromatography by reference to a height factor (calculated from the height of the PTH amino peaks relative to the height of the PTH-norleucine peak in an initial standard run which contained equal n molar concentrations of the PTH amino acids).

For each identification, the number of n moles contained in the sample was then estimated by the equation:-

$$\frac{\text{height of X peak}}{\text{Height of nor-leu peak}} \times \text{height factor of X} \times 50$$

Table 3

Identification of phenylthiohydantoin derivatives of amino acids from the automatic sequencer

The amino acid removed during each Edman degradation step in the automatic sequencer was released as a thiazolinone derivative and identified manually, initially by conversion to the more stable phenylthiohydantoin derivative. Most derivatives were found in the organic phase resulting from the conversion procedure but Arg, His and Cysteic acid were found in the aqueous phase.

Identification of derivatives was achieved by a combination of methods as shown in the table, and application of the criteria proposed by Haslett & Boulter (1976).

Identification of PTH derivatives from Beckman Sequencer

Amino Acid	Found in organic phase	Found in aqueous phase	Identified on Thin Layer Chromatography	Identified by gas chromatography: Programme A 160° - 270°C, 8°/min	Programme B, sample silylated 200° - 270°C, 6°/min	Programme C, sample silylated isothermal	Identified by regeneration
Alanine	✓		✓	✓			✓
Glycine	✓		✓	✓			✓
Valine	✓		✓	✓			✓
Proline	✓		✓	✓			✓
Phenylalanine	✓		✓	✓			✓
Methionine	✓		✓	✓			✓
Leucine	✓		✓	✓		✓ single peak	✓
Isoleucine	✓		✓			✓ double peak	✓
Glutamic acid	✓		✓		✓		✓
Aspartic acid	✓		✓		✓		✓
Asparagine	✓		✓				converted to aspartic acid
Glutamine	✓		✓				converted to glutamic acid
Lysine	✓		✓				converted to bis-histidine
Serine	✓		✓				converted to alanine
Threonine	✓		✓				converted to α-amino butyrate
Tryptophan	✓		✓				✓
Tyrosine	✓		✓				✓
Arginine		✓					✓
Histidine		✓					✓
Cysteic acid		✓					✓

FIGURE 5.

Thin layer chromatography of phenylthiohydantoin derivatives of amino acids

Development was by solvent 1 (heptane: propionic acid: dichloroethane; 58:17:25, by vol)

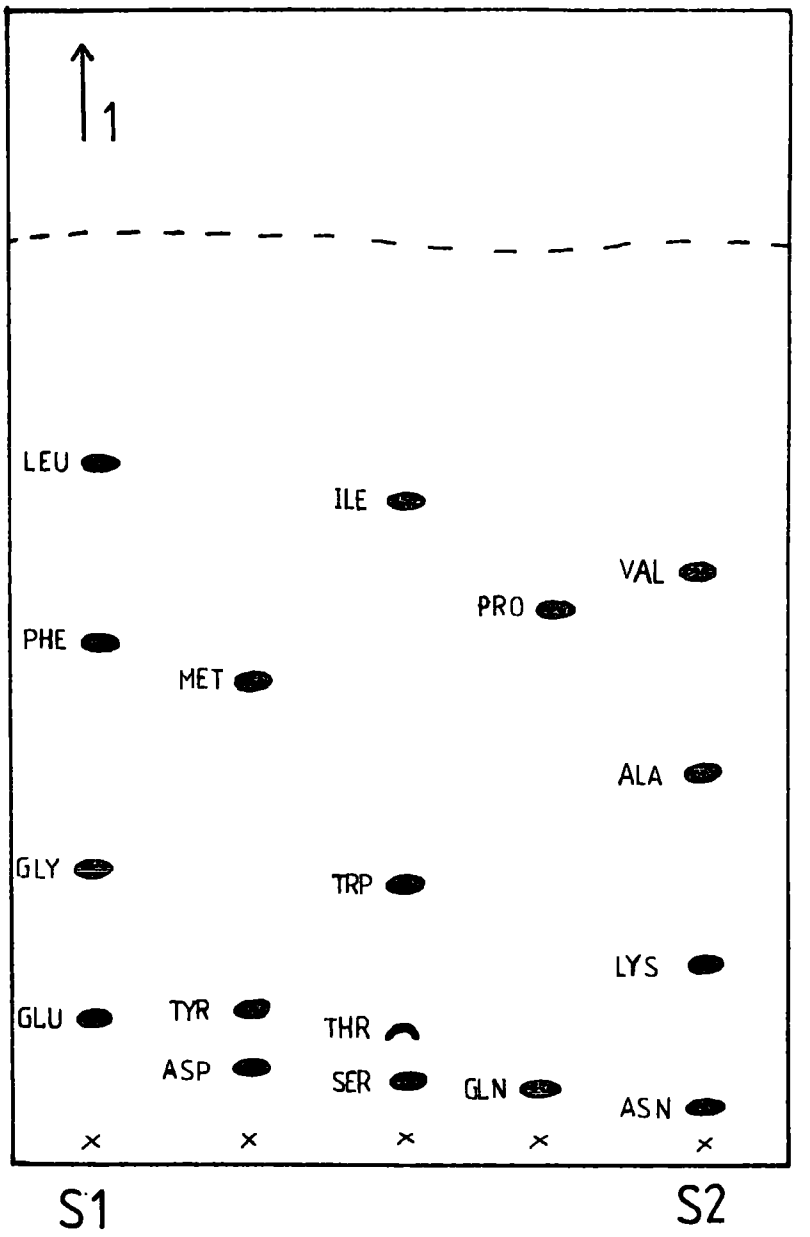

1
LEU
ILE
VAL
PRO
PHE
MET
ALA
GLY
TRP
LYS
GLU
TYR
THR
ASP
SER
GLN
ASN
S1
S2

FIGURE 6.

Thin layer chromatography of phenylthiohydantoin derivatives of amino acids

Development was by solvent 1 (Heptane: propionic acid: dichloroethane, 58:17:25, by vol) followed by solvent 2 (heptane: n-butanol: 75% formic acid, 50:30:9, by vol).

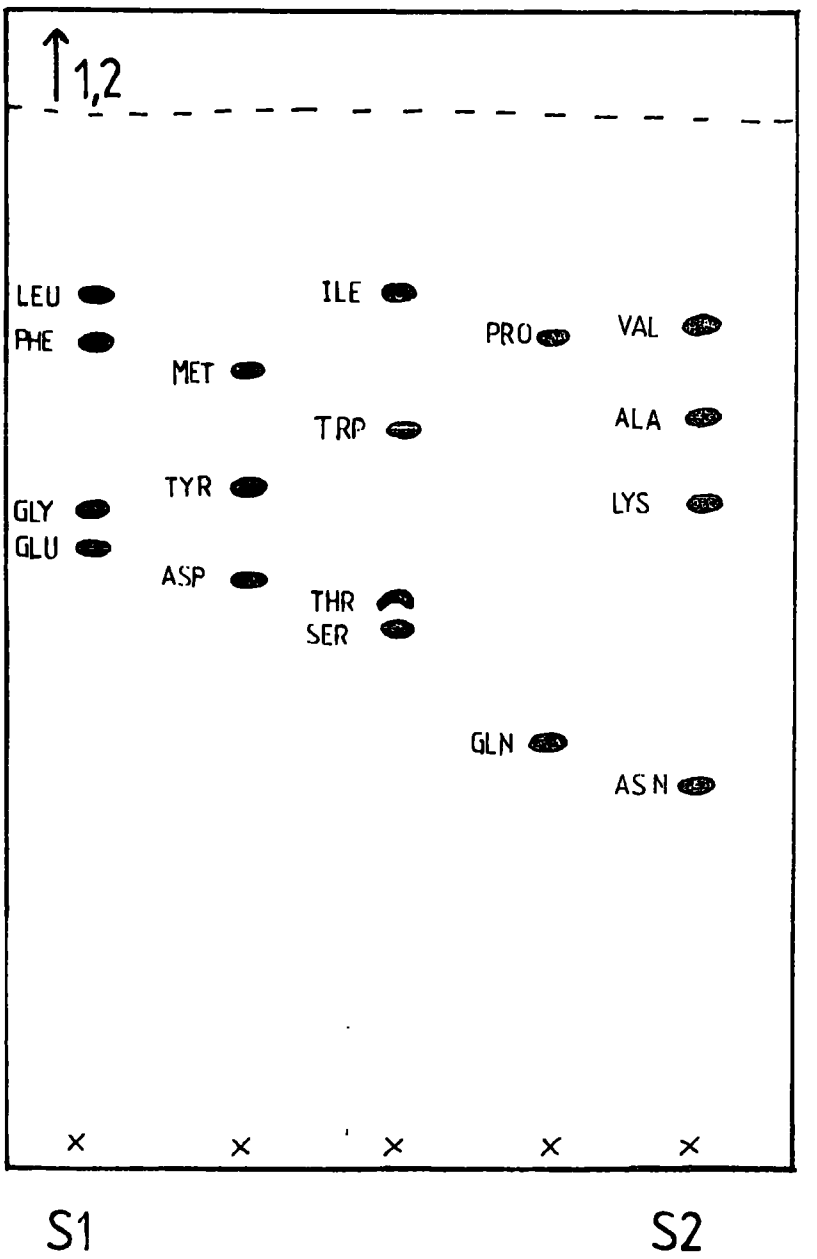

1,2
LEU
PHE
MET
ILE
PRO
VAL
TRP
ALA
TYR
LYS
GLY
GLU
ASP
THR
SER
GLN
ASN
S1
S2

FIGURE 7.

Gas chromatography of PTH derivatives of amino acids

Amino acids identified using programme A (temperature range, 160-270°C; 8° per min).

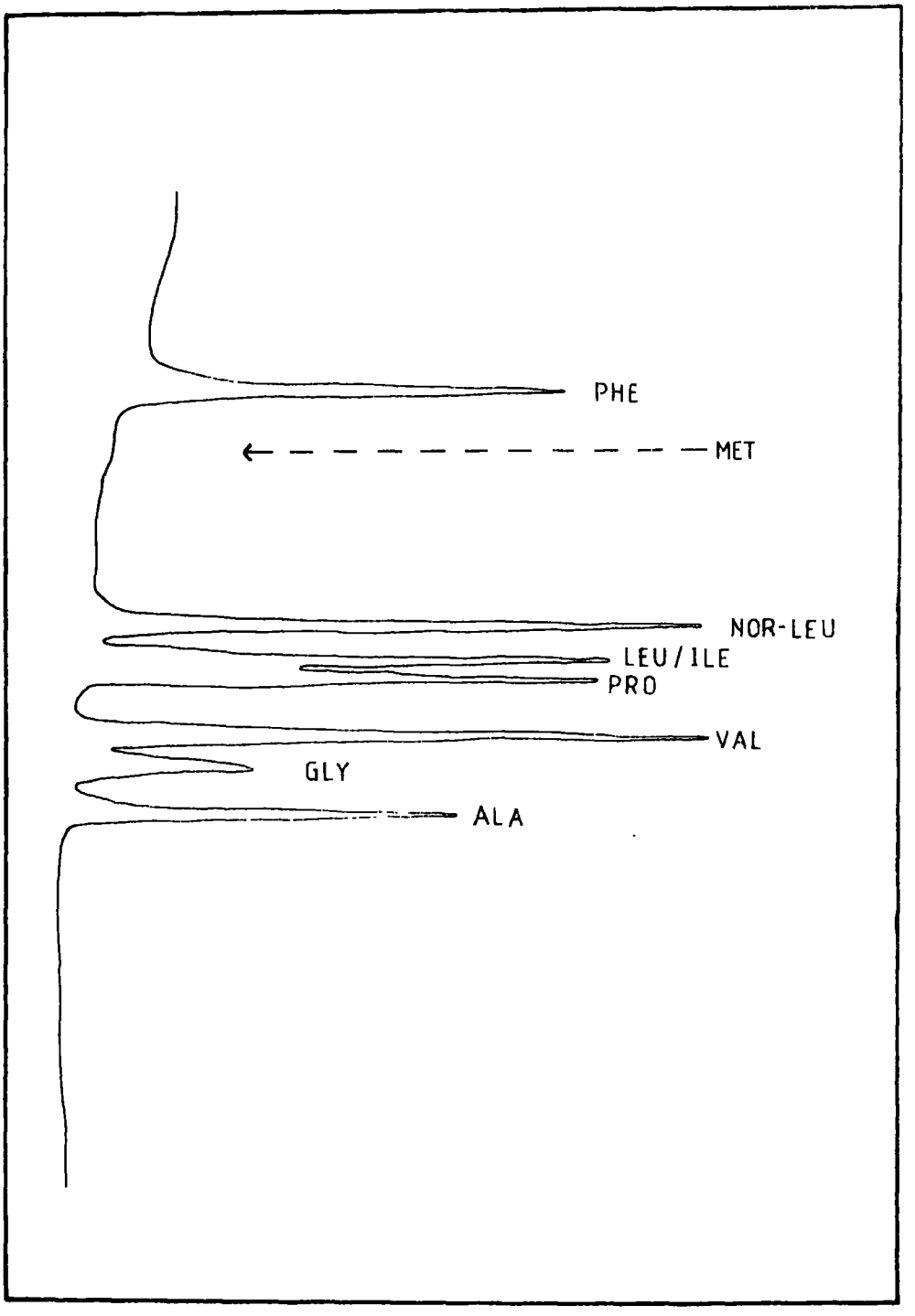

PHE
MET
NOR-LEU
LEU/ILE
PRO
VAL
GLY
ALA

FIGURE 8.

Gas chromatography of PTH derivatives of amino acids

Amino acids identified using programme B (temperature range, 200-270°C; 6° per min; sample silylated).

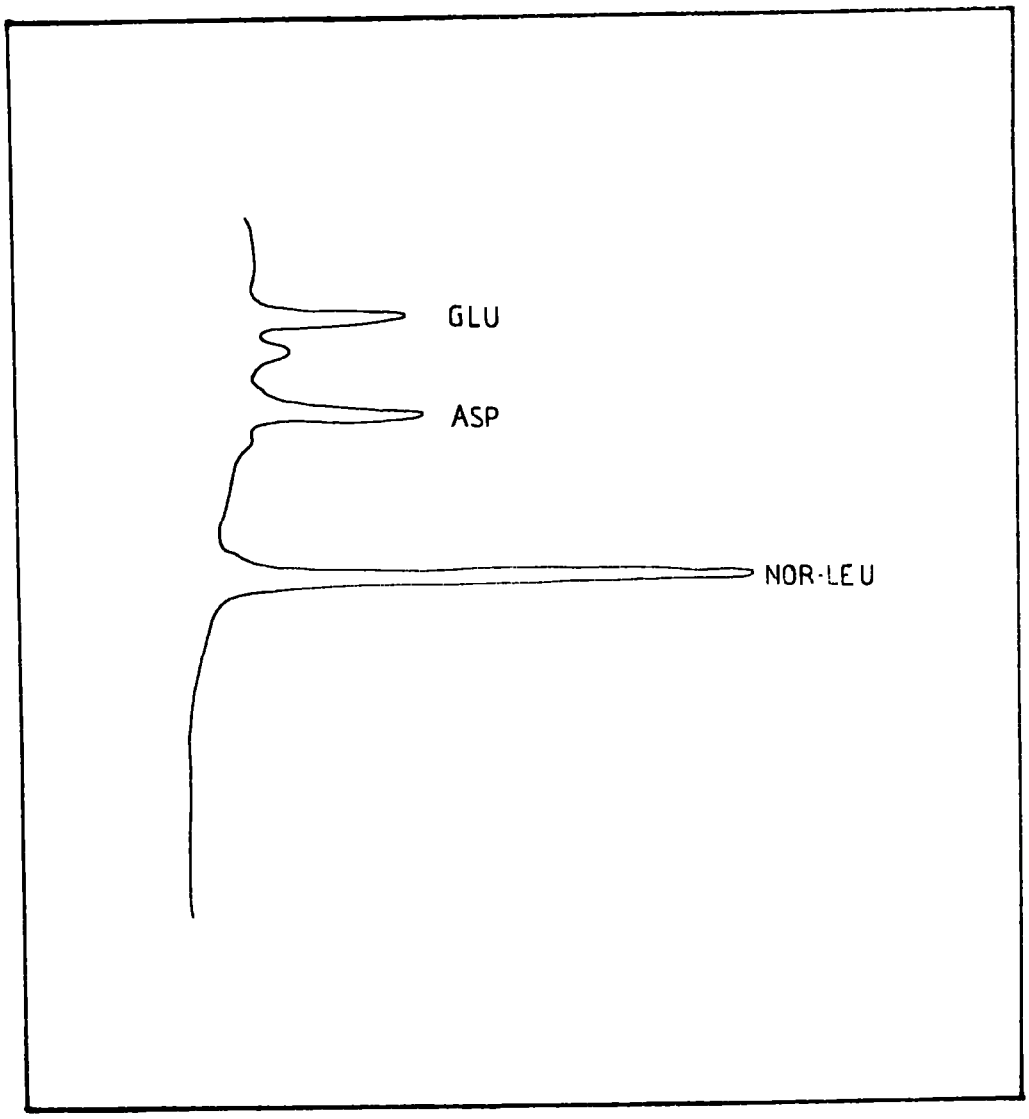
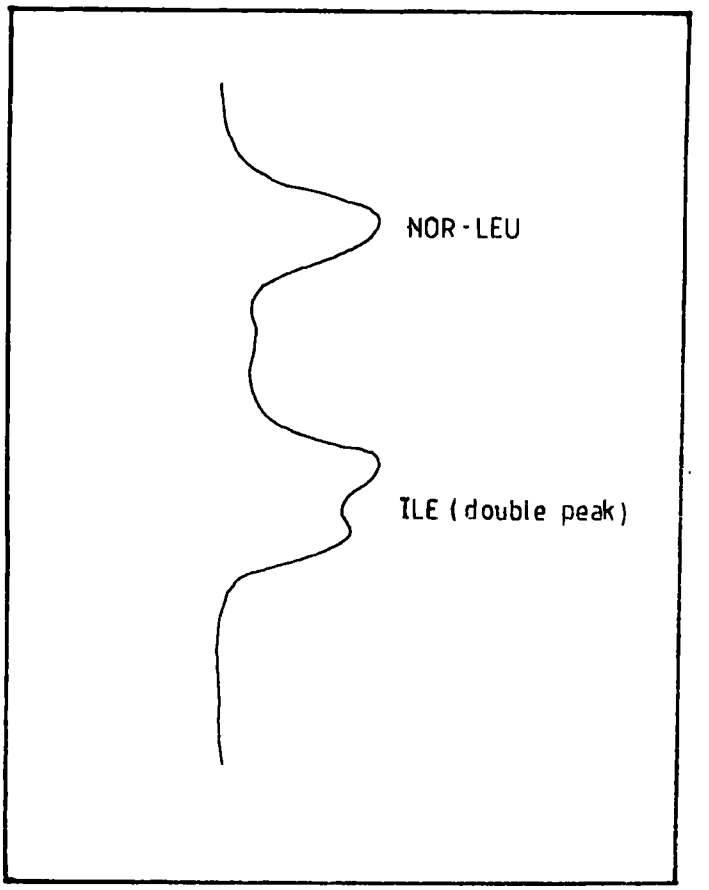
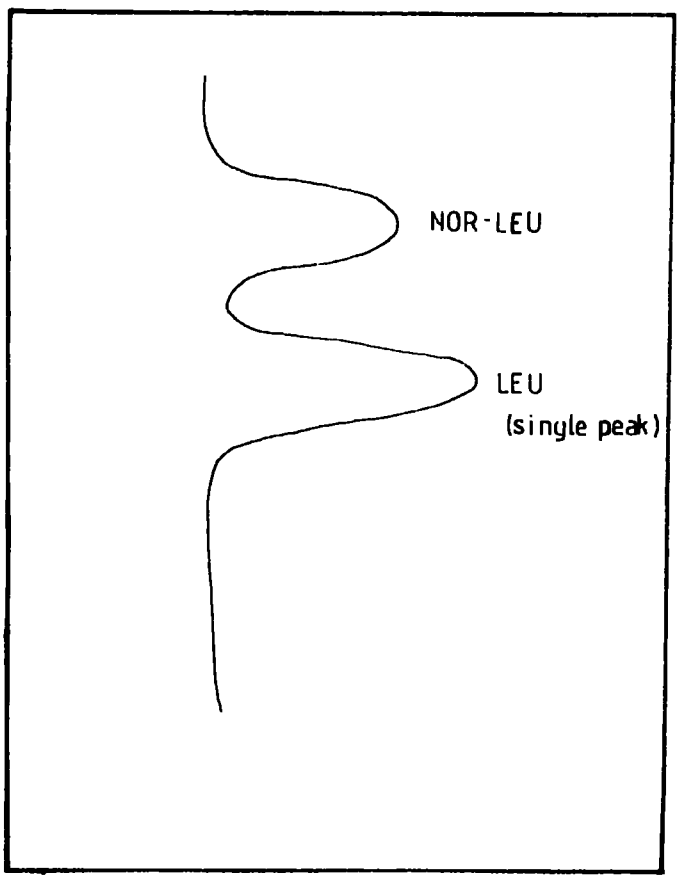

GLU
ASP
NOR·LEU

FIGURE 9.

Gas chromatography of PTH derivatives of amino acids

Amino acids identified using programme C (isothermal, 200°C; sample silylated).


NOR-LEU
ILE (double peak)


NOR-LEU
LEU
(single peak)

Methods of Computer Analysis of Amino Acid Sequence Data

Computer operations were carried out on an IBM 370/168 computer using the first 40 N-terminal amino acid residues of plastocyanin from seven pteridophytes and three gymnosperms. The analysis used programmes by Gleaves (unpublished).

Up to nine species could be examined at one time, using an 'exhaustive search' based on a method by Fitch (1970). This compared all possible topologies, finally selecting the most parsimonious tree (or trees) i.e. which needed the smallest number of amino acid substitutions in its construction. Further species were added using a second programme which compared the effects of moving branches of the original tree produced as above.

Calculation of branch lengths

Branch lengths were calculated using the method of Dayhoff (1972). At residues where heterogeneity was observed the amino acid which gave the most parsimonious result was chosen.

Input of Data

Where there was some doubt as to the identity of an amino acid, for example at position 36 of the *Blechnum spicant* and *Dryopteris filix-mas* sequences the alternative

possibilities were used as though the sequence was heterogeneous. In such cases the computer chose the most suitable amino acid on the basis of parsimony.

RESULTS

Extraction and Purification

No two batches of starting material exhibited exactly the same characteristics, even when using material taken from the same site at the same time and treating them, as far as possible, in exactly the same manner. For this reason, general results only can be given here. Any characteristics unique to one species are shown in Table 4.

Fresh plant material was collected whenever possible, immediately before extraction. Occasionally, if the collection procedure was lengthy, the fresh material was collected the previous day and kept overnight at $4^{o}C$. If frozen material had to be used, this was stored at -20^{o} and used as soon as possible. In all cases, only the softer, fresh, green parts of the plant were used. At all stages in the extraction the buffer was kept below $4^{o}C$ and each step was executed speedily, keeping the length of time that the protein was in contact with acetone to a minimum.

No plastocyanin was extracted during tests using aqueous blending buffer, consequently acetone was included in all blending buffers. Occasionally, during the initial homogenisation a great deal of gelatinous material was produced. This was eliminated by use of a stronger concentration of acetone up to a maximum of 50% acetone.

The nature of the precipitate formed in 80% acetone buffer varied not only from species to species, but between different batches of the same species. If it settled quickly, it was possible to decant off and discard the supernatant. At other times the precipitate was very light and did not settle easily and consequently was separated by centrifugation.

Both the size of the column and also the method of application of the sample to the resin varied during the initial ion-exchange chromatography step. The size of the column was judged by eye, depending upon the volume and colour of the liquid to be applied. If the liquid was viscous, this had the effect of blocking the column, which soon brought the flow rate almost or entirely to zero. In such cases, although the column method was preferred, the sample was batched with the resin, washed thoroughly and then eluted from the column by the normal method.

At this stage, the crude plastocyanin was often contaminated by dark brown materials which coloured the whole of the resin and some also tended to elute off the column at the same time as the plastocyanin. Occasionally these were present in such large amounts that they obliterated the blue colour of the oxidised plastocyanin

completely, and in extreme cases even prevented detection of plastocyanin in the spectrophotometer. In such a case, the darkest brown fractions were pooled and the plastocyanin separated from the contaminants by repeated ion-exchange chromatography and gel filtration. (see Figure 10).

A second DE23 Sephadex ion-exchange column removed an appreciable amount of these contaminants. A dark brown band of polyphenolic materials and ferredoxin was left at the top of the column and the plastocyanin separated as a distinct paler brown band further down the column. In all but one case it was possible to separate the plastocyanin from this pale brown band by washing the column with 50 mM Tris-HCl buffer (pH 7.2) containing 100 mM NaCl. This had the effect of moving the impurities further down the column, so that on eluting with 50 mM Tris-HCl buffer (pH 7.2) containing 200 mM NaCl the brown impurities came off before the plastocyanin. In the case of Cycas revoluta however, washing the column with 50 mM Tris-HCl buffer (pH 7.2) containing 100 mM NaCl actually caused the plastocyanin to elute off.

Further purification by means of gel filtration on Sephadex G75 removed yet more impurities; a blue band of oxidised plastocyanin moved ahead of a band of brown or occasionally yellow contaminants. All fractions containing

plastocyanin were then pooled and final purification was effected by gradient elution of reduced plastocyanin from an ion-exchange column of DEAE-Sephadex. A brown band was left at the top of this column, leaving the final batch of plastocyanin fractions, with an absorbance ratio of less than 3, colourless.

Storage at -20°C did not appear to affect the plastocyanin appreciably except in one case, Ephedra, when the unpurified plastocyanin was lost completely during storage at -20°C for 3 days.

Table 4

Experimental details and yield data for extraction procedures

General results and common problems are discussed in the text.

Details of experimental conditions and specific problems encountered for each species are given here.

Where several batches of the same species were used, separate details are given only if experimental conditions were different; otherwise the total amount of plant material is given and the range of yield of crude plastocyanin is quoted.

Yield data of plastocyanin from lower plants

Species	Concentration of acetone in blending buffer (%)	Total amount of prepared plant material used (Kg)	Yield of crude plastocyanin (mg per Kg plant material)	Conditions of picking and extraction	Preparation for extraction	Comments on extraction procedure	Total crude plastocyanin (mg)	Total pure plastocyanin (mg)
Cycas revoluta	30	2 (40 leaves)	22.8	fresh, same day	Leaflets stripped from mid-rib	Leaves looked hard and leathery but blended well. This plastocyanin eluted from DE23 Sephadex columns at a lower concentration than others i.e. 100 mM NaCl	45.6	22.1
Abies grandis	30	1	1.5	fresh, next day	Leaves stripped from branches	Amount of crude protein recorded could be low due to masking effect of polyphenols or tannins. Yield low and collection difficult, therefore no further extraction attempted.	1.5	-
Picea abies	30	1	trace	fresh, next day	Leaves impossible to strip from branches in quantity. Branches therefore cut into small pieces	A large amount of non-green material was consequently included in the blend. Woody material caused heating problems during blending. Extraction abandoned	trace	-
	30	1	trace	fresh, same day				

Larix decidua	30	1	-	frozen	Leaves stripped from branches in all batches	This was fresh, soft, green material and cells were apparently well ruptured, therefore there was no reason for the failure of the extraction		
	30	2	-	fresh, next day early summer				
	30	2	-	fresh, next day late summer				
	30	1	-	fresh, same day, late summer				
	50	1	-	fresh, same day, late summer				
	aqueous	1	-	fresh, same day, late summer			-	-
Pinus laricio	30	1	-	fresh, next day	Leaves stripped from branches	Blended well, no problems encountered	-	-
Sequoia sempervirens	30	1	?	fresh, next day	Leaves stripped from branches	Blended well. Trace of plastocyanin possibly masked by tannins. Supply limited therefore extraction not continued	-	-

Thuja plicata	30	1	-	fresh, next day	Leaves stripped from branches	Blended well, no apparent reason for failure		
	aqueous	1	-	fresh, next day				
	30	1	-	fresh, frozen overnight			-	-
Cupressus lawsonia	30	1	trace	fresh, same day early summer	Branches chopped into small pieces	Homogenate was extremely slimy. Extraction therefore abandoned		
	30	1	trace	fresh, same day late summer				
	30	1	-	fresh, frozen overnight			-	-
	50	2	0.5	fresh, same day, late summer	"	Mucus materials were precipitated in 50% acetone buffer	1	-
Araucaria araucana	30	4	2	fresh, same day	Leaves removed from branches.	Mucus type substances present which caused extraction problems		
	50	12	4	fresh, next day	"	Mucus materials precipitated in 50% blending buffer. No other extraction problems	40.0	39.2
Taxus baccata	30	1	2	frozen for one week	Leaves removed from all branches except new year's growth	Best yields obtained when young shoots used. Freezing overnight did not affect the yield, but made the leaves easier to blend. Freezing for longer periods caused a deterioration in yield		
	30	18	5-7	fresh, next day. Batches containing older leaves were frozen overnight			104	59.4

Ginkgo biloba	30	375	trace	fresh, same day	Leaves removed from branches	Leaves were soft, fresh and bright green. Cells appeared to have been well ruptured. A small amount of plastocyanin may have been present which was masked by tannins. A limited amount of material was available therefore further extraction was not feasible	trace	-
Ephedra sp.	30	4	5	fresh, next day	Green parts of the plant were stripped from the non-green or very woody parts	Some freezing of material was inevitable, due to method of collection and this plant obviously did not store well. No plasto-cyanin was detected on attempting to purify samples after only one week in store at -20°C.		
	30	4	3	fresh, after two days				
	30	2	1	frozen for three days			29.0 lost	-
Selaginella spp.	30	6.6	0-12	fresh, next day?	Washed and non-green parts removed	Yields varied with age and condition of material. Transport problems meant some samples were not very fresh.	16.6	-
Equisetum arvense	30	26.5	4-9	fresh, samd day	No preparation required		147.6	71.2
Equisetum sylvaticum	30	10	4-6	fresh, same day	No preparation required		54.6	26.0
Angiopteris palmiformis	30	0.6	4	fresh, next day	Mid-rib removed from fronds	The amount of plant material required was not available	2.4	-
Osmunda regalis	30	1	22	fresh, next day	No preparation required			
	30	1	28	fresh, same day			52	45.9
Dicksonia antarctica	30	3	-	fresh, next day	Mid-rib removed from fronds	Material was tough and temperature of blend rose above 4°C. No more material was available	-	-

Thelypteris erubescens	30	0.2	-	fresh, next day	No preparation required	Material blended well, but insufficient quantity available	-	-
Pteridium aquilinum	30	12	0-7	fresh, same day, early summer	Mid-rib removed	Homogenate contained a large amount of mucous material. In some batches sample was eventually lost as a result		
	30	20	8-10	fresh, same day, late summer	Mid-rib removed	Mucous materials absent, but soluble dark brown contaminants present	280.4	200.0
Pteris cretica	30	1.75	8	fresh, same day	Mid-rib removed	No further extraction possible due to lack of material	14.2	-
Blechnum spicant	30	8	4-8	fresh, same day	No preparation required		51.6	26.0
Phyllitis scolopendrium	30	1.2	trace?	fresh, next day	No preparation required	Possible trace of plastocyanin masked by tannins. Limited material available	trace	-
Dryopteris dilatata	aqueous	1	-	fresh, same day	Mid-rib removed		-	-
	30	11	2-10	fresh, same day	Mid-rib removed		53	20.8
Dryopteris filix-mas	30	8	6	fresh, same day	Mid-rib removed		49	21.28
Salvinia natans	30	2.5	0-2	fresh, same day	No preparation required	Quantity of material required was not available	1.4	-
Conocephalum conicum	30	3	0-4	fresh, next day	Washed and cleaned thoroughly. Only young plants used where possible	Good yields obtained with young fresh green material only	4	-
Marchantia polymorpha	30	1	1	fresh, same day	Washed and cleaned thoroughly. Only young plants used where possible		1	-

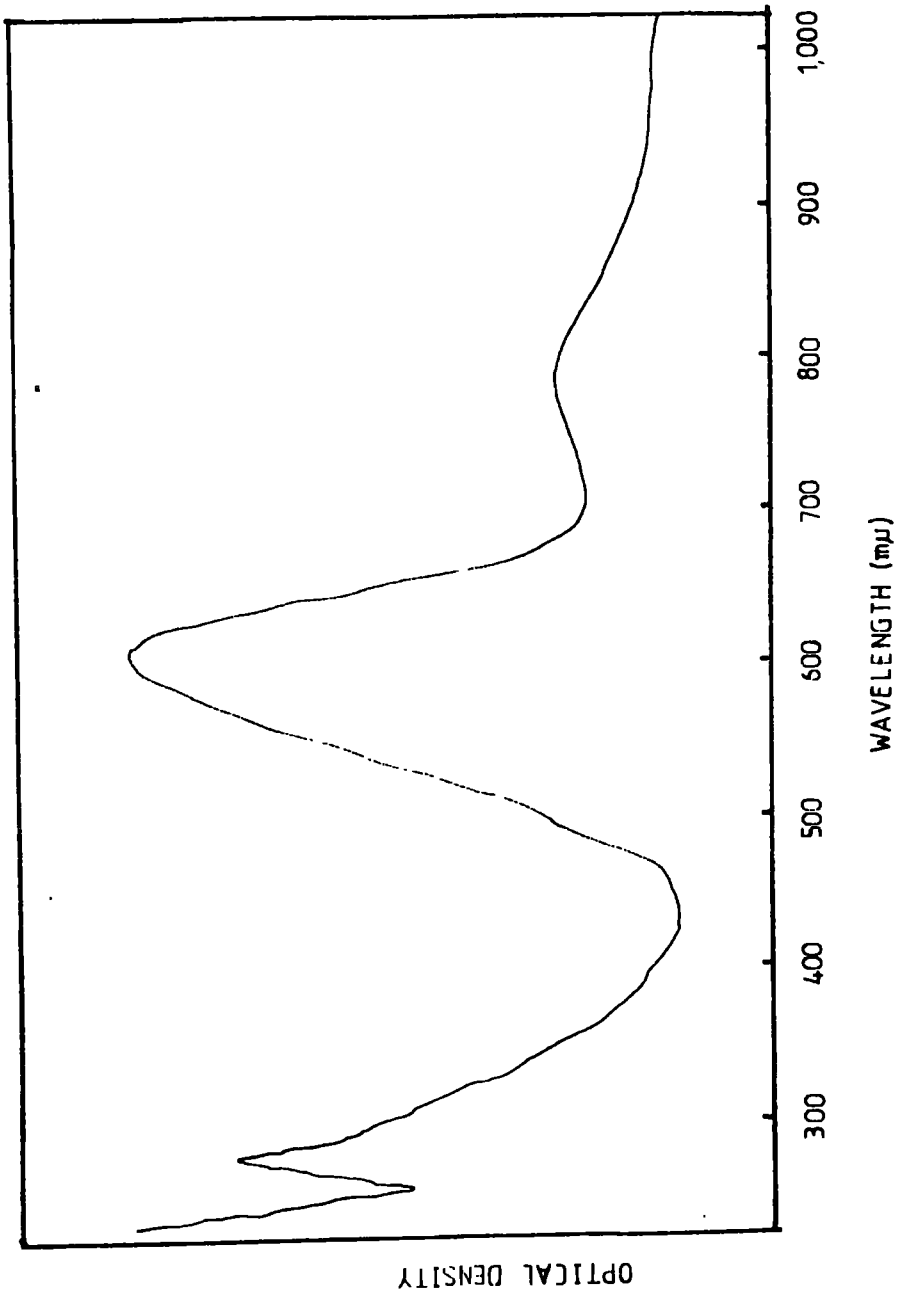
Sphagnum spp.	30	1	trace	fresh, next day	Green shoots only were collected and gently squeezed in a muslin bag to extract excess water		trace	-
Polytrichum spp.	30	5.5	0-2	fresh, next day	No preparation required		2	-
Mnium hornum	30	0.5	trace	fresh, same day	No preparation required		-	-
Fontinalis antipyretica	30	0.3	-	fresh, next day	Green shoots only were used. These had to be cut off individually	Preparation for extraction extremely time consuming	-	-

FIGURE 10

Ultraviolet and visible absorbance spectrum of oxidised plastocyanin.

The peak in the ultraviolet region (centred at 278 nm) is probably due to the aromatic amino acids.

The peak in the visible region (at 597 nm) is characteristic of the molecule and due to the presence of the copper atom. Polyphenolic contaminants present during extraction tended to obliterate this peak.


OPTICAL DENSITY
300
400
500
600
700
800
900
1,000
WAVELENGTH (mμ)

Amino Acid Sequence of Plastocyanin from *Pteridium aquilinum*

The elution behaviour of *Pteridium aquilinum* plastocyanin on Sephadex G75 during purification suggested that it had a molecular weight of approximately 10,500, which is similar to those reported for plastocyanins from higher plants (Boulter *et al.*, 1977).

The complete sequence contained 99 residues, giving a sequence molecular weight of 10,500 and was determined from the evidence of proteolytic digestion of two large CNBr peptides and direct manual analysis of a smaller CNBr fragment, plus data obtained from the Beckman Automatic Sequencer. A total of 20 μmol (200 mg) of protein was used. The sequence is shown in Figure 12, giving the points of cyanogen bromide and enzyme cleavage, together with the overlapping peptides from which the sequence was deduced. The protein was shown to have N-terminal alanine and C-terminal tryptophan. The amino acid composition was obtained from three duplicate 50 μg (5 nmol) samples, hydrolysed for 24, 48 and 72 h respectively and this is shown in Table 5.

Table 5

Amino acid composition of bracken plastocyanin

	Mean values 24 h hydrolysis	Mean values 48 h hydrolysis	Mean values 72 h hydrolysis	Value taken	Sequence
Asp	9.6	12.1	10.7	10.8	7.5
Asn					2
Thr	10.6	10.0	9.1	11.4	11
Ser	4.4	3.7	2.7	5.6	5
Glu	11.1	9.5	9.9	10.2	7.5
Gln					3
Pro	5.6	6.2	5.8	5.9	6
Gly	10.9	11.6	10.8	11.1	11
Ala	12.3	13.2	12.4	12.6	9
Cys	0.8	1.0	1.0	0.9	1
Val	8.5	9.2	9.7	9.7	10
Met	0.8	1.1	0.9	0.9	2
Ile	3.4	3.6	3.8	3.8	4
Leu	3.7	4.2	3.8	4.2	4
Tyr	2.8	2.9	2.6	2.8	3
Phe	4.4	4.7	4.4	4.5	5
His	2.1	2.2	2.2	2.2	2
Lys	6.4	4.6	4.7	5.2	5
Arg	0.2	0.4	0.3	-	-
Trp	not determined quantitatively			+	1
			Total		99

The results are expressed as residues/mol. Values taken were the average of 24 and 72 h hydrolysates except Thr and Ser were corrected for losses (Moore & Stein, 1963) and maximal values were taken for valine, isoleucine and leucine. Tryptophan was not determined.

Cyanogen bromide cleavage (Figures 11A and 11B)

CNBr partially cleaved S-carboxymethylated plastocyanin into three major fragments, which were purified by gel filtration and named X-1, X-2 and X-3 in order in which they eluted from the column after the uncleaved protein.

N- and C-terminal analyses of these peptides were compared with similar analyses of uncleaved protein with the result that the fragments were unequivocally positioned in the same order in the protein sequence (Table 7). Amino acid analyses of the three fragments were in reasonable agreement with analysis of uncleaved protein (Table 6). Tryptophan was shown to be present in the total protein and present, though in a chemically degraded form, in the X-3 fragment but absent from the X-1 and X-2 CNBr peptides (Table 8).

Similar fragments have been obtained previously by CNBr cleavage of plastocyanin from higher plants (Boulter et al., 1977).

FIGURE 11A

Chromatography of CNBr peptides from bracken plastocyanin

A column (1 cm x 200 cm) of Sephadex G 50 fine equilibrated and eluted with 70% (v/v) formic acid was used. The peptide mixture was applied in a volume of less than 1 ml; blue dextran and methylene blue were included as markers. 2 ml fractions were collected at a flow rate of 20 ml/h in an LKB Ultra rack fraction collector fitted with a Uvicord III absorbance monitor.

_____ absorbance at 280 nm

- - - - - - - - - - - - - absorbance at 206 nm

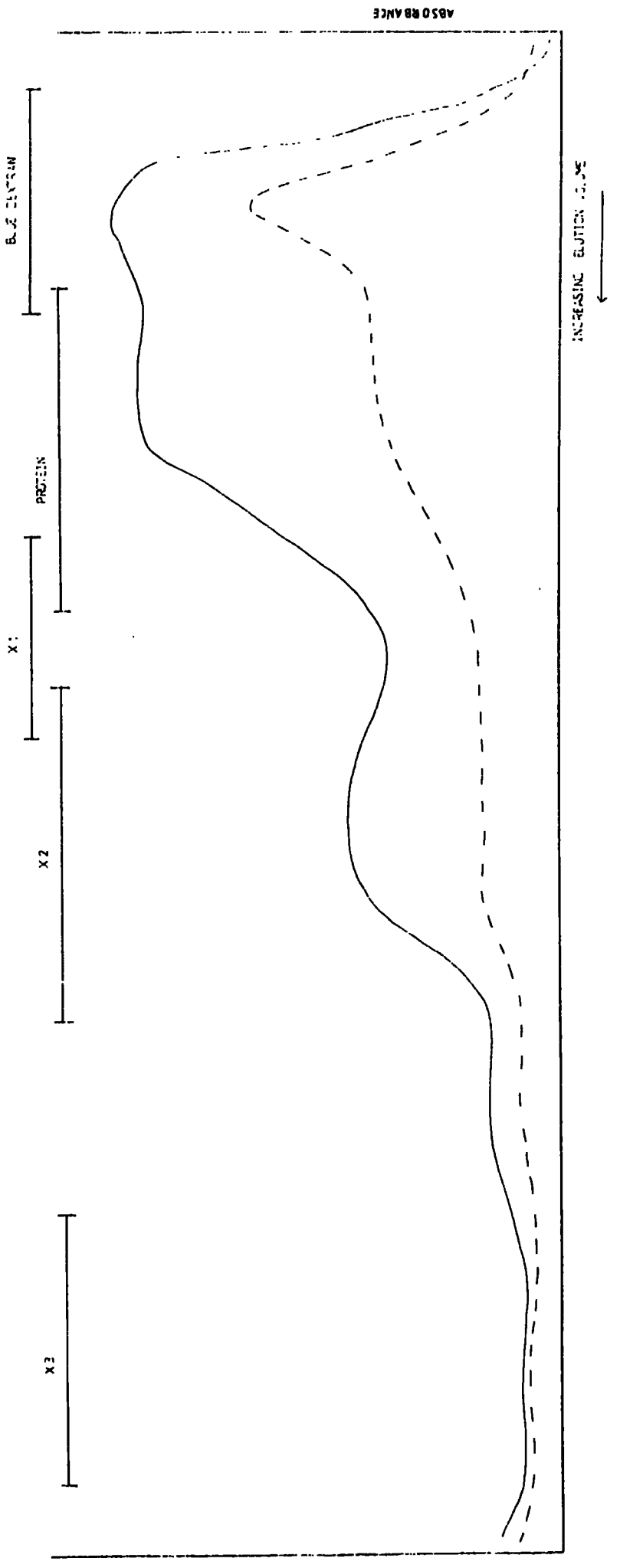

ABSORBANCE
BLUE DEXTRAN
PROTEIN
X1
X2
X3
INCREASING ELUTION VOLUME

FIGURE 11B

Chromatography of CNBr peptides from bracken plastocyanin

A column (1 cm x 200 cm) of Bio Gel P10 fine equilibrated and eluted with 70% (v/v) formic acid was used. The peptide mixture was applied in a volume of less than 1 ml; blue dextran and methylene blue were included as markers. 2 ml fractions were collected at a flow rate of 20 ml/h in an LKB Ultra rack fraction collector fitted with a Uvicord III absorbance monitor

_____absorbance at 280 nm

- - - - - - - - - - - -absorbance at 206 nm

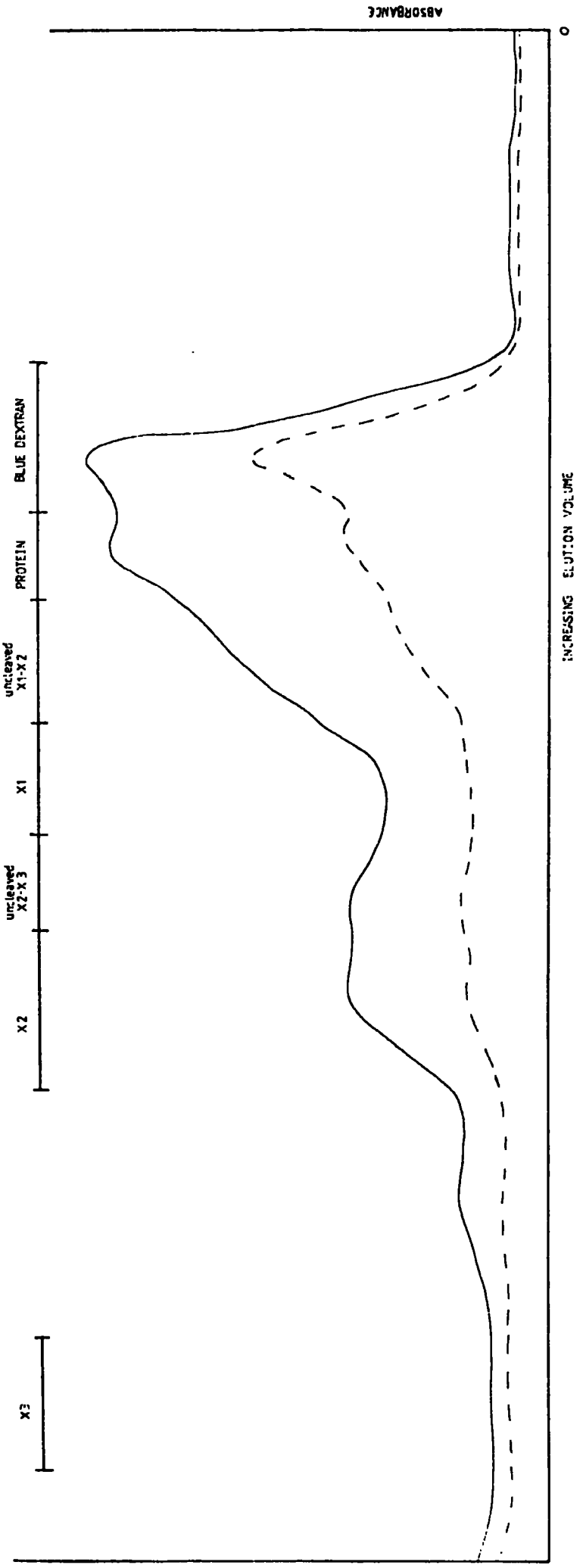

ABSORBANCE
0
INCREASING ELUTION VOLUME
BLUE DEXTRAN
PROTEIN
uncleaved X1-X2
X1
uncleaved X2-X3
X2
X3

Table 6

Amino acid composition of bracken plastocyanin CNBr fragments

	X-1 fragment				X-2 fragment				X-3 fragment			
	24 h	72 h	Value taken	Sequence	24 h	72 h	Value taken	Sequence	24 h	72 h	Value taken	Sequence
Asp	4.7	3.9	4.3	3[a]	5.5	5.3	5.4	4.5[b]				
Asn				1				1				
Thr	4.7	3.7	5.3	5				5	0.9	0.5	1.2	1
Ser	2.6	1.6	3.3	3				2				
Glu	6.0	5.8	5.9	6	3.2	3.9	3.6	1.5[b]				
Gln								3				
Pro	4.3	4.2	4.3	4[a]	1.6	1.5	1.6	2				
Gly	6.8	6.7	6.8	7	2.7	2.5	2.6	3	1.0	1.0	1.0	1
Ala	7.2	7.3	7.3	7	2.3	2.4	2.4	2				
Cys				-	1.0	0.6	0.8	1				
Val	6.5	7.6	7.6	8	1.2	1.4	1.4	1	0.9	1.0	1.0	1
Hse	0.7	0.8	0.8	1	0.5	0.6	0.6	1				
Ile	2.6	2.8	2.8	3				-	1.2	1.3	1.3	1
Leu	2.2	1.9	2.2	2	2.0	2.0	2.0	2				
Tyr				-	2.4	2.8	2.6	3				
Phe	3.7	3.6	3.7	4	1.0	1.0	1.0	1				
His	0.8	0.8	0.8	1	0.4	1.8	1.1	1				
Lys	2.2	2.2	2.2	2	1.5	1.1	1.3	1	1.6	1.7	1.7	2
Trp				-								1

The results are expressed as residues/mol. Values taken were the average of 24 and 72 h hydrolysates except Thr and Ser were corrected for losses (Moore & Stein, 1963) and maximal values were taken for valine, isoleucine and leucine. Tryptophan was not determined.

[a] Heterogeneity at residue 42; Asp was assumed to be present in the largest amount

[b] Heterogeneity at residue 68; Asp and Glu were assumed to be present in equal amounts.

Table 7

N- and C-terminal analyses of plastocyanin from Pteridium aquilinum, and CNBr peptides X-1, X-2 and X-3

Material	N-terminus	Liberated C-terminal residues	Semi-quantitative estimation of liberated amino-acids after various incubation periods (min)*							
			0.50	1.0	3.0	5.0	10.0	30.0	1h	2h
Plastocyanin	Ala	Tryp**			?	tr	tr	+	++	+++
		Val					tr	tr	+	+++
		Ile							+	++
		Lys								+
		Thr								?
X1	Ala	Hse	+	+	++	++	+++	+++	+++	+++
		Ser			tr	+	+	++	++	+++
		Leu					tr	+	++	++
		Glu								tr
X2	Asx	Hse	+	+	++	++	++	+++	+++	+++
		Ala		+	++	++	++	+++	+++	+++
		Gly		+	++	++	++	+++	+++	+++
		Ser		tr	+	+	+	+	+	+
		Asp				?	?	tr	tr	tr
X3	Lys	Tryp		+	++	+++	+++	+++	+++	+++
		Val						tr	+	++
		Ile							tr	++
		Lys								++

*-Amino acids estimated by intensity of dansyl-derivative on polyamide sheet chromatography

**-Underlined residue = C terminus

Table 8

Reaction of native protein and CNBr peptides with Ehrlich reagent

Material	Result	Conclusion
Plastocyanin	purple colour developed	tryptophan present
X-1	no reaction	no tryptophan
X-2	no reaction	no tryptophan
X-3	yellow orange colour developed	chemically degraded tryptophan present

Small samples of protein or peptide were dried on electrophoresis paper, dipped in freshly-prepared Ehrlich reagent (see Methods) and air-dried for 30-60 min. to allow colour to develop.

Amino acid sequence of X-1 CNBr fragment

The sequence of the X-1 CNBr fragment was determined unequivocally by use of the automatic sequencer and dansyl-Edman analysis of chymotryptic and papain peptides. All peptides overlapped, and a notable feature is the heterogeneity found at position 42. In addition the amino acid composition data of the X-1 fragment and some of the enzyme-generated peptides were identical, further confirming the sequence.

Anomalous results were obtained with one batch of lyophilised X1, stored at $4^{o}C$ in which unspecific cleavage occurred at the Ala-Gly bond between positions 23 and 24.

Chymotryptic digest

20 mg of X-2 CNBr fragment were digested with chymotrypsin for 2 hours. The resulting peptides were separated by high voltage paper electrophoresis.

Four major peptides were obtained in good yield and sufficiently pure condition for quantitative amino acid analysis. The major points of chymotryptic cleavage were as expected, although cleavage to the C-terminal side of Lys, Val and Asp was unexpected and the resulting peptides only isolated in small quantitites.

The order of all the chymotryptic peptides was established by use of sequencer data which provided overlaps for every peptide. Confirmation that peptide X1C4 was C-terminal was obtained by C-terminal analysis of the complete fragment.

Chymotryptic peptides

Peptide X1C1 (1-12)

Ala-Lys-Val-Glu-Val-Gly-Asp-Glu-Val-Gly-Ser-Phe

Following eleven Edman degradation steps, dansylation without hydrolysis confirmed phenylalanine as the C-terminal residue. A consideration of the peptide mobility during electrophoresis at pH 6.5 placed residues 4 and 8 as glutamic acid and residue 7 as aspartic acid. Amino acid analysis results were in good agreement with the sequence data.

Peptide X1C1A (1-14)

Ala-Lys-Val-Glu-Val-Gly-Asp-Glu-Val-Gly-Ser-Phe-Lys-Phe

Only a small quantity of this peptide was isolated and dansyl-Edman analysis was inconclusive after the first three degradation steps. Digestion with carboxypeptidase-A for 1 and 4 h followed by dansylation yielded dansyl-phenylalanine and dansyl-phenylalanine plus bis-dansyl-lysine respectively. The remainder of the sequence was deduced from semi-quantitative analysis, electrophoretic mobility data and comparison with sequencer data.

Peptide X1C1B (1-2)

Ala-Lys

Only a small yield of this protein was obtained. Dansylation without hydrolysis after one Edman step yielded α-dansyl-ϵ-PTC-lysine.

Peptide X1C2 (13-29)

Lys-Phe-Thr-Pro-Asp-Thr-Ile-Thr-Val-Ala-Ala-Gly-Glu-Ala-Ile-Glu-Phe

After sixteen Edman degradation steps, phenylalanine was identified by dansylation without hydrolysis. Mobility data from electrophoresis at pH 6.5 indicated that two of the three acidic residues should be amines whereas mobility data from electrophoresis at pH 1.9 indicated that one amine was present. However, examination of automatic sequencer data and electrophoretic mobility data of papain peptides indicated that all three residues were acidic i.e. residue 17, aspartic acid and residues 25 and 28, both glutamic acid. Since no other chymotryptic peptides from this part of the fragment were isolated with alternative mobilities, these anomalous mobilities were probably due to the large size of the peptide and not an indication of polymorphism. Amino acid analysis results were in good agreement with sequence data.

Peptide X1C3 (30-41)

Thr-Leu-Val-Gly-Glu-Thr-Gly-His-Asn-Val-Val-Phe

Dansylation without hydrolysis after twelve Edman

degradation steps confirmed phenylalanine as the C-terminal residue. Examination of the pH 6.5 electrophoretic mobility and automatic sequence data placed residue 38 as asparagine, assuming a value of +1 for His at that pH. Amino acid analysis results were in good agreement with sequence data.

Peptide X1C3A (33-42)

Gly-Glu-Thr-Gly-His-Asn-Val-Val-Phe-Asp

This peptide was isolated in a small quantity. Dansyl-Edman analysis was inconclusive beyond residue 39, the remainder of the sequence being determined from semi-quantitative amino acid analysis and automatic sequencer data. The distribution of acidic residues was determined by examination of electrophoretic mobility data of this and other chymotryptic and papain peptides, plus automatic sequencer data.

Peptide X1C4 (42-57)

Asp-Ile-Pro-Ala-Gly-Ala-Pro-Gly-Pro-Val-Ala-Ser-Glu-Leu-Ser-Hse

Following fifteen Edman degradation steps, homoserine was identified as the C-terminal amino acid by dansylation without hydrolysis. The peptide mobility at pH 6.5 indicated that residues 42 and 54 were aspartic acid and glutamic acid respectively. The amino acid composition of the peptide confirmed the sequence data.

Peptide X1C4A (43-57)

Ile-Pro-Ala-Gly-Ala-Pro-Gly-Pro-Val-Ala-Ser-Glu-Leu-Ser-Hse

Only a small amount of this peptide was isolated. Dansyl-Edman analysis was tentative beyond residue 52 and inconclusive beyond residue 54. Other amino acids were placed from semi-quantitative amino acid analysis and sequence data from peptide X1C4.

Table 9

Peptides from digestion of X1 CNBr fragment using Chymotrypsin

Peptide/Position	Mobility (pH 6.5)	Mobility (pH 1.9)	Dansyl-Edman Results
X1C1 (1-12)	1.11	0.53	Ala-Lys-Val-Glu-Val-Gly-Asp-Glu-Val-Gly-Ser-Phe
X1C1A (1-14)	0.56	0.68	Ala-Lys-Val-(Glu, Val, Gly, Asp, Glu, Val, Gly, Ser, Phe)-Lys-Phe
X1C1B (1-2)	-	1.6	Ala-Lys
X1C2 (13-29)	0	0.18	Lys-Phe-Thr-Pro-Asp-Thr-Ile-Thr-Val-Ala-Ala-Gly-Glu-Ala-Ile-Glu-Phe
X1C3 (30-41)	0	0.50	Thr-Leu-Val-Gly-Glu-Thr-Gly-His-Asn-Val-Val-Phe
X1C3A (33-42)	0.70	0.54	Gly-Glu-Thr-Gly-His-Asn-Val-(Val, Phe, Asp)
X1C4 (42-57	1.08	0.24	Asp-Ile-Pro-Ala-Gly-Ala-Pro-Gly-Pro-Val-Ala-Ser-Glu-Leu-Ser-Hse
X1C4A (43-57)	0.60	0.25	Ile-Pro-Ala-Gly-Ala-Pro-Gly-Pro-Val-Ala-Ser-Glu-(Leu, Ser, Hse)

Papain digest

6 mg of X-1 CNBr fragment were digested with papain for 2 hours. The resulting peptides were separated by high voltage paper electrophoresis.

Unfortunately, due to the relatively non-specific nature of this enzyme 'families' of peptides were found which differed only by one amino acid either at the C- or N-terminal. This made purification and sequence determination difficult. The peptides isolated in the greatest quantities were produced as a result of expected enzymic activity to the C-terminal of Ala, Gly and Ser, but unexpected cleavage also occurred to the C-terminal of Lys, Glu and Asp. Few peptides were obtained in sufficiently pure yield for accurate quantitative amino acid analysis.

The order of all the papain peptides was established by comparison with sequencer data and chymotryptic peptides.

Papain peptides

Peptide X1P1 (1-11)

Ala-Lys-Val-Glu-Val-Gly-Asp-Glu-Val-Gly-Ser

Following eleven Edman degradation steps, dansylation without hydrolysis confirmed serine as the C-terminal residue. A consideration of the peptide mobility during electrophoresis at pH 6.5 placed residues 4, 7 and 8 as glutamic acid, aspartic acid and glutamic acid respectively.

Peptide X1P2 (12-13)

Phe-Lys

Dansylation without hydrolysis after one Edman degradation step yielded α-dansyl-ε-PTC-lysine.

Peptide X1P3 (14-22)

Phe-Thr-Pro-Asp-Thr-Ile-Thr-Val-Ala

After eight Edman degradation steps, alanine was identified by dansylation without hydrolysis. The electrophoretic mobility of the peptide at pH 6.5 indicated the presence of one acidic residue, aspartic acid at position 17.

Peptide X1P4 (23-33)

Ala-Gly-Glu-Ala-Ile-Glu-Phe-Thr-Leu-Val-Gly

Some difficulty was initially encountered as this peptide had an identical N-terminus and mobility during electrophoresis at pH 6.5 as peptide X1P1. Separation was eventually achieved during electrophoresis at pH 1.9.

Following ten Edman degradation steps the C-terminal residue was identified as glycine by dansylation without hydrolysis. Electrophoretic mobilities indicated the presence of glutamic acid at residues 25 and 28.

Peptide X1P4A (29-42)

Phe-Thr-Leu-Val-Gly-Glu-Thr-Gly-His-Asn-Val-Val-Phe-Asp

Only a small quantity of this peptide was isolated.

Sequence determination was inconclusive from residue 37, the remaining sequence being determined from semi-quantitative amino acid analysis, comparison with peptides X1P5 and X1C3A plus a consideration of the mobility during electrophoresis at pH 6.5.

Peptide X1P5 (34-46)

Glu-Thr-Gly-His-Asn-Val-Val-Asp-Pro-Ile-Pro-Ala-Gly

Twelve Edman degradation steps followed by dansylation without hydrolysis established the sequence of this peptide. The electrophoretic mobility of the peptide at pH 6.5 indicated the presence of two acidic residues. Comparison with chymotryptic peptides suggested these were glutamic acid, residue 34 and aspartic acid, residue 41.

Peptide X1P5A (43-46)

Ile-Pro-Ala-Gly

Dansylation without hydrolysis after three Edman degradation steps confirmed glycine as the C-terminal residue.

Peptide X1P5B

Ile-Pro-Ala-Gly-Ala-Pro-Gly-Pro-Val-Ala

Only a small quantity of this peptide was isolated. Sequence determination was tentative for residue 47 and inconclusive beyond residue 49. The remaining sequence was determined from quantitative amino acid analysis and comparison with peptides X1P5A and X1P6.

Peptide X1P6 (47-53)

Ala-Pro-Gly-Pro-Val-Ala-Ser

The C-terminal residue was identified as serine by dansylation without hydrolysis after six Edman degradation steps. The proline residues at positions 48 and 50 were identified by dansylation followed by hydrolysis for 4 h only.

Peptide X1P7 (54-57)

Glu-Leu-Ser-Hse

Homoserine was identified as the C-terminal amino acid after three Edman degradation steps and dansylation without hydrolysis. The N-terminal amino acid was identified as glutamic acid on the basis of the peptide mobility during electrophoresis at pH 6.5.

Table 10

Peptides from digestion of X1 CNBr fragment using papain

Peptide/Position	Mobility (pH 6.5)	Mobility (pH 1.9)	Dansyl-Edman Results
X1P1 (1-11)	1.25	0.55	Ala-Lys-Val-Glu-Val-Gly-Asp-Glu-Val-Gly-Ser
X1P2 (12-13)	1.6	-	Phe-Lys
X1P3 (14-22)	0.74	-	Phe-Thr-Pro-Asp-Thr-Ile-Thr-Val-Ala
X1P4 (23-33)	1.25	0.28	Ala-Gly-Glu-Ala-Ile-Glu-Phe-Thr-Leu-Val-Gly
X1P4A (29-42)	0.70	-	Phe-Thr-Leu-Val-Gly-Glu-Thr-Gly-(His, Asn, Val, Val, Phe, Asp)
X1P5 (34-46)	0.62	-	Glu-Thr-Gly-His-Asn-Val-Val-Phe-Asp-Ile-Pro-Ala-Gly
X1P5A (43-46)	0	0.58	Ile-Pro-Ala-Gly
X1P5B (43-52)	0	-	Ile-Pro-Ala-Gly-Ala-Pro-(Gly, Pro, Val, Ala)
X1P6 (47-53)	0	0.41	Ala-Pro-Gly-Pro-Val-Ala-Ser
X1P7 (54-57)	1.10	-	Glu-Leu-Ser-Hse

Amino acid sequence of X-2 CNBr fragment

The sequence of the X-2 CNBr fragment from residues 58 to 87 was determined unequivocally from dansyl-Edman analysis of chymotryptic, tryptic and thermolysin peptides. The sequence of the region from residues 88 to 92, which is invariant in all higher plant plastocyanins so far sequenced (Appendix IV) was not firmly established apart from the terminal Met. Repeated attempts to find convincing sequence evidence from peptides failed (see Discussion). The only other notable point is that heterogeneity was found at residue 68.

Chymotryptic digest

15 mg of X-2 CNBr fragment were digested with chymotrypsin for 2 hours. The resulting peptides were separated by high voltage paper electrophoresis.

Peptides were isolated from the N-terminal region in good yields and sufficiently pure condition for accurate amino acid analysis. Peptides from the C-terminal region proved difficult to purify and were not isolated in large amounts.

Typical chymotryptic activity to the C-terminal side of Phe and Tyr was observed, but peptides were also formed from cleavage to the C-terminal side of Asp, Lys, Gly and His.

Chymotryptic peptides

Peptide X2C1 (58-78)

Asp-Gln-Asp-Asp-Leu-Leu-Ser-Gln-Asn-Glu-$\frac{Asp}{Glu}$-Asp-Phe-Thr-Ala-Lys-Val-Ser-Thr-Pro-Gly

Dansyl-Edman analysis was inconclusive after residue 73. Digestion with carboxypeptidase-A for 1 h followed by dansylation yielded dansyl-glycine. The remainder of the sequence was deduced from quantitative amino acid analysis data and comparison with thermolytic peptides. The positions of acidic residues were determined from the electrophoretic mobilities of samples of the peptide removed after ten, eleven and twelve Edman degradation steps and comparison with thermolytic peptides. Both aspartic acid and glutamic acid were identified at residue 68, in approximately equal proportions. Quantitative amino acid analysis confirmed the sequence data.

Peptide X2C1A (58-73)

Asp-Gln-Asp-Asp-Leu-Leu-Ser-Gln-Asn-Glu-$\frac{Asp}{Glu}$-Asp-Phe-Thr-Ala-Lys

Dansylation without hydrolysis after fifteen Edman degradation steps confirmed lysine as the C-terminal residue. Acidic residues were positioned using data from the electrophoretic mobilities of peptide X2C1 and thermolytic peptides. Amino acid analysis results were in good agreement with sequence data.

Peptide X2C1B (74-80)

Val-Ser-Thr-Pro-Gly-Thr-Tyr

Tyrosine was confirmed as the C-terminal amino acid by dansylation without hydrolysis following six Edman degradation steps. Proline was identified at residue 77 following 4 h hydrolysis only. Amino acid analysis results confirmed the sequence data.

Peptide X2C1C (58-70)

Asp-Gln-Asp-Asp-Leu-Leu-Ser-Gln-Asn-Glu-$\frac{\text{Asp}}{\text{Glu}}$-Asp-Phe

Phenylalanine was confirmed as the final residue by C-terminal analysis and by dansylation without hydrolysis following twelve Edman degradation steps. Verification of the sequence was obtained from amino acid analysis data. Acidic residues were placed by reference to peptide X2C1 and thermolytic peptides.

Peptide X2C1D (71-78)

Thr-Ala-Lys-Val-Ser-Thr-Pro-Gly

This peptide was isolated in a small yield. Glycine was confirmed as the C-terminal residue by dansylation without hydrolysis after seven Edman degradation steps. The electrophoretic mobility of the peptide at pH 1.9 corresponds with the sequence data, but the mobility at pH 6.5 is incorrect, as the peptide should have exhibited basic properties.

Peptide X2C2 (79-80)

Thr-Tyr

Dansylation without hydrolysis confirmed tyrosine as the C-terminal residue after one Edman degradation step.

Peptide X2C2A (79-92)

Thr-Tyr-Thr-Tyr-Tyr-Cys-Thr-Pro-His-Gln-Gly-Ala-Gly-Hse

This peptide was isolated in a small yield. Dansyl-Edman analysis was inconclusive after residue 83. Ala, Gly and Hse were tentatively identified by C-terminal analysis, the remaining sequence being determined by qualitative amino acid analysis, reference to other chymotryptic and thermolysin peptides and homology.

Peptide X2C3 (81-83)

Tyr-Tyr-Tyr

Tyrosine was confirmed as the C-terminal residue after two Edman degradation steps by dansylation without hydrolysis.

Peptide X2C3A (81-87)

Thr-Tyr-Tyr-Cys-Thr-Pro-His

Sequence determination was tenative for residue 84. Dansylation without hydrolysis confirmed His as the C-terminal amino acid.

Peptide X2C4 (84-92)

Cys-Thr-Pro-His-Gln-Gly-Ala-Gly-Hse

Sequence determination was inconclusive after residue 88.

C-terminal analysis tentatively ascribed glycine and homoserine as the C-terminal residues. The remaining sequence was determined from qualitative amino acid analysis, sequence data from other chymotryptic and thermolytic peptides and homology.

Peptide X2C4A (88-92)

Gln-Gly-Ala-Gly-Hse

This peptide was not isolated, but traces were found contaminating peptides X2C1B, X2C3 and X2C2A.

Table 11

Peptides from digestion of X2 CNBr fragment using Chymotrypsin

Peptide/Position	Mobility (pH 6.5)	Mobility (pH 1.9)	Dansyl-Edman Results
X2C1 (58-78)	streaked, 1.75?	-	Asp-Gln-Asp-Asp-Leu-Leu-Ser-Gln-Asn-Glu-Asp/Glu-Asp-Phe-Thr-Ala-Lys-(Val,Ser,Thr,Pro)-Gly
X2C1A (58-73)	streaked, 2.0?	-	Asp-Gln-Asp-Asp-Leu-Leu-Ser-Gln-Asn-Glu-Asp/Glu-Asp-Phe-Thr-Ala-Lys
X2C1B (74-80)	0	0.38	Val-Ser-Thr-Pro-Gly-Thr-Tyr
X2C1C (58-70)	streaked, over 2.0	-	Asp-Gln-Asp-Asp-Leu-Leu-Ser-Gln-Asn-Glu-Asp/Glu-Asp-Phe
X2C1D (71-78)	0	0.68	Thr-Ala-Lys-Val-Ser-Thr-Pro-Gly
X2C2 (79-80)	0	0.63	Thr-Tyr
X2C2A (79-92)	0	0.40	Thr-Tyr-Thr-Tyr-Tyr-(Cys,Thr,Pro,His,Gln,Gly)-Ala Gly Hse
X2C3 (81-83)	0	0.47	Thr-Tyr-Tyr
X2C3A (81-87)	0.2	-	Thr-Tyr-Tyr-Cys-Thr-Pro-His
X2C4 (84-92)	0.80	0.68	Cys-Thr-Pro-His-(Gln,Gly,Ala)-Gly-Hse
X2C4A (88-92)	0	0.44	Gln-Gly-Ala-Gly-Hse

Tryptic digest

22 mg of X-2 CNBr fragment were digested with trypsin for 2 hours. The resulting peptides were separated by high voltage paper electrophoresis.

Only two peptides were isolated in quantities sufficient for sequence determination, cleavage having occurred as expected at the Lys-Val bond between residues 73 and 74.

Tryptic Peptides

Peptide X2T1 (58-73)

Asp-Gln-Asp-Asp-Leu-Leu-Ser-Gln-Asn-Glu-$\frac{\text{Asp}}{\text{Glu}}$-Asp-Phe-Thr-Ala-Lys

Lysine was confirmed as the C-terminal residue by dansylation without hydrolysis following fifteen Edman degradation steps. The positions of the acidic residues were determined by reference to thermolysin and chymotryptic peptides.

Peptide X2T2 (74-92)

Val-Ser-Thr-Pro-Gly-Thr-Tyr-Thr-Tyr-Tyr-Cys-Thr-Pro-His-Gln-Gly-Ala-Gly-Hse

This peptide was isolated in a good quantity and sufficiently pure for accurate amino acid analysis. Digestion with carboxypeptidase A for 15 min followed by dansylation yielded dansyl-alanine, dansyl-glycine and dansyl-homoserine in approximately equal amounts. Dansyl-Edman analysis gave

good results to residue 83 after which no sequence analysis was possible. The remainder of the sequence was established from amino acid analysis data, comparison with chymotryptic and thermolytic peptides and homology.

Table 12

Peptides from digestion of X2 CNBr fragment using trypsin

Peptide/Position	Mobility (pH 6.5)	Mobility (pH 1.9)	Dansyl-Edman Results
X2T1 (58-73)	1.97	0.38	Asp→-Gln→-Asp→-Asp→-Leu→-Leu→-Ser→-Gln→-Asn→-Glu→-Asp/Glu→-Asp→-Phe→-Thr→-Ala→-Lys→
X2T2 (74-92)	0.49	-	Val→-Ser→-Thr→-Pro→-Gly→-Thr→-Tyr→-Thr→-Tyr→-Tyr→-(Cys,Thr,Pro,His,Gln,Gly)-←Ala-←Gly-←Hse

Thermolysin digest

11 mg of X-2 CNBr fragment were digested with thermolysin for 2 hours. The resulting peptides were separated by high voltage paper electrophoresis.

Few thermolytic peptides were isolated in sufficiently pure condition for an accurate amino acid analysis to be performed. No totally reliable information was obtained from the C-terminal region of the fragment. An attempt to further purify peptides by paper chromatography failed.

Thermolytic peptides

Peptide X2H1 (58-72)

Asp-Gln-Asp-Asp-Leu-Leu-Ser-Gln-Asn-Glu-$\overset{\text{Asp}}{\underset{\text{Glu}}{}}$-Asp-Phe-Thr-Ala

Alanine was confirmed as the C-terminal residue by dansylation without hydrolysis following fourteen Edman degradation steps. Amino acid analysis results were in good agreement with sequence data.

Peptide X2H1A (58-62)

Asp-Gln-Asp-Asp-Leu

Following four Edman degradation steps, dansylation without hydrolysis yielded dansyl-leucine. Samples were removed after each Edman step and the positions of the acidic residues determined from their mobilities during electrophoresis at pH 6.5.

Peptide X2H1B (62-69)

Leu-Leu-Ser-Gln-Asn-Glu-$\overset{Asp}{\underset{Glu}{}}$-Asp

Aspartic acid was identified as the C-terminal residue by dansylation without hydrolysis after seven Edman degradation steps. The positions of the acidic residues were determined from the electrophoretic mobility of the peptide at pH 6.5 together with samples of the peptide after four and five Edman degradation steps.

Peptide X2H1C (70-81)

Phe-Thr-Ala-Lys-Val-Ser-Thr-Pro-Gly-Thr-Tyr-Thr

A positive identification of all residues was achieved by dansyl-Edman analysis and the result confirmed by amino acid analysis.

Peptide X2H2 (73-79)

This peptide was not isolated

Peptide X2H3 (80-88)

Tyr-Thr-Tyr-Tyr-Cys-Thr-Pro-His-Gln

This peptide was not isolated in sufficiently pure condition for accurate amino acid analysis. Identification for accurate amino acid analysis. Identification of the final amino acid by dansylation without hydrolysis was tentative.

Peptide X2H4 (89-92)

Gly-Ala-Gly-Hse

This peptide was isolated in a very small quantity and only the first two residues were positively identified. Residues 91 and 92 were tentatively identified, and the peptide mobility during electrophoresis at pH 1.9 plus semi-quantitative amino acid analysis data supported these results.

Table 13

Peptides from digestion of X2 CNBr fragment using thermolysin

Peptide/Position	Mobility (pH 6.5)	Mobility (pH 1.9)	Dansyl-Edman Results
X2H1 (58-72)	-	0.21	Asp, Gln, Asp, Asp, Leu, Leu, Ser, Gln, Asn, Glu, Asp/Glu, Asp, Phe, Thr, Ala
X2H1A (58-62)	2.75	0.38	Asp, Gln, Asp, Asp, Leu
X2H1B (62-69)	2.0	0.31	Leu, Leu, Ser, Gln, Asn, Glu, Asp/Glu, Asp
X2H1C (70-81)	0.6	0.51	Phe, Thr, Ala, Lys, Val, Ser, Thr, Pro, Gly, Thr, Tyr, Thr
X2H3 (80-88)	0	0.51	Tyr, Thr, Tyr, Tyr, Cys, Thr, Pro, His, Gln
X2H4 (89-92)	0	0.56	Gly, Ala, Gly, Hse

Table 14

Determination of positions of amide residues in N-terminal region of X2 CNBr fragment

Peptide	Number of Edman degradations	Sequence	Mobility	Charge	Residues removed	Remaining sequence
X2H1B	0	Leu-Leu-Ser-Glx-Asx-Glx-$^{Asx}_{Glx}$-Asx	2.00	-3		
	-4	Asx-Glx-$^{Asn}_{Glx}$-Asx	3.20	-3	Gln	
	-5	Glx-$^{Asn}_{Glx}$-Asx	3.60	-3	Asn	Glu-$^{Asp}_{Glu}$-Asp
X2H1A	0	Asx-Glx-Asx-Asx-Leu	2.75	-3		
	-1	Glx-Asx-Asx-Leu	2.4	-2	Asp	
	-2	Asx-Asx-Leu	2.9	-2	Gln	Asp-Asp-Leu

Samples were removed at relevant stages of dansyl-Edman analysis of the peptide and their mobility during high-voltage paper electrophoresis at pH 6.5 measured.

The locations of amide residues were determined using the method of Offord (1966).

Table 15

Amino Acid Compositions of Peptides from Bracken Plastocyanin

	X1C1	X1C2	X1C3	X1C4	X1P5B	X2C1	X2C1A	X2C1B	X2C1C	X2T2	X2H1	X2H1C
Asp	1.4(1)	1.1(1)	1.4(1)	1.3(1)		5.9(5.5)	5.5(5.5)		5.4(5.5)		5.7(5.5)	
Thr		3.3(3)	1.7(2)			1.6(2)	0.8(1)	2.1(2)		4.1(4)	0.7(1)	3.4(4)
Ser	1.2(1)			1.6(2)		1.7(2)	1.1(1)	0.7(1)	1.0(1)	1.1(1)	1.3(1)	1.3(1)
Glu	2.0(2)	2.4(2)	1.3(1)	1.3(1)		3.6(3.5)	3.3(3.5)		3.2(3.5)	1.1(1)	3.2(3.5)	
Pro		(1)		3.0(3)	2.7(3)	.91(1)		0.6(1)		1.6(2)		1.1(1)
Gly	2.3(2)	1.3(1)	2.2(2)	2.1(2)	2.0(2)	1.0(1)		1.3(1)		2.5(3)		1.4(1)
Ala	1.0(1)	2.7(3)		2.6(3)	3.0(3)	1.3(1)	0.9(1)			0.9(1)	0.8(1)	1.3(1)
Cys										0.5(1)		
Val	2.7(3)	0.5(1)	3.0(3)	1.2(1)	1.2(1)	0.6(1)		0.9(1)		1.1(1)		1.4(1)
Hse				0.7(1)						0.6(1)		
Ile		1.5(2)		0.9(1)	0.7(1)							
Leu			1.6(1)	0.8(1)		2.0(2)	2.0(2)		1.9(2)		1.9(2)	
Tyr								0.6(1)		2.6(3)		0.6(1)
Phe	0.6(1)	2.0(2)	1.0(1)			1.1(1)	0.9(1)		0.8(1)		0.6(1)	0.6(1)
His			0.8(1)							1.3(1)		
Lys	0.6(1)	0.6(1)				0.9(1)	0.8(1)					1.1(1)

The composition values shown are molar ratios determined from 24 h hydrolysis of single samples. No correction was made for destruction of certain amino acids.

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

Tryptophan was not determined.

Amino acid sequence of X3 CNBr fragment

The sequence of the X3 CNBr fragment was established unequivocally by dansyl-Edman analysis on the intact peptide. The notable feature is the presence of tryptophan as the C-terminal residue.

Peptide X3 (93-99)

Lys-Gly-Thr-Lys-Ile-Val-Trp

This fragment was further purified by high-voltage paper electrophoresis at pH 6.5. The peptide gave a yellow-orange colour with Ehrlich reagent, indicating the presence of chemically degraded tryptophan. No tryptophan was identified during routine quantitative and semi-quantitative amino acid analyses but its presence was confirmed when the hydrolysis time was reduced to 1 hour and 2% thioglycollic acid was added to the hydrolysate.

Following six Edman degradation steps, tryptophan was identified as the C-terminal amino acid by dansylation without hydrolysis. C-terminal analyses of both native protein and X3 fragment confirmed Lys-Ile-Val-Trp as the final residues.

Table 16

Semi-quantitative amino acid analyses of X3 CNBr fragment*

amino acid/ dipeptide	16 h hydrolysis in 5.7 M HCl	1 h hydrolysis in 5.7 M HCl plus 2% thioglycollic acid
Val	++	?
Ile	++	?
**Ile-Val	++	+++
Bis-Lys	++++	+++
Gly	++	++
Thr	++	++
Trp		++

*Liberated amino acids determined by dansyl method

**ran approximately in dansyl-phenylalanine position

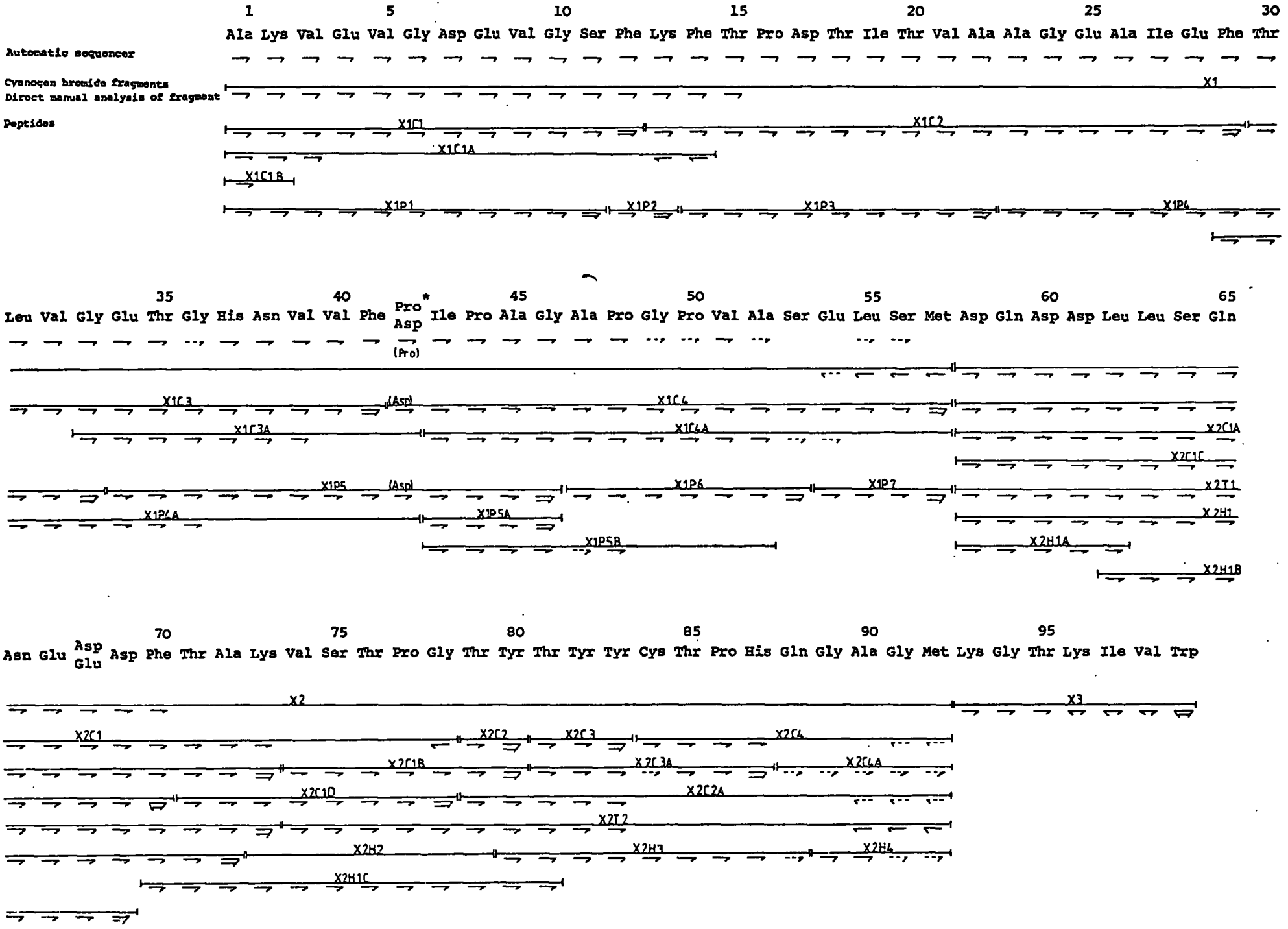
+ to ++++ degree of intensity of dansyl derivatives.

FIGURE 12

The complete sequence of plastocyanin from Pteridium aquilinum

Complete arrows (⟶) indicate amino acids confirmed by dansyl-Edman analysis or identification of PTH derivatives from the automatic sequencer and dotted arrows (⇢) indicate tentative identification. Reversed arrows (⟵) indicate positions confirmed by carboxypeptidase A followed by dansylation, and double arrows (⟹) indicate confirmation of C-terminal residue by identification of the free amino acid by dansylation. Other residues were determined by peptide composition and the order determined by other evidence.

* Heterogeneity at position 42. Pro was identified unambiguously from automatic sequence data (see Appendix II). Asp was identified by dansyl-Edman analysis of peptides obtained from proteolytic digestion of the X 1 fragment.


1 5 10 15 20 25 30
Ala Lys Val Glu Val Gly Asp Glu Val Gly Ser Phe Lys Phe Thr Pro Asp Thr Ile Thr Val Ala Ala Gly Glu Ala Ile Glu Phe Thr
Automatic sequencer
Cyanogen bromide fragments
Direct manual analysis of fragment
Peptides
X1
X1C1
X1C2
X1C1A
X1C1B
X1P1
X1P2
X1P3
X1P4
35 40 45 50 55 60 65
Leu Val Gly Glu Thr Gly His Asn Val Val Phe Pro* Asp Ile Pro Ala Gly Ala Pro Gly Pro Val Ala Ser Glu Leu Ser Met Asp Gln Asp Asp Leu Leu Ser Gln
(Pro)
X1C3
(Asp)
X1C4
X1C3A
X1C4A
X2C1A
X2C1C
X1P5
(Asp)
X1P6
X1P7
X2T1
X1P6A
X1P5A
X2H1
X1P5B
X2H1A
X2H1B
70 75 80 85 90 95
Asn Glu Asp Glu Asp Phe Thr Ala Lys Val Ser Thr Pro Gly Thr Tyr Thr Tyr Tyr Cys Thr Pro His Gln Gly Ala Gly Met Lys Gly Thr Lys Ile Val Trp
X2
X3
X2C1
X2C2
X2C3
X2C4
X2C1B
X2C3A
X2C4A
X2C1D
X2C2A
X2T2
X2H2
X2H3
X2H4
X2H1C

N-Terminal sequence determination

The N-terminal sequences of plastocyanin from pteridophytes and gymnosperms are shown in Figure 13. No real difficulties were encountered in obtaining these sequences by means of the spinning cup sequenator. The criteria used for the identification of the residues were those established by Haslett and Boulter (1976). In general the higher ferns proved the most difficult to sequence because of a tendency to be washed out of the cup more than was the case with the other plastocyanins, thus making identification of the final residues more difficult. Consequently, unambiguous identification was not possible at position 36 in the sequences of plastocyanin from Blechnum spicant and Dryopteris filix-mas, as both Pro and Gly can be difficult to identify positively because of low yields (Haslett and Boulter, 1976). The detailed results of each sequence (i.e. analyses of each peptide) are given in Appendix II.

FIGURE 13

N-terminal amino acid sequence data

The sequences were determined using the Beckman automatic spinning cup sequencer. The procedure for identification of the resulting PTH derivatives is given in Methods.

Unambiguous identification of the residue occurring at position 36 was not possible in the sequences of plastocyanin from Blechnum spicant and Dryopteris filix-mas.

	5	10	15	20	25	30	35	40
Equisetum sylvaticum	Ala Glu Val Ile Met Gly Leu Asp Asp Gly Ser Leu Lys Phe Leu Pro Asp Lys Val Glu Val Lys Ala Gly Glu Lys Ile Thr Asn Val Asn Phe Ala Gly Phe Pro His Asn Val Val							
Equisetum arvense	Ala Glu Val Ile Met Gly Leu Asp Asp Gly Ser Leu Gln Phe Asn Pro Lys Glu Val Val Val Lys Ala Gly Glu Lys Ile Thr Phe Ile Asn Asn Ala Gly Phe Pro His Asn Val Val							
Osmunda regalis	Ala Asp Val Ile Met Gly Gly Asp Asp Gly Ser Leu Ala Phe Ile Pro Asn Lys Ile Val Val Ser Val Gly Glu Pro Ile Thr Phe Lys Asn Asn Ala Gly Phe Pro His Asn Val Val							
Pteridium aquilinum	Ala Lys Val Glu Val Gly Asp Glu Val Gly Ser Phe Lys Phe Thr Pro Asp Thr Ile Thr Val Ala Ala Gly Glu Ala Ile Glu Phe Thr Leu Val Gly Glu Thr Gly His Asn Val Val							
Blechnum spicant	Ala Lys Val Glu Val Gly Asp Glu Val Gly Asn Phe Lys Phe Tyr Pro Glu Thr Ile Thr Val Ala Ala Gly Glu Ala Val Glu Phe Thr Leu Val Gly Glu Thr Pro? Gly? His Asn Ile Val							
Dryopteris filix-mas	Ala Lys Val Glu Val Gly Asp Glu Val Gly Asn Phe Lys Phe Tyr Pro Asp Ser Ile Thr Val Ser Ala Gly Glu Ala Val Glu Phe Thr Leu Val Gly Glu Thr Pro? Gly? His Asn Ile Val							
Dryopteris dilatata	Ala Lys Val Glu Val Gly Asp Glu Val Gly Asn Phe Lys Phe Tyr Pro Glu Ala Ile Thr Val Ala Ala Gly Glu Ser Ile Glu Phe Thr Leu Val Gly Glu Thr Gly His Asn Ile Val							
Cycas revoluta	Ile Glu Val Leu Leu Gly Gly Asn Gly Gly Glu Leu Ala Phe Ile Pro Asp Lys Phe Glu Val Ala Pro Gly Glu Glu Ile Val Phe Lys Asn Asn Ala Ala Phe Pro His Asn Val Val							
Araucaria araucana	Ala Glu Val Leu Met Gly Gly Asn Gly Gly Glu Leu Ala Phe Ile Pro Ser Glu Phe Ser Val Ala Pro Gly Glu Thr Ile Thr Phe Lys Asn Asn Ala Gly Phe Pro His Asn Val Ile							
Taxus baccata	Leu Glu Val Leu Met Gly Gly Asn Gly Gly Glu Leu Val Phe Ile Pro Ser Glu Phe Gln Leu Thr Ala Gly Asp Thr Ile Val Phe Lys Asn Asn Ala Gly Phe Pro His Asn Val Val							

FIGURE 14

N-terminal amino acid sequence data

N-terminal amino acid sequences from lower plants compared with the sequences of two green algae (*Enteromorpha intestinalis* and *Chlorella fusca*). The additional N-terminal residues found in the algal sequences have been omitted. Residues invariant in all higher plant sequences so far determined are also included for reference.

	5	10	15	20	25	30	35	40
Equisetum sylvaticum	Ala Glu Val Ile Met Gly Leu Asp Asp Gly Ser Leu Lys Phe Leu Pro Asp Lys Val Glu Val Lys Ala Gly Glu Lys Ile Thr Asn Val Asn Phe Ala Gly Phe Pro His Asn Val Val							
Equisetum arvense	Ala Glu Val Ile Met Gly Leu Asp Asp Gly Ser Leu Gln Phe Asn Pro Lys Glu Val Val Val Lys Ala Gly Glu Lys Ile Thr Phe Ile Asn Asn Ala Gly Phe Pro His Asn Val Val							
Osmunda regalis	Ala Asp Val Ile Met Gly Gly Asp Asp Gly Ser Leu Ala Phe Ile Pro Asn Lys Ile Val Val Ser Val Gly Glu Pro Ile Thr Phe Lys Asn Asn Ala Gly Phe Pro His Asn Val Val							
Pteridium aquilinum	Ala Lys Val Glu Val Gly Asp Glu Val Gly Ser Phe Lys Phe Thr Pro Asp Thr Ile Thr Val Ala Ala Gly Glu Ala Ile Glu Phe Thr Leu Val Gly Glu Thr Gly His Asn Val Val							
Blechnum spicant	Ala Lys Val Glu Val Gly Asp Glu Val Gly Asn Phe Lys Phe Tyr Pro Glu Thr Ile Thr Val Ala Ala Gly Glu Ala Val Glu Phe Thr Leu Val Gly Glu Thr	Pro? Gly?	His Asn Ile Val					
Dryopteris filix-mas	Ala Lys Val Glu Val Gly Asp Glu Val Gly Asn Phe Lys Phe Tyr Pro Asp Ser Ile Thr Val Ser Ala Gly Glu Ala Val Glu Phe Thr Leu Val Gly Glu Thr	Pro? Gly?	His Asn Ile Val					
Dryopteris dilatata	Ala Lys Val Glu Val Gly Asp Glu Val Gly Asn Phe Lys Phe Tyr Pro Glu Ala Ile Thr Val Ala Ala Gly Glu Ser Ile Glu Phe Thr Leu Val Gly Glu Thr Gly His Asn Ile Val							
Cycas revoluta	Ile Glu Val Leu Leu Gly Gly Asn Gly Gly Glu Leu Ala Phe Ile Pro Asp Lys Phe Glu Val Ala Pro Gly Glu Glu Ile Val Phe Lys Asn Asn Ala Ala Phe Pro His Asn Val Val							
Araucaria araucana	Ala Glu Val Leu Met Gly Gly Asn Gly Gly Glu Leu Ala Phe Ile Pro Ser Glu Phe Ser Val Ala Pro Gly Glu Thr Ile Thr Phe Lys Asn Asn Ala Gly Phe Pro His Asn Val Ile							
Taxus baccata	Leu Glu Val Leu Met Gly Gly Asn Gly Gly Glu Leu Val Phe Ile Pro Ser Glu Phe Gln Leu Thr Ala Gly Asp Thr Ile Val Phe Lys Asn Asn Ala Gly Phe Pro His Asn Val Val							
Algae {	Gln Leu Val Glu Leu Gly Gly Asp Asp Gly Ser Leu Ala Phe Val Pro Ser Lys Val Ser Val Ala Ala Gly Glu Ser Ile Glu Phe Ile Asn Asn Ala Gly Phe Pro His Asn Ile Val							
Algae }	Val Thr Val Lys Leu Gly Ala Asp Ser Gly Ala Leu Val Phe Glu Pro Ser Ser Val Thr Ile Lys Ala Gly Glu Thr Val Thr Trp Val Asn Asn Ala Gly Phe Pro His Asn Ile Val							
Higher plants	Leu Gly	Asp Gly	Leu	Phe	Pro	Phe	Gly	Ile	Phe Lys Asn Asn Ala Gly Phe Pro His Asn	Val
(invariant residues only)								

Computer Analysis of the N-terminal data

Figures 15 and 16 show the two alternative topologies (showing maximum parsimony) produced as a result of computer analysis of the N-terminal sequences of plastocyanin from pteridophytes and gymnosperms shown in Figure 13.

Figures 17 and 18 show the two alternative topologies (showing maximum parsimony) produced using the data shown in Figure 14, which includes the sequences from pteridophytes and gymnosperms plus two green algae.

Figure 19 summarises the alternative topologies (showing maximum parsimony) produced if the N-terminal sequences of two angiosperms (*Bellis perennis* and Magnolia) are included in the data set.

Figure 20 summarises alternative topologies produced using the data in Figure 14 which need only one amino acid substitution more than the minimum to explain their construction.

FIGURE 15

Phylogenetic relationships of pteridophytes and gymnosperms

A computer-generated phylogenetic tree, constructed by the exhaustive search and ancestral sequence methods, relating the sequences shown in Figure 13.

Minimum number of amino acid substitutions needed to explain this topology = 67.

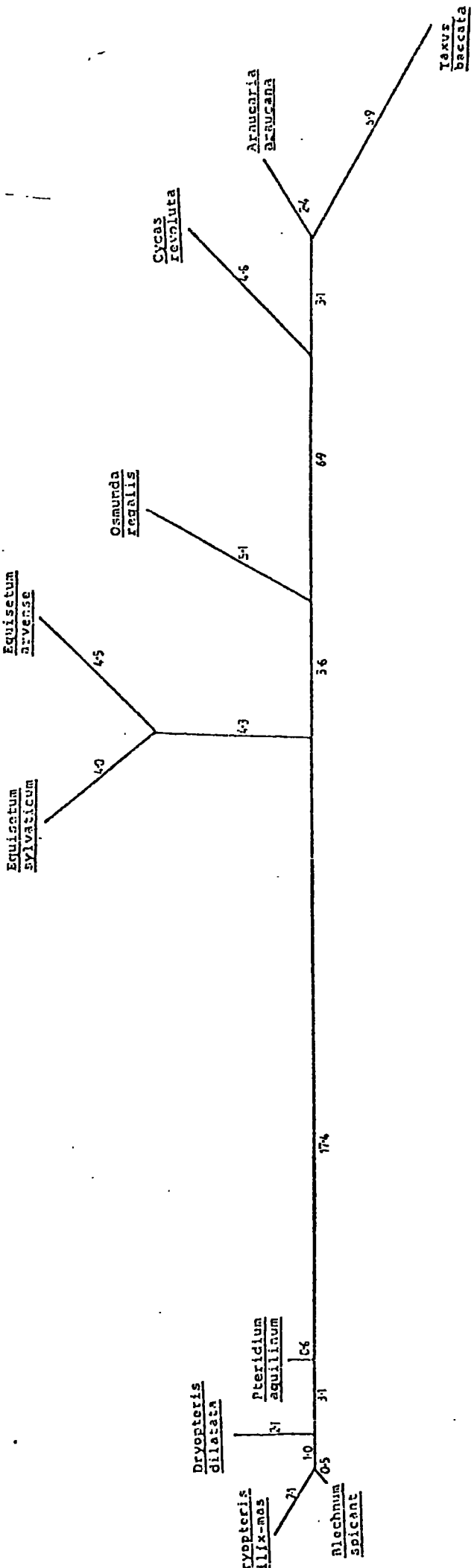

Taxus baccata
Araucaria araucana
Cycas revoluta
Osmunda regalis
Equisetum arvense
Equisetum sylvaticum
Pteridium aquilinum
Dryopteris dilatata
Dryopteris filix-mas
Blechnum spicant
3·9
2·4
4·6
5·1
4·5
4·0
4·3
3·6
6·9
3·1
17·4
3·1
0·6
3·1
1·0
0·5
2·1

FIGURE 16

Phylogenetic relationships of pteridophytes and gymnosperms

A computer-generated phylogenetic tree, constructed by the exhaustive search and ancestral sequence methods, relating the sequences shown in Figure 13.

This is the only topology produced as an alternative to that of Figure 15 which requires the same minimum number of amino acid substitutions (= 67) to explain its configuration.

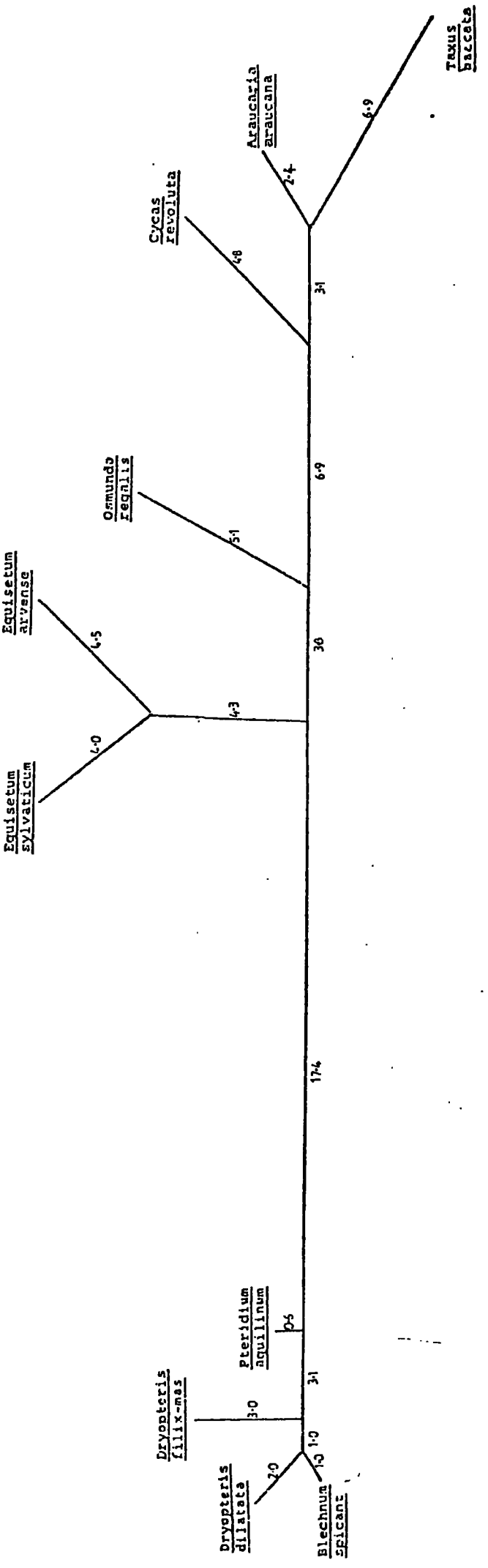

Taxus baccata
Araucaria araucana
Cycas revoluta
Osmunda regalis
Equisetum arvense
Equisetum sylvaticum
Pteridium aquilinum
Dryopteris filix-mas
Dryopteris dilatata
Blechnum spicant
6·9
2·4
4·8
5·1
4·5
4·0
4·3
0·5
3·0
2·0
1·0
3·1
17·4
3·0
6·9
3·1

FIGURE 17

Phylogenetic relationships of pteridophytes, gymnosperms and green algae

A computer-generated phylogenetic tree, constructed by the exhaustive search and ancestral sequence methods, relating the sequences shown in Figure 14.

Minimum number of amino acid substitutions needed to explain this topology = 74.

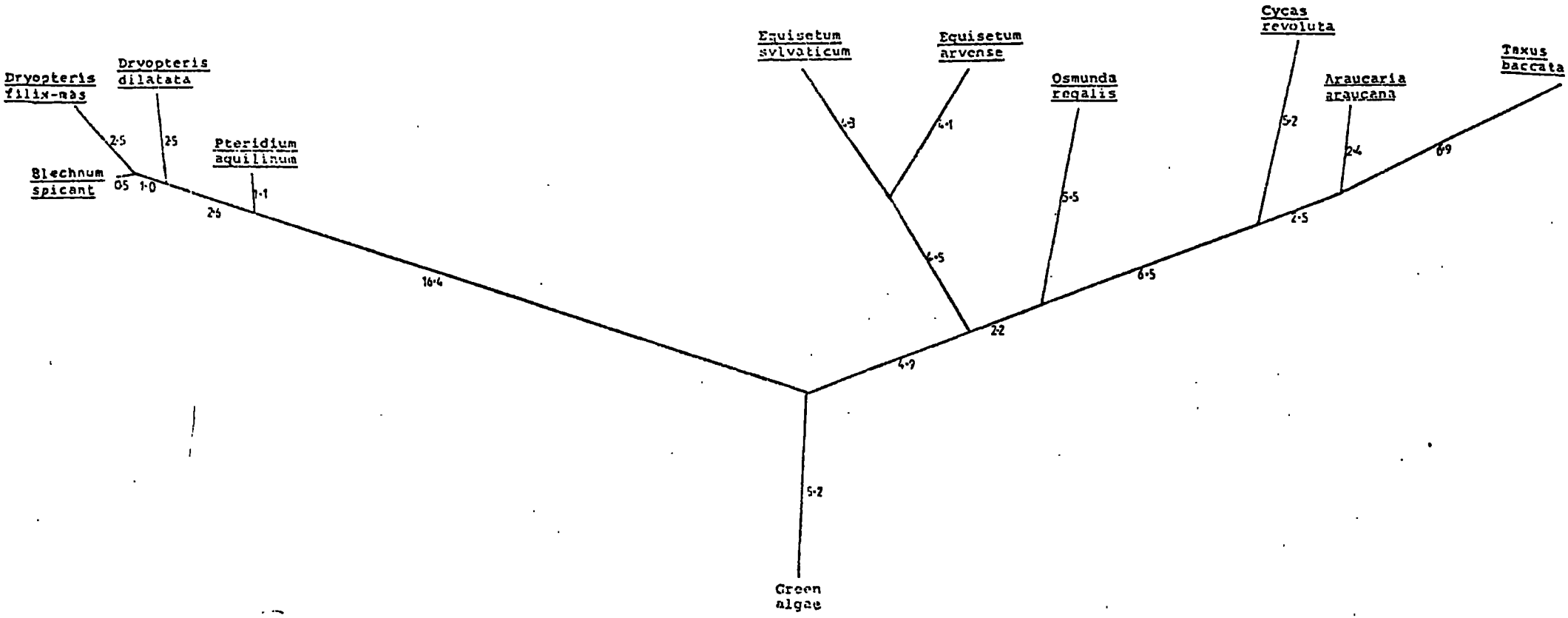

Dryopteris filix-mas
Dryopteris dilatata
Blechnum spicant
Pteridium aquilinum
Equisetum sylvaticum
Equisetum arvense
Osmunda regalis
Cycas revoluta
Araucaria araucana
Taxus baccata
Green algae
2·5
2·5
0·5
1·0
2·6
7·1
16·4
4·3
4·1
4·5
2·2
4·9
5·2
5·5
6·5
5·2
2·4
6·9
2·5
5·2

FIGURE 18

Phylogenetic relationships of pteridophytes, gymnosperms and green algae

A computer-generated phylogenetic tree, constructed by the exhaustive search and ancestral sequence methods, relating the sequences shown in Figure 14.

This is the only topology produced as an alternative to that of Figure 17 which requires the same minimum number of amino acid substitutions (= 74) to explain its configuration.

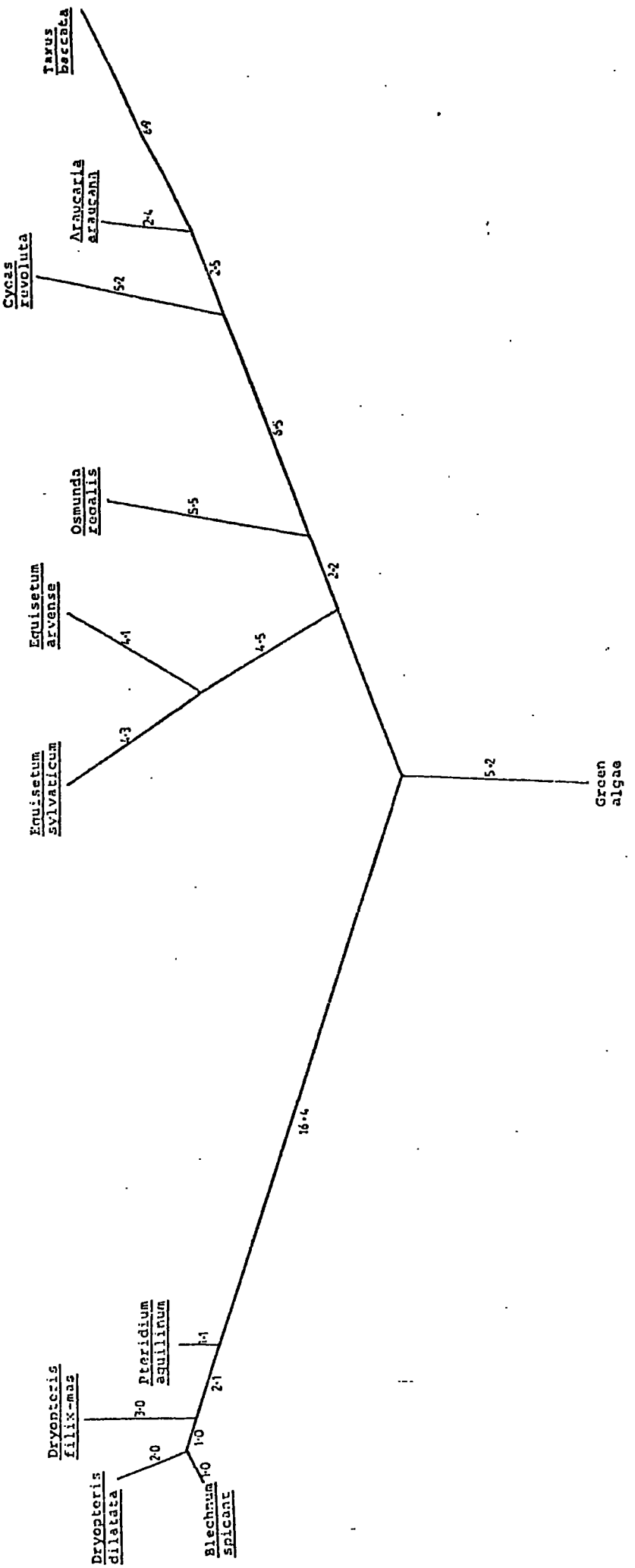
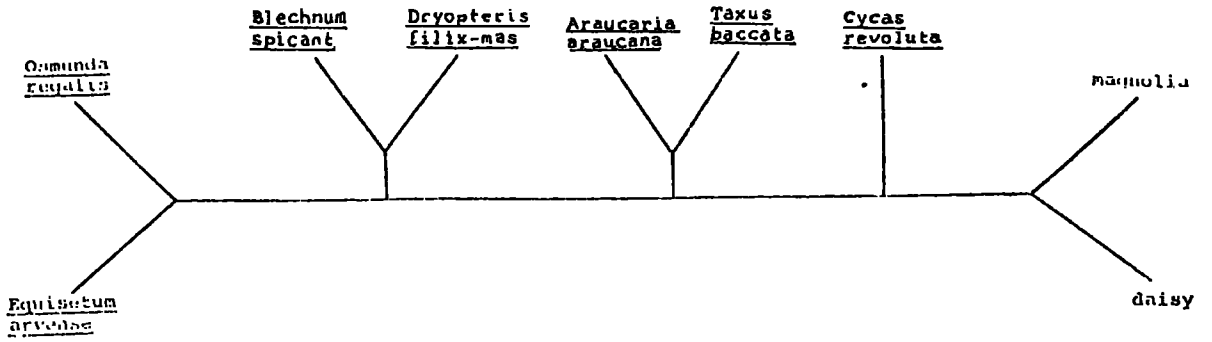

Taxus baccata
6·9
Araucaria araucana
2·4
2·5
Cycas revoluta
5·2
3·5
Osmunda regalis
5·5
2·2
Equisetum arvense
4·1
4·5
Equisetum sylvaticum
4·3
5·2
Green algae
16·4
2·1
1·1
Pteridium aquilinum
3·0
1·0
Dryopteris filix-mas
2·0
Dryopteris dilatata
1·0
Blechnum spicant

FIGURE 19

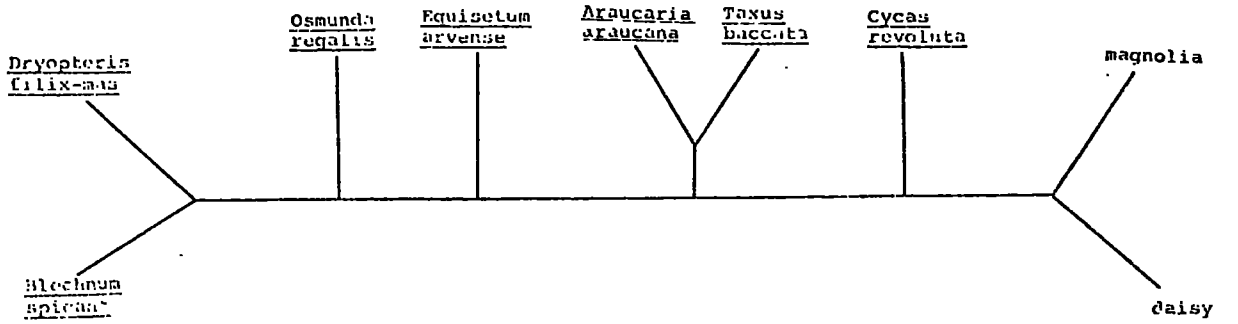
Angiosperm origins

The five alternative topologies which showed maximum parismony produced when two angiosperm sequences (magnolia and daisy) were included in the N-terminal sequence data set for computer analysis.

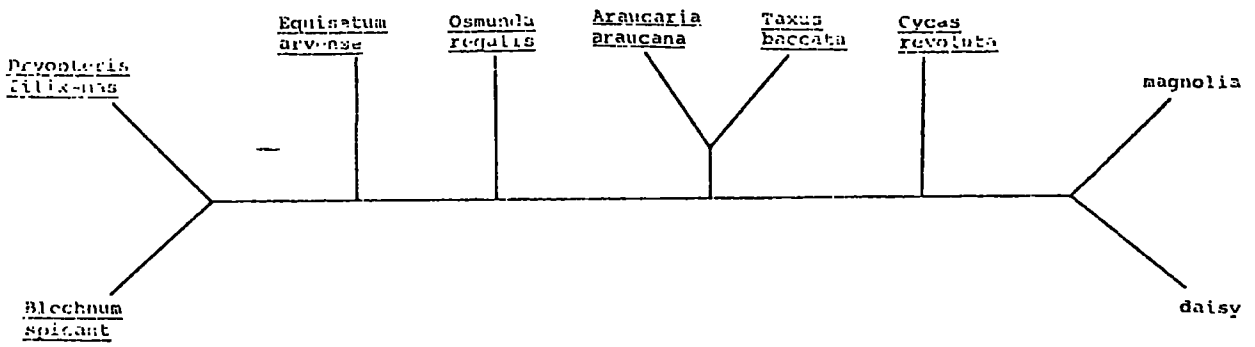
A


Blechnum spicant
Dryopteris filix-mas
Araucaria araucana
Taxus baccata
Cycas revoluta
Osmunda regalis
magnolia
Equisetum arvense
daisy

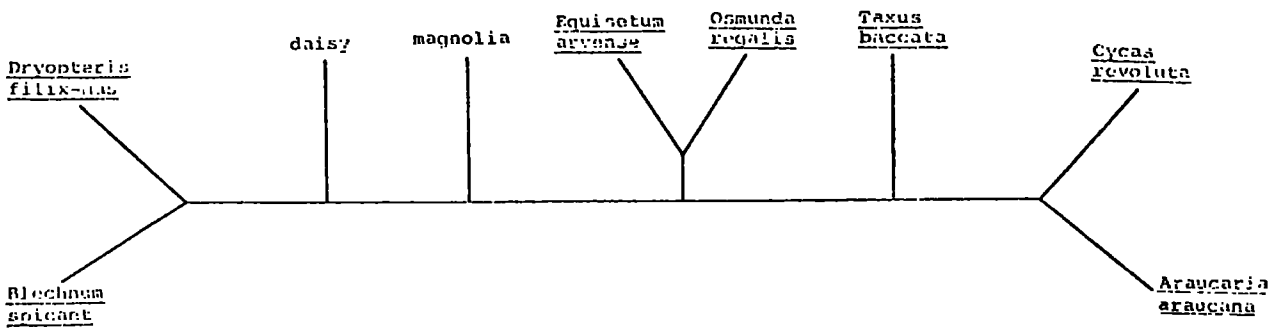
B


Osmunda regalis
Equisetum arvense
Araucaria araucana
Taxus baccata
Cycas revoluta
Dryopteris filix-mas
magnolia
Blechnum spicant
daisy

C


Equisetum arvense
Osmunda regalis
Araucaria araucana
Taxus baccata
Cycas revoluta
Dryopteris filix-mas
magnolia
Blechnum spicant
daisy

D


daisy
magnolia
Equisetum arvense
Osmunda regalis
Taxus baccata
Dryopteris filix-mas
Cycas revoluta
Blechnum spicant
Araucaria araucana

E

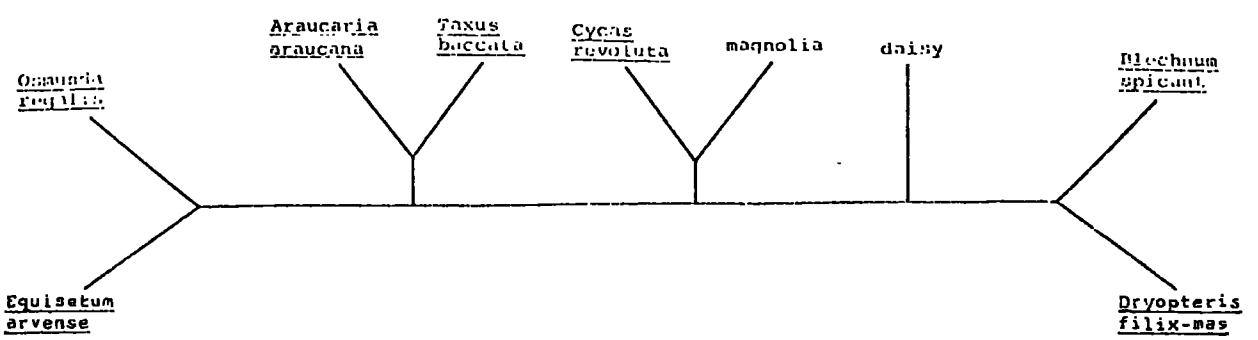
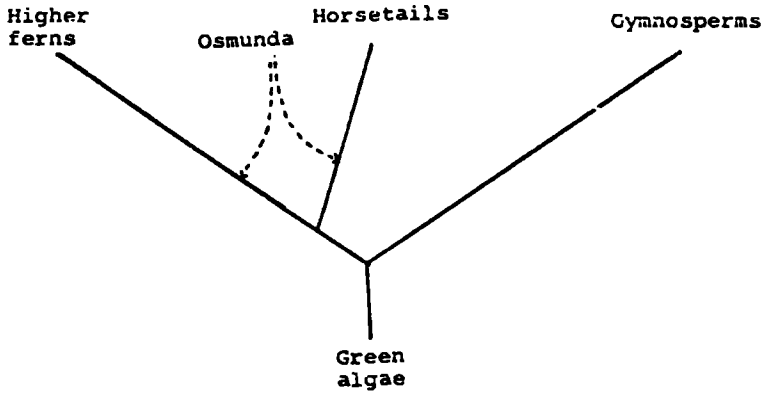
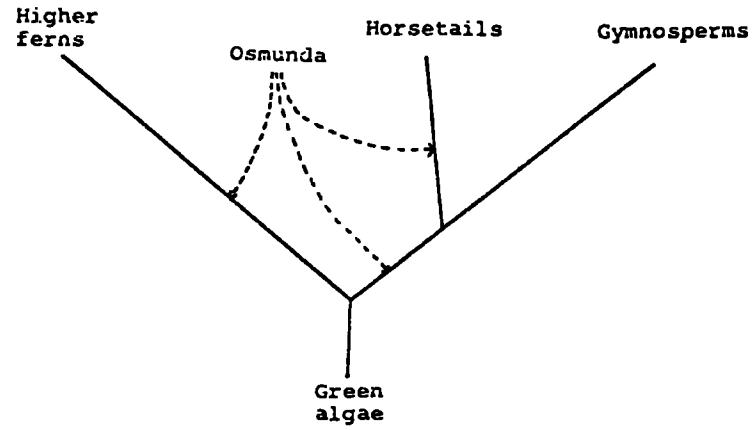

Araucaria araucana
Taxus baccata
Cycas revoluta
magnolia
daisy
Osmunda regalis
Blechnum spicant
Equisetum arvense
Dryopteris filix-mas

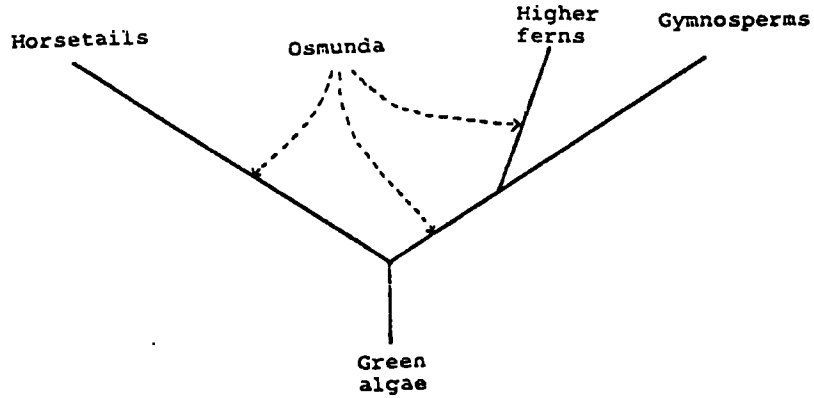
FIGURE 20

Possible alternative positions of Osmunda

A summary of the alternative topologies produced which all needed a minimum number of 75 amino acid substitutions to explain their configuration.


Higher ferns
Osmunda
Horsetails
Gymnosperms
Green algae


Higher ferns
Osmunda
Horsetails
Gymnosperms
Green algae


Horsetails
Osmunda
Higher ferns
Gymnosperms
Green algae

DISCUSSION

Choice of Materials

Many of the materials used for extracting proteins for amino acid sequence studies have been chosen for ease of supply rather than because of their potential phylogenetic significance. In this investigation, the possible phylogenetic importance was the first criterion of choice but even so, selection of material proved difficult. Although initially it was hoped to obtain data from species ranging from the most 'primitive' bryophytes to the most 'advanced' gymnosperms, in practice the range of suitable material was much narrower.

Bryophytes posed especially difficult problems. They are found in relatively inaccessible sites, and tend not to grow in sufficient quantities for collection of 1 to 3 Kg batches of such tiny plants not to significantly damage the ecosystem (Watson, 1968). Once collected, the material had to be carefully cleaned and sorted for two reasons. The fresh green plant material (considered best for plastocyanin extraction) is virtually inseparable from the older, dark green material when collecting in the field. This was more easily accomplished in the laboratory. In addition, liverworts usually grow on a sandy or stony substratum which was difficult to remove from the rhizoids. Nevertheless, despite such a thorough collection procedure

the percentage of fresh, actively photosynthesising material in the batches of plants which were finally extracted was probably still quite low. Consequently, collection and preparation of bryophyte material for extraction was extremely inefficient in terms of the time taken to obtain such small quantities of starting material. With the exception of Polytrichum spp. it is difficult to conclude whether the failure to extract plastocyanin was due to a fault in the extraction procedure or insufficient quantities of suitable starting material.

Pteridophytes proved the easiest group of plants to extract successfully. Although the majority of species are found in tropical and sub-tropical regions (Sporne, 1974) and of the British species a few prefer the milder and wetter west coast climate (e.g. Phyllitis scolopendrium), those that do occur are plentiful, easily identified and provide very few extraction problems (Equisetum spp., Blechnum spicant, Dryoptens spp., Pteridium aquilinum). Two exceptions are Osmunda regalis and Asplenium filix-femina. The latter tends to occur as single plants rather than in clumps and is difficult to distinguish from Dryopteris dilatata. Osmunda regalis is rare so supplies were limited. However one plant supplied on average 1 Kg of fresh green fronds without causing severe damage, which gave a good yield of plastocyanin,

so extraction was still feasible. Preparation for extraction was unnecessary, or in the case of some of the Pteropsida, limited to removal of the tough mid-rib of the fronds - a very quick process.

Gymnosperms presented more extraction problems than pteridophytes. Again the majority of species in this group are not native (Sporne, 1967), but some can be found in gardens (Araucaria araucana, Ginkgo biloba, Cupressus lawsonia) and others are plentiful in forestry plantations (Abies grandis, Picea abies, Larix decidua). Most species are found in tree form so that availability of material did not initially appear to be a problem. Preparation for extraction was often time consuming. The woody stems were removed if possible (Taxus baccata, Araucaria araucana), or otherwise cut into short pieces (Picea abies, Cupressus lawsonia) but in the majority of cases, very little or no plastocyanin was extracted. The appearance of gymnosperm foliage could not be relied upon to give a good indication of the potential yield of plastocyanin. For example, the fresh green leaves of Larix decidua and Ginkgo biloba yielded no plastocyanin, whereas the dull, dark green leaves of Cycas revoluta, Araucaria araucana and Taxus baccata yielded sufficient plastocyanin to allow purification. However, batches of plant material containing fresh, bright green

leaves of the new years growth of Taxus baccata gave a better yield than batches containing mostly older leaves, as would be expected. All species from which plastocyanin was finally obtained were collected locally except for Cycas revoluta and 1 Kg of Osmunda regalis fronds. The living plant of Cyas revoluta was sent complete and very little time lost in transportation. The Osmunda regalis fronds were extracted the day after collection and only a small loss in yield was observed compared to fronds collected locally on the day of extraction.

It was generally felt that extraction of plastocyanin from plants collected from non-local sources was not feasible. The unavoidable delay between collection and extraction coupled with the small yields of plastocyanin obtained from most pteridophytes and gymnosperms meant that any deterioration in the protein content of the plant material so caused would result in non-detection of the plastocyanin. During collection, if several batches of one species were required they were as far as possible collected from the same plant or the same locality to ensure that any heterogeneity which may be found in the sequence was due to polymorphism within the plant or community as opposed to intercommunity variation or even misidentification. Normally a test extraction was performed on a small pilot batch of

plant material, no more than 1 Kg to assess the yield before proceeding with further extractions.

The plastocyanins from which the sequence data were obtained were therefore extracted from particular species not solely because of their significance in the classical picture of evolution, but also because of their availability and convenience of collection or relative ease with which plastocyanin could be extracted from them.

Extraction and Purification

The general extraction procedure followed was that of Plesnicar and Bendall (1970). At all times the extraction buffer was kept as cold as was practicable, usually below $4^{o}C$ and the length of time that the crude extract was in contact with acetone (especially 80%) was kept to a minimum. Both precautions were necessary to prevent denaturation, precipitation of the protein or formation of apoplastocyanin.

In general, fresh material was preferred if available, as in certain cases storage at $-20^{o}C$ resulted in a much lower yield (or even no plastocyanin) as compared with fresh material. A possible reason for this could be the rupturing of cell membranes during freezing and consequent release of enzymes causing autodigestion or denaturation of the plastocyanin within the cell. In other cases, if the plant material was very tough, foliage collected the day

prior to extraction and frozen overnight was preferable. Tough stems and leaves caused the temperature of the blend to rise, in extreme cases as high as 20 to 30°C. Freezing therefore helped to keep the temperature of the homogenate below 4°C and occasionally had the effect of softening the material itself (e.g. Taxus baccata). When freezing material, the foliage was well spread within the sack to prevent possible bacterial decay in the centre.

During the initial homogenisation a great deal of gelatinous material was produced, possibly mucus, carbohydrate or polysaccharides (Stahmann, 1963). This occurred especially at the beginning of the growing season and did not appear to have any connection with growing conditions such as large or small amounts of rainfall or sunshine, variations in temperature or growing sites. Time did not permit a thorough investigation into the nature and causes of appearance of these contaminants, but it was found that extraction in 40 or 50% acetone blending buffer precipitated and so removed them. In each case the lowest possible concentration of blending buffer was used for reasons of safety and economy. Removal of these contaminants at this stage was important as no other such thorough method was found of eliminating them - otherwise the extraction and purification steps which followed were complicated by the presence of such sticky materials; the

precipitate formed in 80% acetone buffer would not settle and columns ran slowly or even stopped because the resin was blocked. Occasionally the liquid applied to the first DE23 Sephadex column was viscous, despite the use of stronger blending buffer. In such cases the sample was applied to the resin batchwise. This did not appreciably affect the yield of crude plastocyanin although this method of application of the sample to the resin meant that the protein was not eluted in a single band, but in a larger volume and consequently contained more impurities. This in turn meant that more purification procedures were needed, leading to an eventual loss in yield of pure plastocyanin.

The amount of material used for each extraction varied with the texture of the plant. On average 1 Kg was blended in 3L of buffer so that the maximum amount of plant material conveniently handled in one day was 3 Kg; with the poor yields of plastocyanin obtained from most species this meant that the crude plastocyanin had to be stored at -20°C until a sufficient quantity had been extracted for purification. Storage at -20°C did not make any appreciable difference to the plastocyanin except in the case of Ephedra spp., when no plastocyanin was detected in the crude extract after only three days. This was presumably due to formation of apoplastocyanin, but there was no apparent

reason why this should have happened to the protein from only the one species.

Following the initial extraction procedure, purification of the crude plastocyanin was not attempted until 30-40 mg had been collected. All purification steps were accomplished as quickly as possible, and the material was stored at -20^{o}C between steps, if necessary, to prevent growth of bacteria. Generally the time taken for purification was a maximum of 48 h at 4^{o}C. A second ion-exchange step, using DE23 Sephadex, was found to be extremely useful as an initial purification stage. Several litres of crude plastocyanin could be concentrated and purified in one step, so removing the need for a lengthy concentration procedure and eventually leading to a larger yield of a purer product. On average 100-200 ml of crude plastocyanin was collected from this column and the following concentration step reduced to 1-3 h (as opposed to 12-36 h without the second ion-exchange step). In addition a great deal of polyphenolic material was normally removed, so that the resulting concentrate was also much less contaminated.

The literature reports that the repeated use of ion-exchange chromatography only for purifying plastocyanin samples results in a lower yield than if gel filtration steps are included also (Borchert and Wessels, 1970;

Scawen et al., 1975). In this study it was considered most suitable to purify the protein initially by gel filtration on Sephadex G75 followed by a final ion-exchange separation by gradient elution from DEAE Sephadex to remove deamidation products and apoplastocyanin which would not be removed during gel filtration (Ramshaw et al., 1973). The plastocyanin was applied to the DEAE Sephadex in reduced form, to prevent the elution of two separate peaks of reduced and oxidised plastocyanin. Apart from the viscous materials already mentioned, the other major contaminants of most samples were the brown materials, thought to be polyphenolic in nature (Pirie, 1959; Loomis and Battaile, 1966). These behave similarly to plastocyanin during ion-exchange chromatography and gel filtration and were only eliminated by repeated use of these steps (Borchert and Wessels, 1970; Milne and Wells, 1970; Scawen et al., 1975). Contaminants other than polyphenols removed during the purification procedures included any mucus-type materials not removed in the extraction, bound ferricyanide originating from the potassium ferricyanide added during earlier procedures (Graziani et al., 1974) plus apoprotein and deamidation products formed during the initial extraction.

Loss in yield was inevitable during all purification steps, and the precautions outlined previously were considered

all that were practicable to keep these losses to a minimum. Nevertheless, only approximately half the original quantity of crude plastocyanin was usually obtained as pure protein. As with all stages of the extraction, no general rule was found, but it was felt that the higher the yield of plastocyanin per Kg of starting material, the purer would be the crude extract and hence easier to purify. This hypothesis was not substantiated by all the results. For example, 22.8 mg/kg of crude plastocyanin were extracted from Cycas revoluta, yet only about half this was recovered as pure protein. Occasionally, a much higher yield of the pure product was obtained than expected (e.g. Araucaria araucana and Osmunda regalis). This was probably not due to fewer purification losses, but rather because the initial estimate of crude plastocyanin was low, due to partial obliteration of the absorbance peak at 597 nm during spectrophotometry by the absorbance of contaminating materials around the same wavelength.

There is no obvious reason for the high number of poor yields or total failure to find plastocyanin in the species examined. There is no doubt that it must be present (see Introduction) so that, if time had been available, four separate lines of enquiry into the problem could have been followed. Firstly the degree of tissue comminution, cell

rupture and chloroplast fragmentation may have been low (Lyddiatt, 1975). However, the use of acetone blending buffer makes this possibility most unlikely. No microscopic examination of the blended material was carried out, but the dark green colour of the buffer and the pale green, almost white colour of the waste plant material suggested a high degree of cell rupture. Similarly, this could well be the reason for the failure of the extractions using aqueous blending buffer, as there was much less evidence of efficient cell rupture having taken place, despite longer blending times. A second possibility is that the plastocyanin was either denatured by the acetone blending buffer and precipitated at this initial stage or was adsorbed onto particulate fractions (Newcomb, 1963) when in both cases it would be discarded with the cell debris. Spectrophotometrical analysis at 597 nm of a small portion of this precipitate resuspended in a 1% solution of Tween 80 in 20 mM Tris-HCl buffer (pH 7.2), with a small amount of potassium ferricyanide added proved negative. However, the denatured protein, if present would be unlikely to absorb at this wavelength, so no conclusion can be drawn at this stage.

The formation of apoprotein is probably the most feasible explanation of the failure to extract plastocyanin.

Since this was one of the greatest hazards during purification it seems logical to assume that it could happen at some stage during the extraction also. Spectrophotometrical analysis of the 30% acetone buffer after blending and removal of cell debris and also the redissolved protein prior to application to the column to detect any holoprotein still present both proved negative. Two conclusions can be drawn. Either the plastocyanin was extracted and the apoprotein formed almost immediately or that the presence of many other substances (including the acetone) completely obscured the absorption peak of any holoprotein present, which would necessarily be weak, due to the large volumes of liquid involved. A final and less likely explanation is that the plastocyanin was not absorbed onto the resin of the initial DE23 Sephadex. It seems extremely unlikely that the plastocyanins from only certain pteridophytes and gymnosperms would differ sufficiently in amino acid sequence to alter the overall charge of the protein (as is the case of plastocyanins from blue-green algae) to the extent that they were no longer acidic (Lightbody and Krogmann 1967; Aitken, 1975).

The tremendous variation in yields obtained from the different species is also difficult to explain. Although some of the highest yields were extracted from fresh, soft,

light green foliage picked the same day as the extraction (e.g. Osmunda regalis) it did not follow that such characteristics were indicative of a good yield of the protein (e.g. Larix decidua). Plesnicar and Bendall (1970) also noted this variation in yields of plastocyanin from higher plants. They found that the yields of plastocyanin from pea, spinach and green barley chloroplasts were very similar when expressed as the number of chlorophyll molecules per atom of plastocyanin copper, whereas orache and tobacco chloroplasts showed significantly different contents of plastocyanin even though the tissues apparently photosynthesised in the same manner. They concluded that the number of chlorophyll molecules per molecule of electron transport component was not a rigid one.

Plastocyanin has no easily measurable activity. Consequently, the ratio of the absorbance A_{278}/A_{597} at the two major absorbance peaks (reduced and oxidised forms respectively) of the protein was used (Katoh et al., 1962). Plastocyanin samples with an absorbance ratio below 3 were considered to be sufficiently pure for sequence studies to be carried out. Typical values for pure plastocyanin reported in the literature lie in the range 1.2 - 1.4. The lowest figure yet reported is 0.8, but this is apparently unstable (Katoh et al., 1962). Other sources report that

samples with purity ratios as high as 2.0 can still appear to be perfectly homogenous on examination of the copper content or by electrophoresis (Gorman and Levine, 1966). Such high values are thought to be due in part to remaining small quantities of bound polyphenols or ferricyanide (Boulter et al., 1977) and, to some extent, to variations in the aromatic amino acid content of the protein, but mainly to formation of apoprotein.

Determination of complete sequence of plastocyanin from bracken (*Pteridium aquilinum*)

Purity of Sample

The availability of bracken and its high yield of plastocyanin made it appear suitable for sequence studies. However, many extraction difficulties were encountered in early summer when slimy mucus-like components in the extracts eventually led to a low recovery of the protein. Results of later experiments suggested that these components would not have complicated the extraction procedure if 40-50% acetone blending buffer had been used (see Extraction discussion). Extractions carried out during late summer showed no traces of mucus-like substances, but that soluble dark brown materials, possibly polyphenols were present. These did not effect the extraction procedures unduly, but did cause problems during purification. Following repeated gel filtration and ion-exchange purification steps, the final sample of protein which was eventually used for sequence studies had a purity ratio of only 5 by which time purification was halted because the plastocyanin solution turned colourless, indicating the formation of apoprotein. There are two possible reasons for this apparent low purity; either unacceptable levels of impurities were still present or apoprotein was present in larger amounts than normally encountered. The presence of contaminants could explain

some problems during sequencing; cyanogen bromide cleavage was poor, insoluble substances were formed during enzymic and cyanogen bromide cleavage and analysis results for the protein and X2 CNBr fragment showed anomalies. The reasons for spontaneous loss of copper from the molecule are unknown, although plastocyanins from certain species have previously been observed to form apoplastocyanin faster than others, leading to the conclusion that differences in primary structure may be responsible.

Sequence Determination

The sequence of bracken plastocyanin is shown in Figure 12 together with the point of enzymatic and cyanogen bromide cleavage. No overlap is shown between the cyanogen bromide peptides but the ordering of peptides has previously been established in Chlorella and French bean plastocyanin (Kelly & Ambler, 1974; Milne & Wells, 1970). The molecule resembles higher plant rather than algal sequences in as much as there are no additional residues at the N- or C-termini and probably no additions or deletions within the sequence.

Elution patterns of protein and CNBr fragments from Sephadex G50 were similar to those of higher plants,

suggesting a similar molecular size of both protein and peptides. This is in keeping with results for other plastocyanins which have been demonstrated to be monomeric (Boulter et al., 1977). A large amount of unreacted protein remained and some insoluble material formed after treatment with CNBr. Despite careful control of experimental conditions (reduction of formic acid concentration from 75% to 70% and use of parafilm instead of a stopper on the container) little improvement was obtained.

Sequence determination of the X1 CNBr fragment did not pose any unusual problems. The only position at which identification of the amino acid was not immediately straight forward was residue 42, where heterogeneity was observed. Although amino acid analysis data would suggest that mostly Asp occurs in this position, unambiguous identification of Pro was obtained from sequenator data and similar proof for the occurrence of Asp was obtained from chymotryptic and papain peptides. Peptides containing Pro at this position were not isolated during manual sequencing, possibly because they were present in small concentrations but more likely because they would be large and not easily eluted from the paper after electrophoresis. Polymorphism has been demonstrated in several plastocyanin sequences (Boulter et al., 1977). It has usually been chemically conservative, involving

alternatives such as Ile-Leu or Glu-Asp, however polymorphism such as Glu-Val has also been observed. That this heterogeneity occurs within a single plant, as opposed to different individuals of the same species was demonstrated in the case of Malva (Boulter et al., 1977) indicating the presence of two different plastocyanins in the same plant. The polymorphism observed in bracken at position 42 is particularly interesting, as Asp has previously been invariant at this position. Since Pro is not a chemically conservative substitution this may have some effect on the functioning of the molecule.

Anomalous cleavage at the Ala-Gly bond (positions 23-24) was observed several times, both prior to separation of CNBr peptides and whilst a lyophilised sample of the X1 fragment was stored at $4^{o}C$ for about eight weeks. In the former case, the peptide formed from residues 24-57 eluted from the column with the X2 fragment, whereas the peptide composed of residues 1-23 eluted off the column after the X2 peak and before the X3. The cause of this spontaneous breakage of the peptide bond is unknown.

Diffulties were encountered in the identification of the complete X2 CNBr peptide sequence. Unambiguous results were obtained for all residues except residues 88-91. Incompatible amino acid analyses for total protein and the

X2 fragment plus poor C-terminal results on the complete fragment mean that little can be concluded from this information. The amino acid analysis of peptide X2T2 (Table 15) supports the sequence as shown in Figure 12. The possibility cannot be overlooked that additions or deletions have occurred here, although this seems unlikely as this region of the polypeptide chain is highly conserved and is suspected of performing an important function (see later Discussion). Should an alteration in the primary structure be finally proved in this position it may explain the apparent ease with which apoplastocyanin was formed during purification.

Heterogeneity was shown to occur at residue 68. Evidence from dansylation and amino acid analyses suggest aspartic acid and glutamic acid occur in approximately equal amounts, although estimation by dansylation was difficult as both residues tend to 'carry over' during dansyl-Edman analysis.

Dansyl-Edman analysis of the X3 fragment indicated the presence of both Ile and Phe at position 97, however a longer hydrolysis time (20 hours) and amino acid analysis results revealed that the misidentified Phe spot observed after thin layer chromatography was in fact an Ile-Val dipeptide. The occurrence of Trp at the C-terminus is interesting as this amino acid is absent from all higher plant sequences so far determined and is found only in the sequence of the green

alga *Chlorella fusca* at position 29.

Other difficulties encountered in the determination of the sequence were routine and have been discussed many times previously (Ramshaw, 1972; Meatyard, 1974; Lyddiatt, 1975; Valentine, 1976; Takruri, 1979).

Complete sequence studies

The complete sequences of plastocyanins from higher plants, algae and bracken are shown in Appendix IVA.

The N-terminal sequences of other plastocyanins are shown in Fig. 14. Appendix IVC summarises the invariant residues common to each major group on the basis of this N-terminal data.

Before elucidation of the tertiary structure of any plastocyanin molecule, the structural and functional significance of invariant or conserved residues could only be speculated. Since the production of crystallographic data for plastocyanin from poplar (Populus nigra var. italica) by Colman et al., (1978), more significance can be placed on these results. A projection of the plastocyanin molecule based on the poplar data is shown in Figure 21, and there is good reason to suggest that other plastocyanins have essentially the same structure. For example, the NMR spectra of a series of plant plastocyanins prove that the environment of the copper atom is highly conserved (Freeman et al., 1978). Comparisons with other plastocyanin primary structures indicate that many residues invariant in the sequences so far determined are mostly those which are thought to have specific functions on the basis of the x-ray data. Also, additions and deletions which occur in algal sequences

(Boulter et al., 1977) would, if applied to the poplar sequence have apparent minimal effects since they occur in chains lying on the outside of the molecule. For these reasons, and in the absence of any evidence to the contrary, the tertiary structure of bracken plastocyanin is assumed to approximate to that of poplar.

Prior to the availability of the x-ray analysis data, various predictions had been made about the molecule. An irregular co-ordination geometry for the copper atom (as a tetrahedral distortion from square-planar co-ordination) had already been forecast to explain the unusual features of type 1 copper-protein EPR spectra (Blumberg, 1966; Brill and Bryce, 1968). The location of the copper site had been forecast as being 'relatively inaccessible to solvent' after proton relaxation measurements (Blumberg and Peisach, 1966). Several amino acids had been proposed as possible copper ligands, confirmed by x-ray analysis as being His 37, Cys 84, His 87 and Met 92 (see Figure 25). The first three ligands were predictable from experimental data from the protein and model compounds (Katoh and Takamiya, 1964; Graziani et al., 1974; Markley et al., 1975) and sequence homology (Boulter et al., 1977). Suggestions that the fourth ligand was formed from the phenolate oxygen of a Tyr side chain (Amundsen et al., 1977)

or the depronated nitrogen of a peptide group elsewhere in the molecule (Hare et al., 1976) are now known to be incorrect, but the only evidence to suggest the involvement of Met came from experimental data from model compounds (Jones et al., 1975). An attempt was also made to predict the most probable values of all the torsion angles between the backbone peptide groups, and hence the three-dimensional structure (Wallace, 1976). These results are now seen to have been successful at the level of the secondary structure, but less so, as predicted, at the tertiary level of structure.

The plastocyanin molecule resembles a slightly flattened cylinder, formed from eight strands of the polypeptide chain, although the exact positions of the hydrogen bonds between these strands have not yet been published. The core of the poplar plastocyanin molecule is hydrophobic and notably aromatic, due to a clustering of six of the seven Phe residues with their side chains on the interior of the molecule (see Figure 22). In contrast, acidic and basic chains are mostly on the exterior of the molecule (see Figures 23 and 24). Most higher plants have six Phe and three Tyr residues, poplar has seven Phe and two Tyr residues. The other Phe residue at position 35 does not form part of the hydrophobic core, instead the side chain

projects into the surrounding medium. It is possible that this may play some part in preventing the formation of apoplastocyanin, providing a possible explanation as to why if this was absent, the copper atom apparently drops out of some plastocyanins more readily than others. Bracken plastocyanin contains five Phe and three Tyr residues, and their approximate distribution is shown in Figure 27, and although no conclusion can be reached as to the position of the side chains a similar clustering is apparent. Of the six Phe residues conserved in higher plants (at positions 14, 19, 29, 35, 41, 82) only three are conserved in bracken (at positions 14, 29 and 41). However Phe 19 and Phe 82 are substituted by Ile (hydrophobic) and Tyr (aromatic) respectively - both relatively conservative substitutions. Similarly, Tyr 80 and Tyr 83 remain invariant. In plastocyanin from bracken and other higher ferns Phe 35 is absent, however an additional Phe residue at position 12 may compensate for this, as it lies in a loop of the polypeptide strand which is also adjacent to the copper atom.

The five lysine residues are randomly distributed through the poplar plastocyanin molecule, and a similar pattern is apparent in higher plants and bracken (see Figures 24 and 28). In contrast, the acidic residues are not evenly distributed. Poplar plastocyanin has an elongated negative

region on the exterior of the molecule extending from the carboxylate group of residues 42-44 to those of residues 59-61. Five of these six residues are conserved in other higher plant plastocyanins, indicating a similar pattern. A notable difference exists between these molecules and plastocyanin from bracken, in which the acid region has moved to residues 59-61 and 67-69. This is not such a drastic step as it may seem, as examination of Figures 29 and 30 show that these differing acidic regions are in fact in almost the same position on adjacent strands of the polypeptide chain. The reason for these differences can only be speculated about, especially as the function of this negative region is uncertain. It may be pure chance, or possibly a difference in function or even due to a change in pattern of corresponding molecules with which the plastocyanin molecule comes into contact (e.g. a membrane).

A feature of the poplar plastocyanin molecule which may be significant is the hydrophobic patch formed by residues on the loops of the polypeptide chain surrounding the opening underneath which is the copper site. These are residues 10, 12, 35, 36 and 89-91. Examination of all plastocyanin sequences determined to date shows that with the exception of residue 35 (previously discussed) amino acids occuring at these positions, (See Appdx.IV8) whilst not all invariant,

all exhibit hydrophobic properties, strongly suggesting some functional importance. This is in line with the evidence that solvent molecules cannot approach nearer than 6 Å to the copper atom.

An unusual folding of the polypeptide chain is found at position 16 in poplar plastocyanin where a Pro residue is involved in a cis bond. This residue is invariant in all sequences suggesting a similar configuration is present in all, possibly due to some structural constraint because of the copper site. Other invariant residues include the copper ligands and several glycine residues (at positions 6, 10, 24, 78, 89, 91 and 94) which may be present because of their chemical properties or for structural reasons, the absence of a large side chain being advantageous, for example, at a tight bend.

As more widely differing species are added to the data set, fewer amino acid positions are shown to be truly invariant. If only the higher plant sequences are compared, the number of invariant residues is as high as 50. Addition of plastocyanin sequences from blue-green and green algae reduces this number to 27 (Boulter et al., 1977) and addition of lower plant data further reduces it to 19.
This is important since the longer an amino acid remains invariant, despite additions of widely differing sequences

to the data set, the stronger is the indication of a specific function for that amino acid, as in the case of the copper ligands.

If the sequences of plastocyanins and azurins are compared (see Introduction) even fewer residues are invariant (Colman et al., 1978), as shown in Figure 31. It is interesting that data from the higher ferns has shown that Gly 67 and Val40 are no longer invariant. In the bracken sequence, position 67 is Glu and in monkey puzzle plastocyanin, position 40 is Ile, therefore leaving the only invariant residues around the copper site. From their results, Colman et al. concluded that since azurins and plastocyanins have both similar biological roles and similar copper sites, their molecular structure must also be similar.

This evidence demonstrates the value of obtaining sequence data from species which are not closely related when investigating structure-function relationships of a protein, in contrast to evolutionary studies in which sequence data from more closely related species is desirable.

The mechanism of electron transport within the molecule, to and from the copper atom is still uncertain. The concept of 'outer-sphere' interaction (Moore and Williams, 1976) would apply if plastocyanin reacted with another

metalloprotein, when the metal-ligand bonds would remain unbroken in both molecules. Electron transfer could take place via the imidazole ring of His 87 which is the only structure separating the copper atom from its surroundings at one end of the molecule. A second theory involves transfer of the electron via pathways formed from hydrophobic side chains within the molecule. Of many possible such pathways, Colman *et al*. suggest one which (in poplar plastocyanin) starts at the periphery of the molecule in the acidic region and is formed from residues Phe 82, Val 93, Gly 94 and Phe 14 all of which are invariant in higher plants. Unfortunately the bracken plastocyanin data does not fit so well, in particular Lys occurs at position 94 instead of Val and the displacement of the acidic region, which has already been discussed, suggest that either this pathway is not the correct one or that the functional regions of the two molecules have evolved in different directions.

The functions of other invariant or highly conserved residues within the plastocyanin sequences so far determined is unclear. Possibly they have no specific function, but more probably they could be sites of recognition between the plastocyanin molecule and molecules in solution or in the membrane.

FIGURE 21

Projection of the poplar plastocyanin molecule

The circles represent the positions of the α-carbon atoms of the component amino acids, which are numbered following the scheme of Boulter *et al.*, (1977) for higher plant plastocyanin sequences (see Appendix IVA). The letters N and C represent the NH_2-terminal and COO-terminal residues respectively. The approximate directions of the Cu-ligands are indicated.

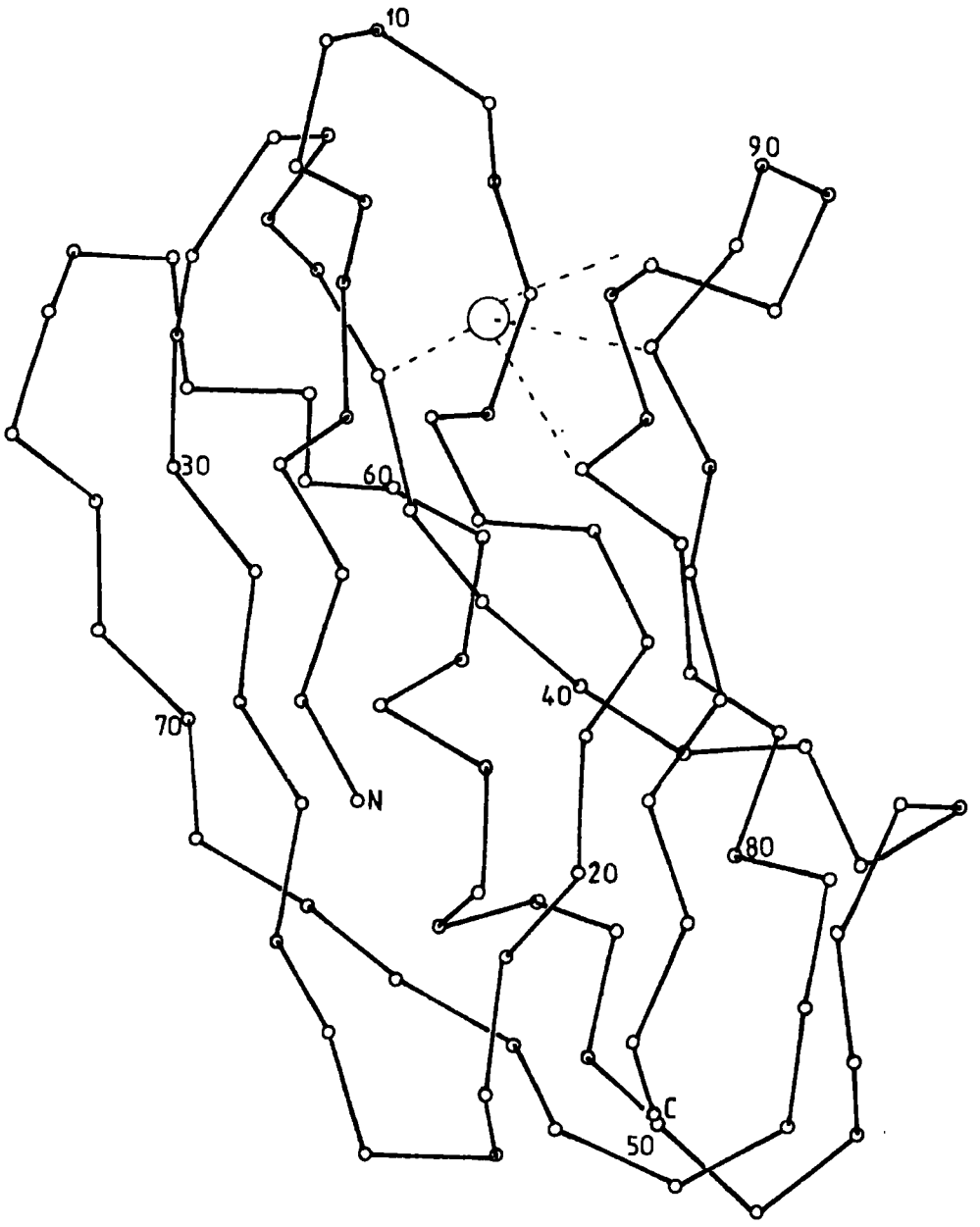

10
90
30
60
40
70
N
20
80
C
50

FIGURE 22

Aromatic residues in poplar plastocyanin

Computer drawing of the poplar plastocyanin molecule based on x-ray crystallographic data (Colman et al., 1978) showing the positions of side chains which affect the aromatic region of the spectrum.

Side chains in red represent phenylaline residues.

Side chains in green represent two histidine residues (around the copper site) and two tyrosine residues.

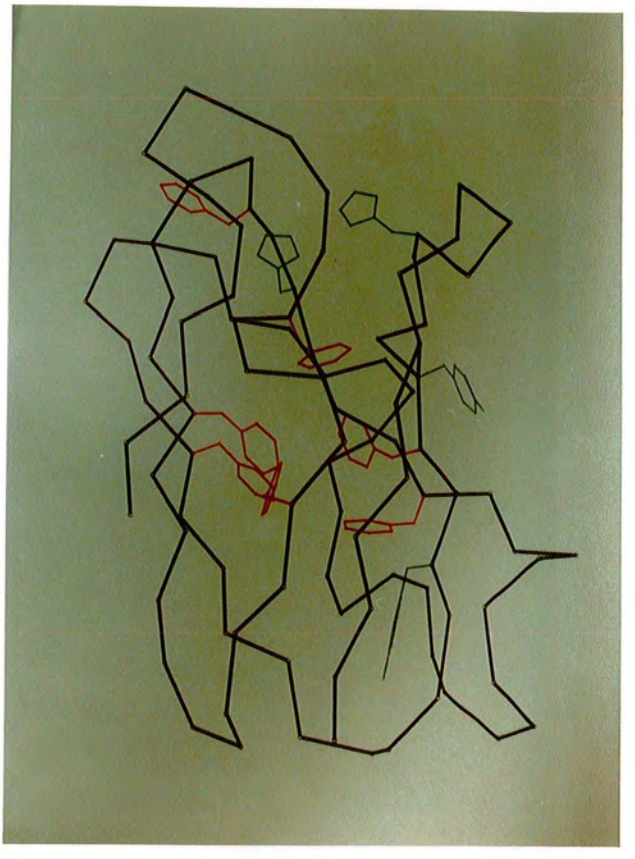


FIGURE 23

Conserved acidic residues in higher plant plastocyanin

Computer drawing of the plastocyanin molecule showing the positions of conserved acidic side chains found in the higher plant plastocyanins sequenced to date (Boulter et al., 1977).

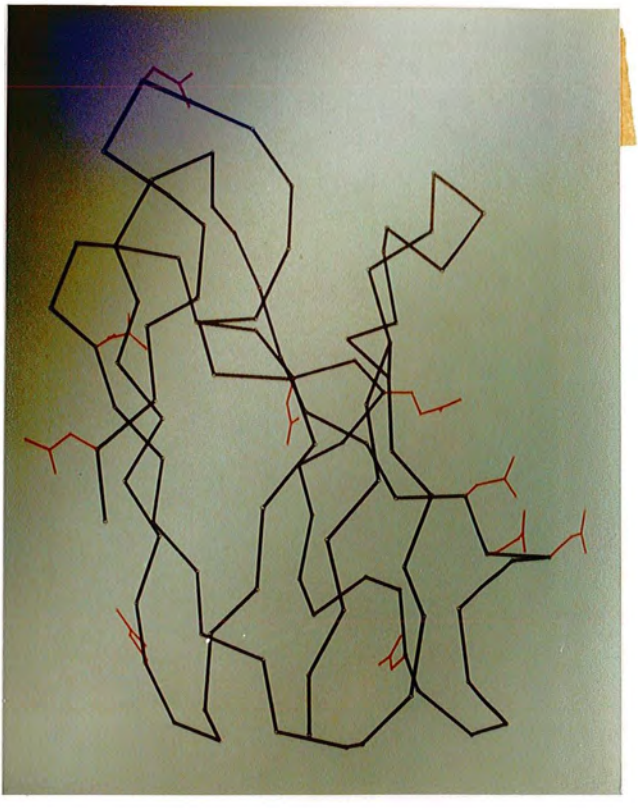


FIGURE 24

Basic residues in higher plant plastocyanin

Computer drawing of the plastocyanin molecule showing the position of lysine side-chains which are either invariant (shown in blue) or nearly always invariant (shown in green) in all higher plant plastocyanin sequences determined to date.

FIGURE 25

Copper ligands in the plastocyanin molecule

Computer drawing of the plastocyanin molecule showing side chains involved in binding of the copper atom (His 37, Cys 84, His 87 and Met 92).

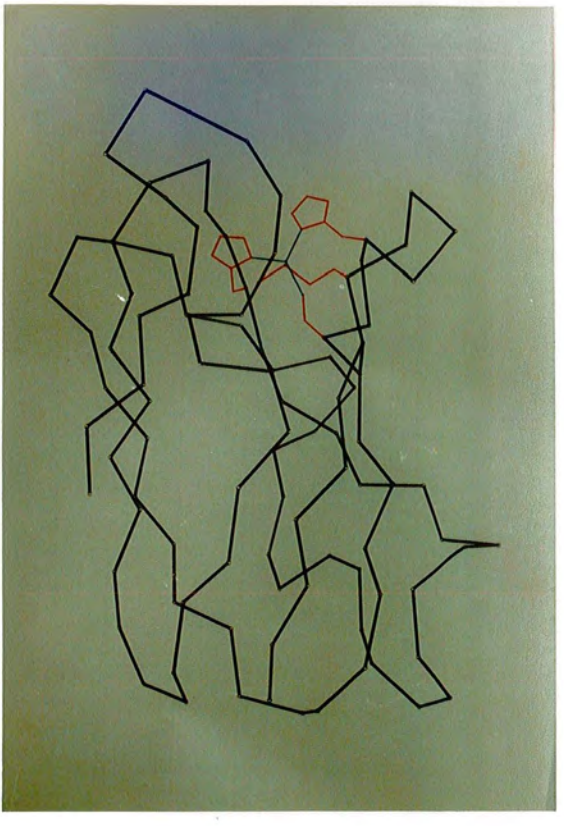


FIGURE 26

Poplar plastocyanin molecule

Computer drawing of the popular plastocyanin molecule showing the positions of all side chains.

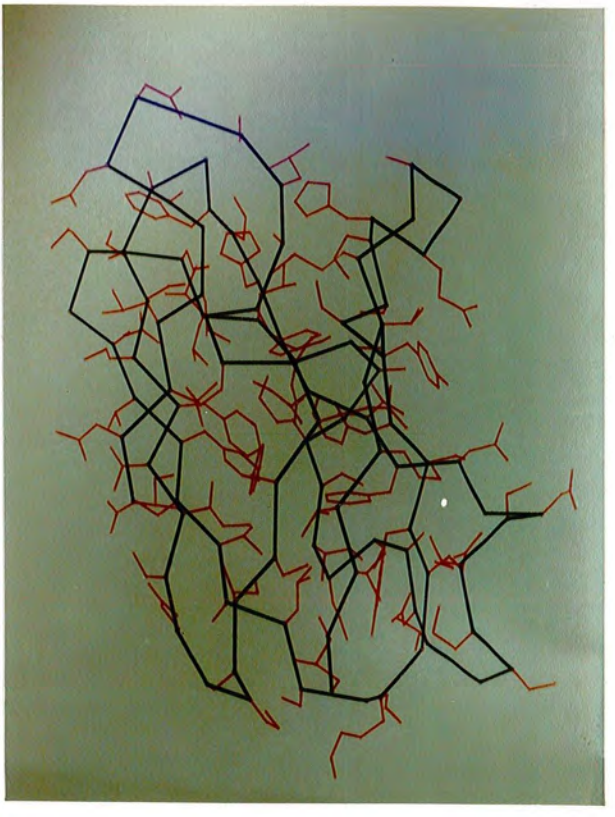


FIGURE 27

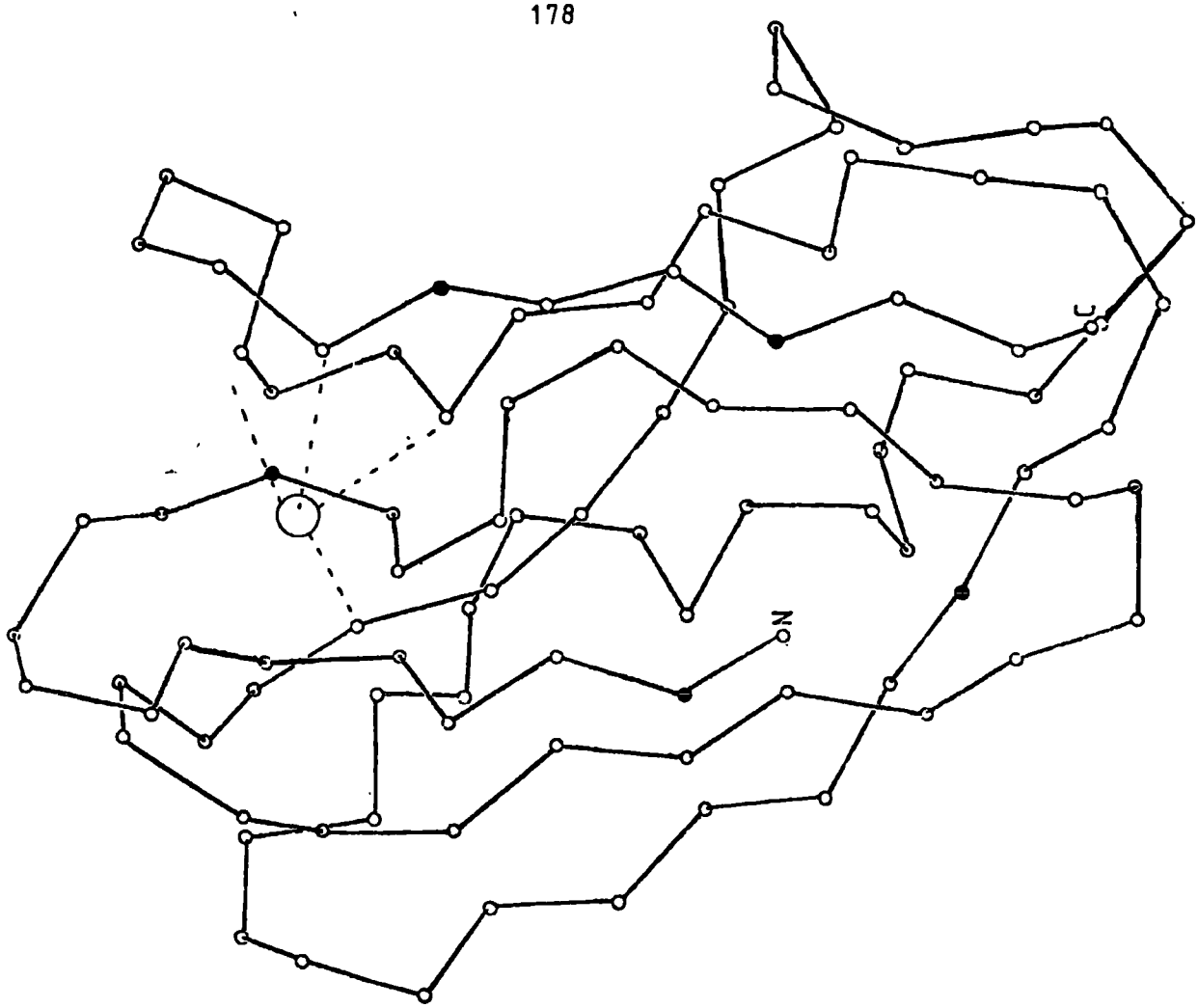
Aromatic residues in bracken plastocyanin

Drawing showing the positions of phenylalanine (red) and tyrosine (green) residues in bracken plastocyanin, assuming the configuration of the polypeptide chain is similar to that of poplar plastocyanin.

FIGURE 28

Lysine residues in bracken plastocyanin

Drawing showing the positions of lysine (blue) residues in bracken plastocyanin, assuming the configuration of the polypeptide chain is similar to that of poplar plastocyanin.


C
N

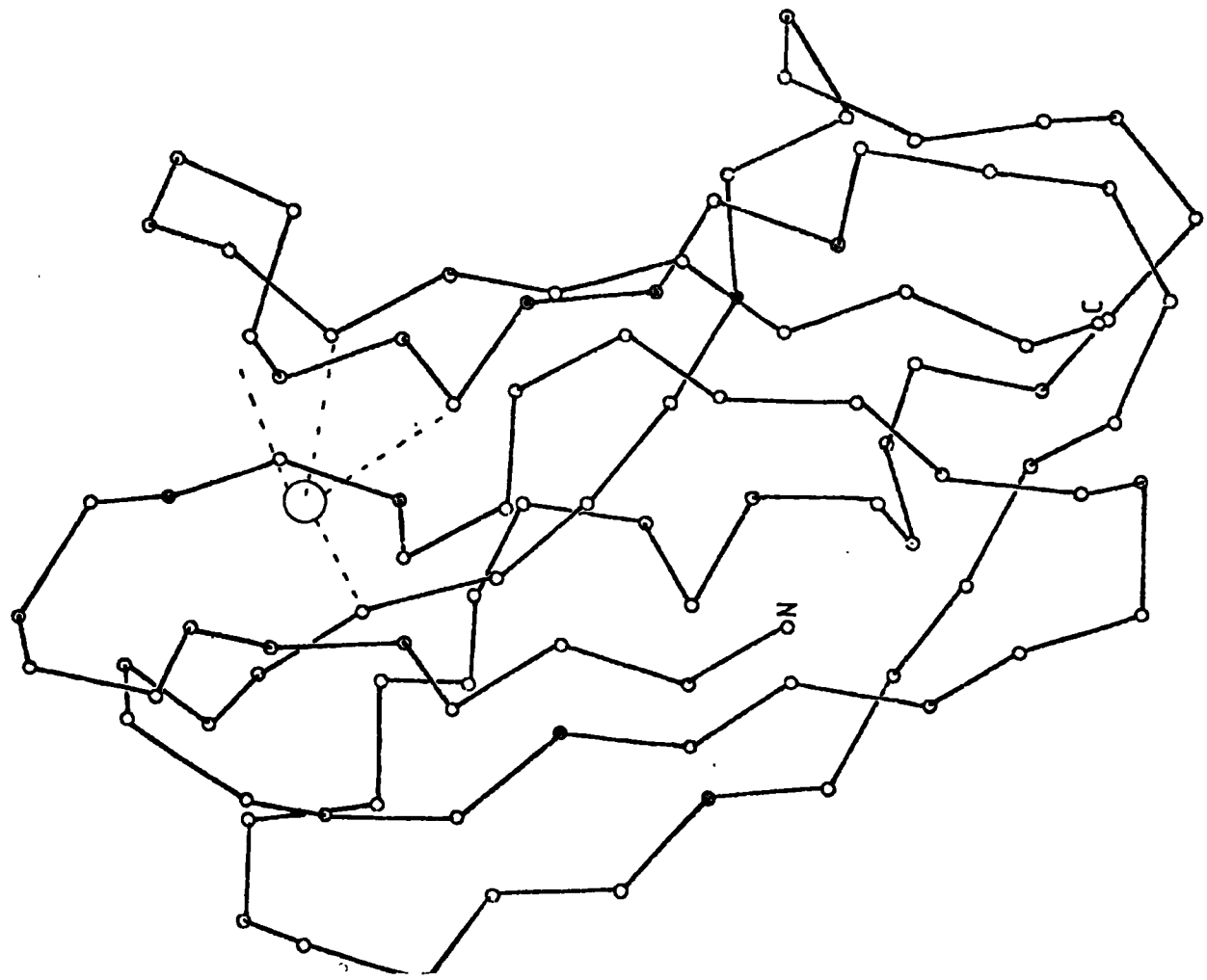

C
N

FIGURE 29

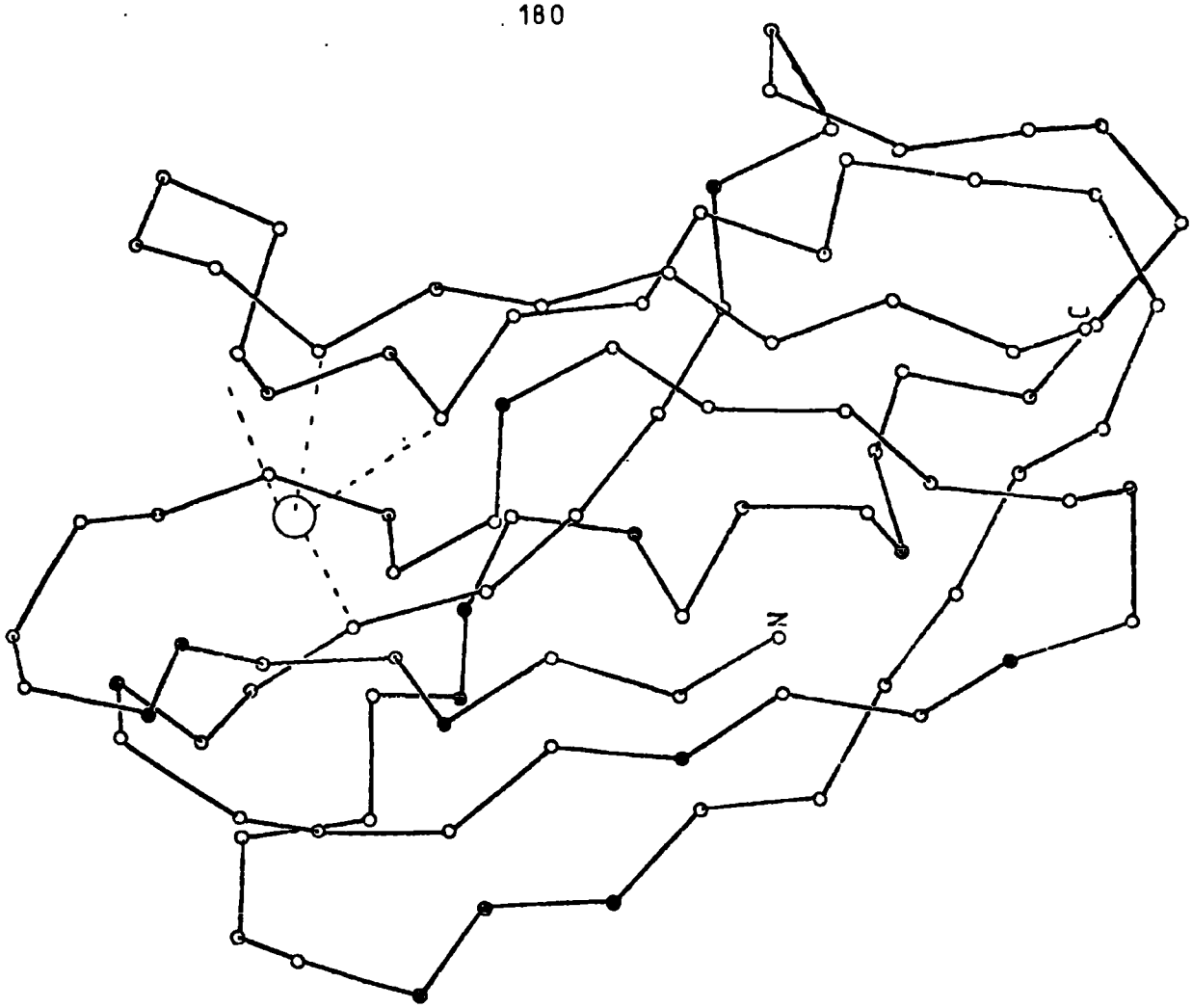
Acidic residues in higher plant plastocyanin

Drawing showing the positions of acidic residues (coloured in red) conserved in all higher plant plastocyanin sequences determined to date.

FIGURE 30

Acidic residues in bracken plastocyanin

Drawing showing the positions of acidic residues (coloured in red) in bracken plastocyanin.


C
NO

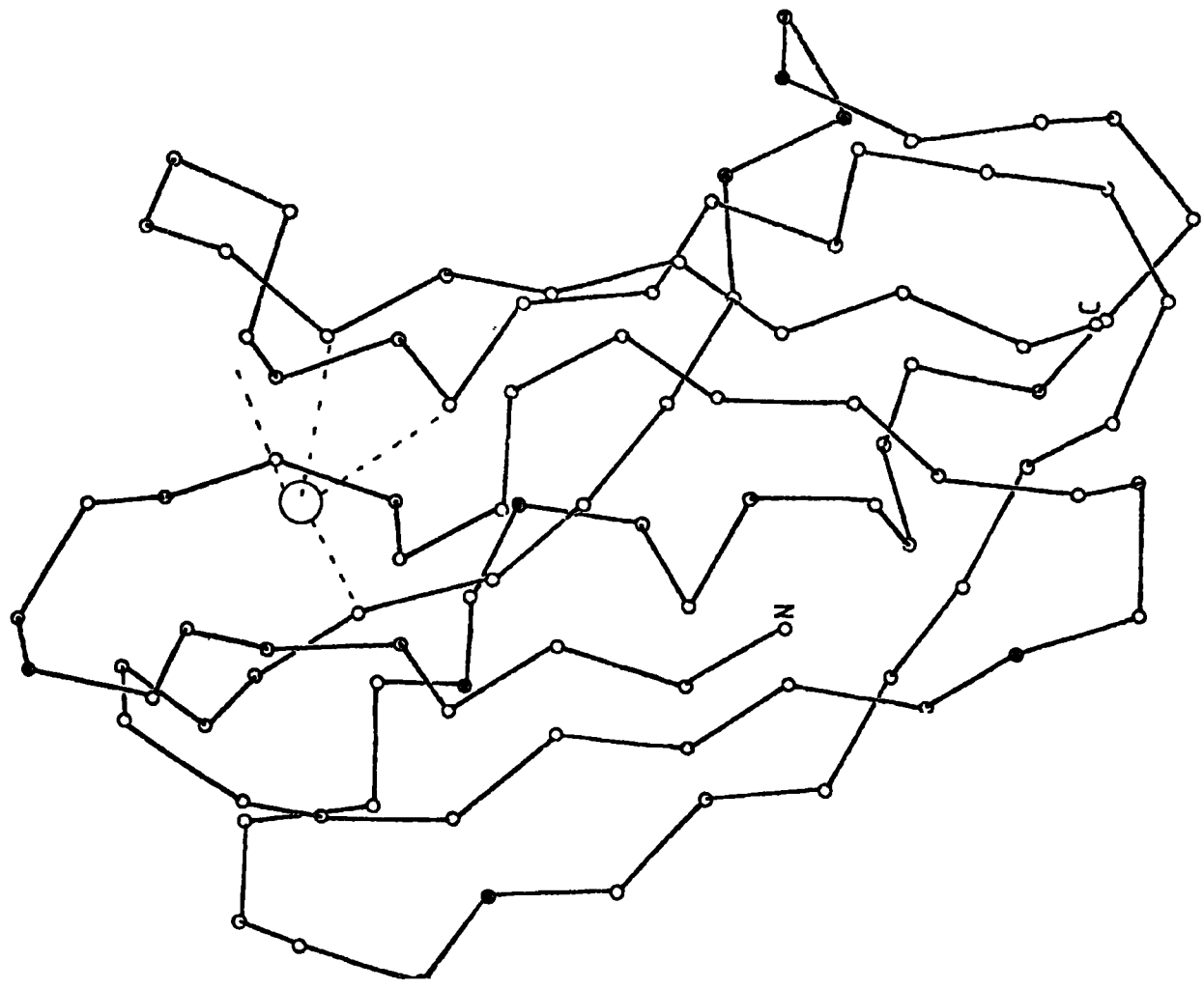
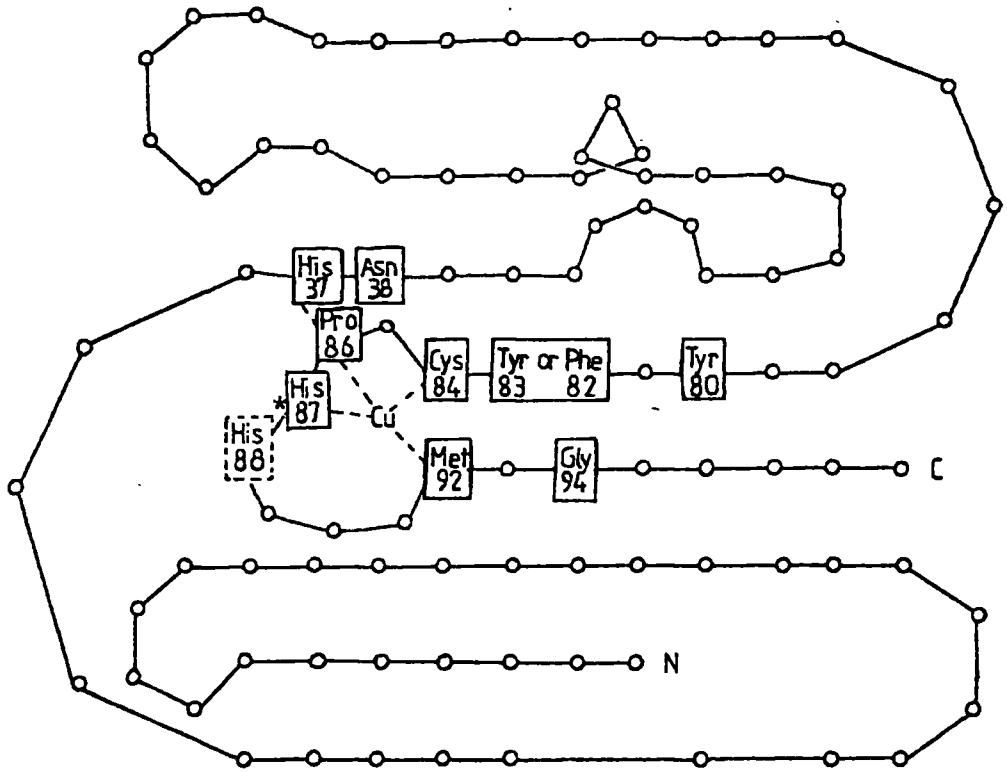

C
NO

FIGURE 31

Invariant residues common to all plastocyanins and all azurins

Topological diagram for plastocyanin showing residues (in boxes) which are invariant in all plastocyanins and all azurins.

The diagram is a modification of that of Colman *et al.*, 1978, as sequence data from *Pteridium aquilinum* and *Araucaria araucana* indicate that two residues previously considered to be conserved, Gly at position 67 and Val at position 40 are variable.



* His conserved at residue 87 in plastocyanins (Boulter et al., 1977) and residue 88 in azurins (Ryden and Lundgren, 1976).

N-terminal sequence determination

An inspection of the N-terminal data set (Figure 14) reveals that there are, as expected, greater similarities between members of the same natural group than between members of different groups. For example, the two Equisetum spp. sequences are remarkably similar, differing at only seven positions. This is also the case for the sequences from the two Dryopteris spp. which have only 5 differences between them. However comparison between Equisetum sylvaticum and Dryopteris dilatata reveals 25 dissimilar positions. When the whole data set are considered only 7 residues remain invariant throughout all the groups (Appendix IVC). A more meaningful analysis of this data was obtained using computer methods.

Construction of phylogenetic trees from N-terminal protein sequences

In order to construct a phylogenetic tree from protein sequences, the group of proteins comprising the data set must be shown to be similar due to homology (Fitch and Margoliash, 1967; Fitch, 1970; Fitch and Markowitz, 1970; Needleman and Wunsch, 1970; Boulter, 1973a), as opposed to convergence or simply chance.

Data acquisition can be a major problem, involving the use of large amounts of plant material, and the development of new protein extraction and purification methods. Such problems can be made easier by the careful selection of the protein to be sequenced and by the use of recognised rapid and accurate automatic sequencing techniques (Haslett and Boulter, 1976). At the present time automated methods using spinning cup and solid phase sequences are at a stage where to determine the complete sequence is difficult and requires considerable time. However the first forty residues of each sequence can be very rapidly obtained. It is of interest therefore that Boulter *et al.* (1977) showed that the use of the full sequence rather than the first forty residues in the case of plastocyanins would be on average only 1.6 times as sensitive and that a good correlation exists (Haslett *et al.*, 1977; Peacock and Boulter, 1975). Since the rate of data acquisition by

use of automated techniques is many times greater than if manual methods were employed, and the amount of sequence data available is in itself a limiting factor in the construction of phylogenetic trees (Mayer et al., 1953), then acquisition of a large data set of partial sequences is more desirable than a smaller set of complete sequences.

Some manipulation of the data may be required before meaningful comparisons of sequences can be obtained. If deletions or additions have occurred, the sequences may have to be realigned, so involving the possibility of error or differences of opinion. Several successful mathematical models have been proposed to minimise such problems (Moore and Goodman, 1977), which however were not encountered in analysis of the plastocyanin data from the lower plants (i.e. deletions were absent).

Handling of sequence data is still in its infancy, and is where most of the criticisms of molecular evolutionary techniques usually fall. Because the human mind cannot process several amino acid sequences at once, the use of mathematical and computer methods, to which this sort of data is well suited, is essential. Originally, analysis of sequence data was by numerical matrix methods (Dayhoff, 1972), which assumed a constant rate of evolution and therefore the greater the number of amino acid differences between two

protein sequences, the most distantly related the species they represent. A more sophisticated method involves the comparison of various phylogenetic trees constructed by the computer in turn until the tree (or trees) is found which, taking into account the hypothetical ancestral sequences at each of the nodes, would require the fewest amino acid changes (i.e. maximum parsimony) to explain the topology (Boulter et al., 1972; Boulter, 1973b). The advantages of this ancestral sequence method are that it is essentially an objective approach and so convergent and back mutations can often be discerned (but see later), unlike traditional subjective methods using fossil evidence. In addition, even if different rates of mutation have occurred along the various lines of descent, leading to differences in the relevant branch lengths, the positions of the actual branch points should nevertheless be correct (Boulter, 1974). Boulter and Peacock (1975) assessed the accuracy of this method by using a model and simulated cytochrome c sequences and obtained good correlation with the actual results. The most obvious criticism of the ancestral sequence method is of the assumption the evolution proceeds by the most parsimonious route, which presumably may not always be correct. However, this is insufficient reason for rejecting the method, but rather a challenge to be overcome by alternative data handling methods yet to be developed, and a

slight reservation to be borne in mind when interpreting the results. It can be argued (Cronquist, 1976) that because the method is slightly complicated, it is hard to perceive inaccuracies. However, the interpretation of fossil evidence has never been straightforward, otherwise the need for other taxonomic aids such as protein evolution would not have arisen. Certainly, some proteins are more suitable for such studies than others, depending on the closeness of the supposed taxonomic relationship to be investigated. A slowly evolving protein will not show sufficient sequence differences for any relationships to be determined in, for example, intra-familial studies. Similarly a protein which is evolving at an unsuitably fast rate will possess so many differences between the protein sequences from the individuals under study that back, parallel and convergent mutations will not be detected, even by this method, and consequently the true relationships will be obscured. Since the rate of evolution of a protein cannot be known until a few sequences have been established, some inaccuracies may at first appear, until sufficient data has been accumulated. From investigations completed to date, it is clear that a certain number of reasonably closely related species are required before a tree becomes meaningful and that the resulting trees are more accurate near the branch ends and less accurate the further into the tree one proceeds

(Boulter and Peacock, 1975) - a logical conclusion since the inner branches depict the most distant relationships and therefore contain the highest proportion of undetected back, parallel and convergent mutations.

Possibly the best example of a phylogenetic tree obtained from molecular data is that based on the complete sequences of cytochrome c from vertebrates (Dayhoff, 1972), which is in general agreement with the phylogenetic relationships induced from comparative morphology, serology and the fossil record (Romer, 1945). Recent attempts to construct phylogenies applicable to controversial areas of classical invertebrate phylogeny have been only partly successful (Lydiatt et al., 1978) because of the small numbers of highly variant sequences involved. Phylogenies based on complete sequences of cytochrome c from higher plants have been constructed (Boulter et al., 1972) prompting a great deal of discussion, because lack of fossil evidence as regards angiosperm origins means no firm evidence can be provided for any of the diverse opinions on the subject (Davis and Heywood, 1963; Cronquist, 1968; Takhtajan, 1969; Thorne, 1968). Molecular evolution within the algae has been investigated using cytochrome c (Meatyard, 1974) and ferredoxin (Takruri, 1979). The trees produced from ferredoxin sequences show good correspondence with the main

outline of the traditional views of evolution within the plant kingdom. Thus the blue-green and red algae come on a separate branch to the green algae and higher plants, whilst the "fern allies" are seen as being quite distinct in evolutionary terms.

The relatively high yield of plastocyanin and ease of extraction suggested that it would be an ideal protein for evolutionary studies (Ramshaw et al., 1973). However, it has been estimated that plastocyanin evolves approximately twice as fast as cytochrome c, therefore incorporating twice the errors into any phylogeny constructed from plastocyanin sequences compared with a similar phylogeny constructed from cytochrome c sequences from the same species. Nevertheless, some useful information has been obtained, especially in intrafamilial studies, for example, among the Compositae (Boulter et al., 1977)

The widely accepted view regarding the evolution of plants other than the algae is that the green algae gave rise to a primitive group of plants known as psilophytes and from these primitive pteridophytes, the main divisions of plants known today evolved along separate paths. One not widely held theory (Church, 1919) is that the main divisions of the plant kingdom had been evolving along separate pathways even earlier, since the days of their algal ancestors.

In any event the precise nature of these pathways as determined from the fossil record is still under much discussion. It was hoped that computer analysis of the amino acid sequences obtained in this study might throw some light on the problem.

Figures 15 to 18 represent the topologies obtained from the sequence data (see Results). To avoid undue complication, Figures 16 and 18 will not be discussed separately as they show only a slight difference (in the relative positions of the two Dryopteris spp.) from Figures 15 and 17 respectively.

It is useful for a phylogenetic tree like Figure 15 to have an 'origin'. Rooting such a tree (i.e. finding the position of the hypothetical common ancestor) can be done in one of several ways. Most simply, the mid-point of the tree can be assumed to represent this origin. If the mid-point of the two longest branches shown in Figure 15 is taken, Figure 32 is the result - suggesting that the three major plant groups, the ferns (with the exception of Osmunda), horsetails and gymnosperms separated at the same time. Such a method assumes that a constant rate of molecular evolution must have taken place at least along the major lines of evolution. This is a very large assumption to make, as the horsetail branch is obviously much shorter than the other two.

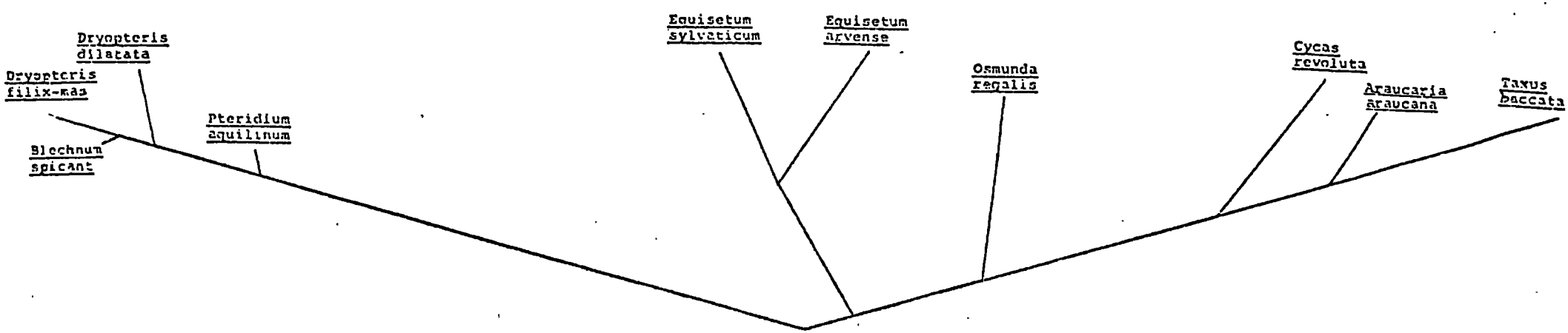
To avoid this assumption, the sequence of a protein from

FIGURE 32

'Rooting' a phylogenetic tree

An exercise in rooting the phylogenetic tree shown in Figure 15.

The mid-point of the two longest branches was used to root the tree.


Dryopteris dilatata
Dryopteris filix-mas
Blechnum spicant
Pteridium aquilinum
Equisetum sylvaticum
Equisetum arvense
Osmunda regalis
Cycas revoluta
Araucaria araucana
Taxus baccata

species (in this case, the green algae) which are regarded as being taxonomically ancestral to the other species in the data set have been included (see Figure 17).

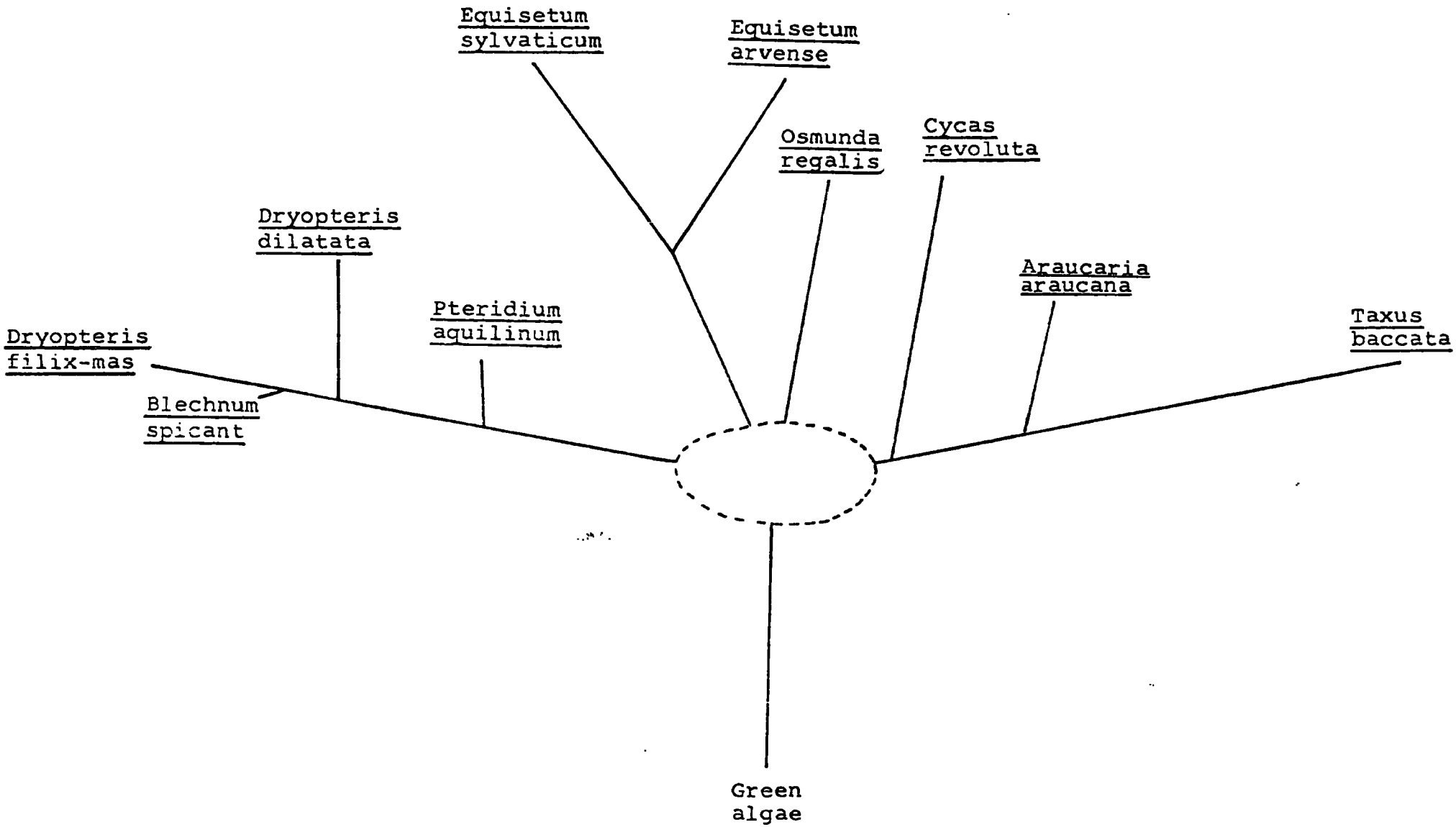
When this is done, the mid-point of the two branches is found to be almost at the same point as the algal branch meets the tree (i.e. Figures 17 and 32 are in good agreement). It must be remembered that the plastocyanin sequences of the algal species used are of present day species and as such should not be regarded as being ancestral to any other hence the topology drawn in Figure 17 may be misleading. It would possibly be more correct to draw the algal branch upwards to avoid such an implication, but the visual impact of the tree would then be lost. Because of the limit on the number of species which could be handled by the computer, the data from the green algal sequences used was treated as if it was one species showing heterogeneity (see Methods). This had the effect of 'forcing' the two sequences onto one branch.

A third method of rooting such a phylogenetic tree is more subjective - a logical use of the fossil record and contemporary taxonomic opinion. Figure 33 is the best tree which can then be extrapolated from the limited data available in this study. In this tree, the point at which the algal branch meets the tree is assumed to be the position

FIGURE 33

A summary of the topologies produced by computer analysis of N-terminal sequence data, taking into account information from the fossil record

The dotted circle represents a basic primitive plant stock from which, at unknown times, the four lines separated.


Equisetum sylvaticum
Equisetum arvense
Osmunda regalis
Cycas revoluta
Araucaria araucana
Taxus baccata
Dryopteris dilatata
Pteridium aquilinum
Dryopteris filix-mas
Blechnum spicant
Green algae

of a common ancestor of the ferns and gymnosperms. From the traditional viewpoint, this was probably a member of the Psilophyta , but it has already been mentioned that this divergence could have taken place earlier, in which case the common ancestor would have been an alga. Because plastocyanin has been found to be a fast evolving protein, and the range of species chosen in this study was wide, there is a large likelihood that many of the differences between the proteins are too large for significant comparisons to take place. Consequently, the positions of the nodes of the Equisetum sp. and Osmunda regalis branches are assumed to be inaccurate as they are based on insufficient data. For the same reasons the branch lengths were considered to be inaccurate when constructing this version of the tree. Consequently, the points of divergence of the four branches have not been represented exactly, but rather the dotted circle represents a basic primitive plant stock, from which, at unknown times the four branches separated.

It is therefore clear that whilst assumptions have been made in the construction of these trees, further assumptions are necessary for their interpretation. For example, a time scale (linear or otherwise), over which the evolutionary events depicted by the tree supposedly occurred, must be assumed. If no significant importance is to be

placed on the branch lengths of the tree, then more reliance must be given to classical taxonomic opinion based on fossil evidence, in which case some of the impartiality of the method is lost.

Two approaches have been made in interpreting the trees represented in Figures 17, 32 and 33. Firstly, a subjective examination of the trees with regard to contemporary taxonomic opinion and secondly a more objective examination of the trees in terms of their branch lengths and consequent rates of evolution.

A. Calamophyta

At one time it was thought that the ferns, lycopods and horsetails were sufficiently related to be grouped together in one natural division of the plant kingdom. However, Jeffrey (1902) showed that there are two 'stocks' of vascular plants - the Pteropsida (ferns, gymnosperms and angiosperms) and the Lycopsida (lycopods and horsetails). As a consequence, Scott (1909) was the first to give the horsetails the rank of division. Any of the three interpretations of the phylogenetic tree so far discussed confirms this idea that there is no evidence of a recent common ancestor between the ferns and horsetails, and the elevation of these groups to the level of subdivision (Swain, 1974) or division (Smith, 1955) of the plant

kingdom is justified.

Most modern theories therefore assume a completely independent line of development of the Calamophyta since the time of the psilophyte ancestors (Takhtajan, 1969). Only Scott (1920) mentions the possibility of the Calamophyta diverging from the gymnosperm line of descent (as indicated in Figures 17 and 32) when reviewing and eventually rejecting (on the grounds of parallel evolution) a suggestion by Renault that some primitive horsetails, the calamariae, bore seed-like structures.

B. Pterophyta

The earliest known fossil 'ferns' are intermediate in type between psilophytes and present ferns and are 350 million years old (Devonian). Many fossil ferns are known from the carboniferous but most are very different from present-day ferns and are thought to have died out during the Upper Permian. Some present-day ferns are directly related to these carboniferous ferns, but others have no known fossil ancestors, (Smith, 1955) see Appendix III. Eusporangiate ferns are considered to be the most primitive of present day ferns and Leptosporangiate ferns more recent, with some evidence to suggest placing Osmunda in an intermediate position (Bower 1935). In this study it was hoped to obtain data from at least one Eusporangiate fern, also

Osmunda regalis and representatives of several families of Leptosporangiate ferns, including the two heterosporous orders. Unfortunately, only data from Osmunda regalis and four representatives of the Polypodiaceae (an artificial family) were obtained. Consequently, little can be concluded from the relative positions of the five species.

The four members of the Polypodiaceae apparently form a tight group, although it is interesting that the first member to diverge is Pteridium aquilinum. One not widely held theory (Bower 1935) is that the ferns evolved along two separate lines in terms of their fructification - whether the sporangia are marginal or superficial. If this scheme is followed, a pattern as shown in Appendix III B is obtained, with the Pteroids (Pteridium aquilinum) being separated from the Blechnoids (Blechnum spicant) and Dryopteroids (Dryopteris spp.). This theory assumes that the two series, Marginales and Superficiales, have been distinct from each other since Palaeozoic times. This will be discussed later. It is possible, however, that the results correspond with the theory, purely coincidentally, because of the few species involved.

The two Dryopteris spp. represented in the tree do not appear on a single branch as might have been expected.

However analysis of the sequence data suggests this may be an artefact due to back mutation which would be removed if further closely related species were included.

The anomalous position of Osmunda regalis, on the gymnosperm branch instead of the fern branch, is hard to account for, although as already explained, the branch details in this part of the tree are unreliable due to the fast-evolving nature of plastocyanin and the taxonomically wide range of species involved. There is good documented fossil evidence for Osmunda and its ancestors (Scott 1920; Smith 1955), and since none of the accepted authorities in fossil botany have ever doubted that the primitive ferns of the Carboniferous are the direct ancestors of all the present-day Pterophyta, then it would be rash and unscientific to place any significance on its position in this tree. Indeed, if other alternative trees are investigated, in which the positions of all species except Osmunda regalis remain the same, then it is found that only one extra substitution is required to obtain several trees, summarised in Figure 20. This suggests that the position of Osmunda regalis in the tree is uncertain, due to an equal degree of dissimilarity, rather than similarity, with all other species in the tree. Two alternative explanations are obvious. Firstly, that plant evolution does not necessarily proceed

by the most parsimonious route, and one of the basic assumptions of this method is therefore false, or secondly a burst of evolution has taken place along the path to the higher ferns, so that the Osmunda regalis sequence is so unlike any other on the tree that it behaves as if it is unrelated to any, and consequently its true position is not indicated. It is possible that if the plastocyanin sequences from species more closely related to Osmunda regalis were obtained and included in the data set, then its position in the phylogenetic tree would become clear. The behaviour of Equisetum arvense illustrates just this point. If Equisetum sylvaticum is omitted from the data set, the position of Equisetum arvense is no longer fixed and several topologies are produced, in which its position varies slightly. However, when the sequence of Equisetum sylvaticum is included only one position provides the most parsimonious solution.

Should further information become available which gives more weight to the reliability of the topology represented in Figure 17, then several conclusions are possible, regarding the relative positions of the higher and primitive ferns, both to each other and to the remaining species in the tree. From the evidence of this tree alone, it would appear that the higher and primitive ferns did not share a common ancestry amongst the Carboniferous ferns, but have followed

separate lines of descent since the Psilophytes or earlier - an argument which must surely be rejected on the basis of the fossil evidence. If, as more widely believed, both groups of ferns shared a common ancestry in the Carboniferous ferns, then, since the time of their divergence, a rapid rate of molecular evolution has taken place along the line of the higher ferns. A reason for this could be a modification of the function of plastocyanin. If at the same time, a degree of convergent evolution has taken place within the molecule, so that the Osmunda regalis sequence has become more similar to the gymnosperm sequences, then the true phylogetic relationship between the two groups of ferns could have become obscured.

C. Gymnosperms

The occurrence of fossil gymnosperms is shown in Appendix III C, where no attempt has been made to show the hypothetical evolutionary routes because of the diversity of opinions as to their accuracy.

It is generally accepted that present-day gymnosperms represent two major groups, the Cycadopsida and Coniferopsida, plus a smaller group the Gnetales. It was only possible to obtain sequence information from one member of the Cycadopsida (Cycas revoluta) and two members of the Coniferopsida (Araucaria araucana and Taxus baccata).

The fossil record indicates all the groups of gymnosperms have been separate for a very long time and there has been much discussion as to whether these two major gymnosperm lines are monophyletic or diphyletic in origin, and where the ancestral gymnosperm groups (in particular the seed ferns or Pteridospermales) should fit into the evolutionary pathway.

Cronquist (1960) believes that the gymnosperms are diphyletic, the Cycadopsida and Coniferopsida having no common ancestor short of the Psilophytinae, with the Pteridospermales as the direct ancestors of the Cycads (Cronquist, 1968).

In contrast, Sporne (1967) has recently suggested that this long-held idea of a diphyletic origin may be incorrect, as it implies that the seed-habit evolved twice. He suggests the two groups may have had a common ancestor in the Progymnosperms, some members of which Smith (1955) recognises as members of the Primofilices. Sporne also suggests a possible blurring of distinction between the Cycadopsida and Coniferopsida, on the basis of Cordaitalean seed-structure (Smith, 1964). In addition, he refutes the arguments for linking the Pteridospermales with the Bennettitales or Cycads, as does Arnold (1953) who stresses that cycads are not modern Pteridosperms.

Consequently, most of the recent evidence regarding gymnosperm evolution has served to disprove links between ancestral and present-day gymnosperms, without producing any positive evidence.

The phylogenetic tree based on the amino acid sequences is not very useful. It confirms the divergence of Cycas revoluta from the main gymnosperm line of descent earlier than the two members of the Coniferopsida, but does not give reliable evidence as to when this happened. If the results are taken at their face value and if the divergence of the fern and gymnosperm node is assumed to be phylogenetically correctly placed then the two major groups of gymnosperms diverged late. However unlikely, it cannot be overlooked that an alternative explanation is possible i.e. that the geological period of the psilophytes corresponds to a point further along the branch leading to the gymnosperms (i.e. the higher ferns and gymnosperms had separate algal ancestors).

Even if further evidence allowed an accurate phylogenetic tree and time-scale to be established due to lack of data, many problems would remain and in practice it would not solve the problem, of which of the ancestral gymnosperm groups gave rise to the present-day species.

D. Angiosperms

At various times, nearly every group of gymnosperms has been proposed as the possible ancestral stock for the angiosperms. Present opinion falls into two main schools of thought. Some, like Strasburger (1978), Takhtajan (1969), Cronquist (1968) and Smith (1955), visualise the Pteridospermales as angiosperm ancestors, whereas others such as Hughes (1976) firmly believe that a closer inspection of the fossil evidence only reveals superficial similarities.

An investigation into the molecular evolution of plastocyanin from angiosperms was not part of this study; however, the sequences of daisy and magnolia were included in one exhaustive search (see Figure 19). Although two of the five trees (Figures 19A and B) showing maximum parsimony were in reasonable accord with the fossil evidence the remaining trees (Figures 19C to E) were so at variance with accepted biological ideas as to force the conclusion that the sequences of the angiosperms were again too different from the other species in the tree to make any conclusion possible.

Time Scales

In considering branch lengths as indications of evolutionary distances, i.e. as linear time scales, it must be possible to fix at least one evolutionary event, with some degree of accuracy, from the fossil record. In the phylogenetic tree of the lower plants, the most reliable event is the point of divergence of the Taxus baccata/ Araucaria araucana lines of descent. If this is assumed to be 175 million years ago (Sporne, 1967), a value for the unit evolutionary period (Ramshaw et al., 1972) of 25.4 million years, i.e. $\frac{175}{6.9}$ is obtained. Corresponding times of divergence for the other branches are shown in Table 17.

It is interesting to note that the divergence of the Cycas/Coniferopsida line would then have occurred approximately 240 million years ago, at the beginning of the Permian. This is later than most theories based on fossil evidence would allow. However, within the group of higher ferns, the results show some agreement with the fossil record since the Pteridium aquilinum/Dryopteris spp. divergence would have occurred approximately 150 million years ago - at a time when the first polypodiaceous ferns appeared in the fossil record. Further evolutionary lines within the higher ferns would then have appeared 90 and 60 million years ago, corresponding with the very beginning

of the Coenozoic, when the majority of fossil ferns of this type were beginning to appear (Smith, 1955).

The figure of 175 million years was based on the time when Taxus-like fossils appeared in the fossil record. However Araucarian type fossils are known as far back as the beginning of the Permian (Sporne,1967). If this reference is used the time value is extended to 250 million years. This would then give a value of 340 million years for the divergence of the Cycas revoluta/Coniferopsida line, a figure in good agreement with the ideas of Sporne (1967) and Strasburger et al. (1978). This figure would then cause the figures for fern evolution to be in disagreement with the fossil record (Smith, 1955).

If the many assumptions of this method are accepted, these results can be interpreted as an indication that different rates of evolution took place along the two separate lines of descent (the one to the ferns and the other to the gymnosperms). In the original method a correction was applied for back mutations (Margoliash & Fitch, 1968), which are more numerous over longer geological periods; however this was not considered worthwhile applying in this instance as the fossil datings were not so accurate.

Table 17

An exercise in establishing the approximate times of divergence of lower plant lines

The species considered are those used in the construction of the lower plant phylogenetic tree in which the branch lengths are assumed to be an indication of the length of time elapsed during the course of evolution of the respective species.

The point of reference to the fossil record is the divergence of the Araucaria/Taxus lines (see text).

* No correction has been made for back mutations.
** Times of divergence taken from fossil evidence.

Times of divergence of lower plant groups

Groups compared	Branch length taken	Time of divergence * if UEP = 25.4 (x 10^6 years)	Time of divergence * if UEP = 36.3 (x 10^6 years)
Araucaria/Taxus	6.9	175**	250**
Cycas/Coniferopsida	9.4	240	340
Osmunda/Gymnosperms	15.9	400	580
Equisetum/Osmunda	18.0	460	650
Algae/Equisetum/ Higher Ferns	23.0	580	830
Pteridium/ Dryopteris spp.	6.1	150	220
Dryopteris dilatata/ Dryopteris filix-mas	3.5	90	130
Dryopteris filix-mas/ Blechnum spicant	2.5	60	90

Validity of Methods of Molecular Phylogeny

Some classical taxonomists such as Cronquist (1976) strongly disagree with these methods of basing phylogenies on molecular data. Such people prefer to group species together on such grounds as the morphological similarity of a very few characters (from fossil or present-day organisms), whilst ignoring others, the choice of characters being totally dependent on the views of each individual taxonomist. These taxonomists will discount mathematical models of all kinds (which include the ancestral sequence method) and yet as Felsenstein (1975) pointed out, the mental processes involved in attributing greater significance to one morphological character rather than another are no more than sophisticated mathematical modelling in the form of (intuitive) statistical inference. Cronquist himself has almost admitted to this when he says, "As we develop our schemes of classification ..... we also develop of necessity a progressive bias in the interpretation of new evidence. The new evidence that fits a pre-existing scheme is considered to be important and confirmatory and that which does not fit is discarded as not being of taxonomic importance." (Cronquist, 1976). Despite saying this, Cronquist does not go on to explain the method he adopts to prevent this bias happening.

Mathematically-based models are mostly objective, and

consequently any assumptions or drawbacks in the method can be relatively easily distinguished, and even molecular evolutionists will admit that the method is not yet perfect. The thinking behind phylogenetic schemes based on fossil and morphological evidence however, is not as easily examined, being much more subjective. Because the drawbacks to this approach cannot be so easily spotted and because the method has been, of necessity, the only one used for many years, there tends to be a feeling that it must be right.

In grouping together species on the basis of similarity, for example, of ovule structure the classical taxonomist cannot easily distinguish between a similarity of characters due to a recent common ancestry and a similarity brought about simply by chance, convergence or the retention of primitive characters. This argument is particularly apt if fossil evidence is lacking or incomplete, as for example in the case of the angiosperms.

If five hypothetical species are examined, 3 of which have a similar ovule structure, X and two have a contrasting structure Y, then assuming this morphological characteristic is considered important by the classical taxonomist, the three species with ovule structure X will be grouped together and the two others in a separate group,

giving an impression of a phylogeny such as in Figure 34A. This may not be accurate. A possible alternative explanation is shown in Figure 34B. If the similarity in structure is due to a common ancestor, as in the case of the two Y individuals, then the assumption of the previous tree is correct. However, X may be the characteristic found in the original common ancestor of all five individuals in which case, the similarity of the two X individuals 1 and 2 in Figure 36B is due to retention of the primitive character (making them as distantly related to each other as they are to the Y individuals) and back mutation in individual 3.

Cronquist (1976) uses a similar model, but without substantiating his argument in any way states that the ancestral sequence method does not detect back mutations, in an attempt to give a general impression that his methods are far superior. A critical analysis of the two methods quickly reveals that the reverse is more likely to be true. Although not all back mutations are detected by the ancestral sequence method, given sequences from species which are closely related, then the ancestral sequence method is very efficient, because of the objective way in which a large amount of information is handled. For example, if each of the amino acids found in the N-terminal sequences of plastocyanins (Figure 14) are thought of instead as 40 different

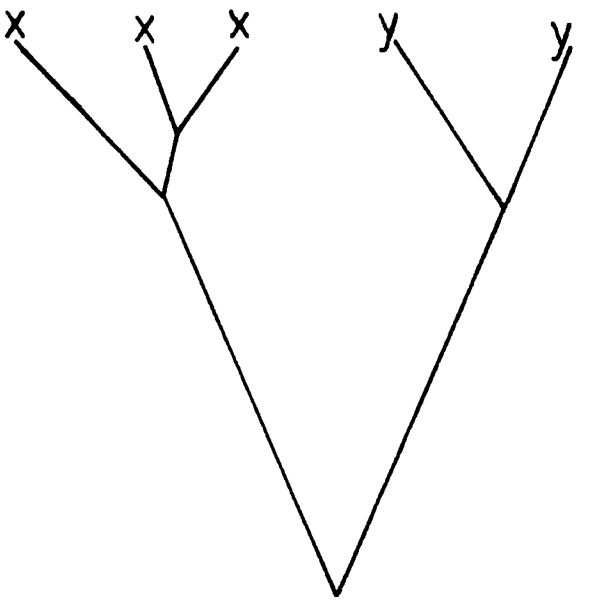
FIGURES 34A and B

Interpretation of fossil evidence

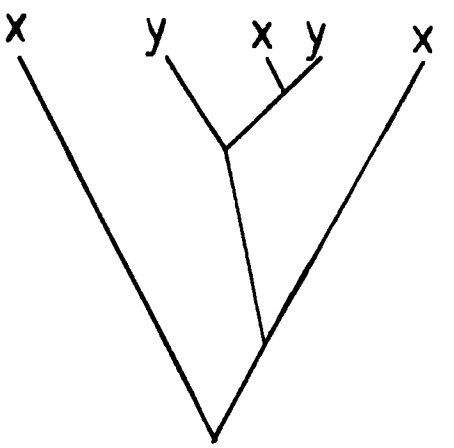
Two possible alternative phylogenies based on two hypothetical variations of ovule structure, x and y.

For further details, see text.

A



B



morphological characteristics such as leaf venation, ovule structure etc. which classical taxonomists utilise, then it is clear that the human mind could make very little sense of the data as it stands. Some characters would have to be considered more important than others. For example the Osmunda regalis sequence appears to be more similar to the Equisetum spp. sequences than those of the higher ferns. A subjective approach would then give the characters linking Osmunda to the Equisetaceae (e.g. positions 4, 8, 9 etc) less importance than those linking it to the higher ferns (e.g. positions 19, 29). In the analysis by the ancestral sequence method such subjective judgements were not made and although Osmunda was consequently not placed in the 'accepted' position in the phylogenetic tree, it was in the correct position as indicated by the results.

REFERENCES

Aitken, A. (1975) Biochem. J. 149, 675-683.

Aitken, A. (1976) Nature 263, 793-796.

Aitken, A. (1977) Evolution of Protein Molecules. Proceedings of the Symposium on Evolution of Protein Molecules (Matsubara, H. & Yamanaka, T. Eds.), 251-263. Japan Sci. Soc. Press, Tokyo.

Ambler, R.P. (1971) Recent Developments in the Chemical Study of Protein Structures, 289-305 Proceedings of Colloque Montpellier, organised by A. Previero, J.F. Pechere and M.A. Coletti-Previero. Inserm, Paris.

Amundsen, A.R., Whelan, J. & Bosnich, B. (1977) J. Amer. Chem. Soc. 99, 6730-6739.

Anderson, E. (1934) Nature 133, 462.

Arnold, C.A. (1953) Phytomorph. 3, 51-65.

Bartnicki-Garcia, S. (1970) In 'Phytochemical Phylogeny' (Harborne, J.B. Ed.), 81-103 Academic Press, London.

Beckman (1972) Instruction Manual for Model 890C Sequencer. Beckman Instruments, Palo Alto.

Beckman (1973) "In Sequence", issue No. 4, p.1. Beckman Instruments, Palo Alto.

Bishop, N.I. (1971) Annu. Rev. Biochem. 40, 197-726.

Blömback, B., Blömback, M., Edman, P. & Hessel, B. (1966) Biochim. Biophys. Acta 115, 371-376.

Blumberg, W.E. (1966) In 'The Biochemistry of Copper' (Peisach, J., Aisen, P. & Blumberg, W.E. Eds.), 49-64. Academic Press, New York.

Blumberg, W.E. & Peisach, J. (1966) Biochim. Biophys. Acta 126, 269-273.

Böhlen, P., Stein, S. & Udenfriend, S. (1974) Arch. Biochem. Biophys. 163, 390-399.

Borchert, M.T. & Wessels, J.S.C. (1970) Biochim. Biophys. Acta. 197, 78-83.

Boulter, D. (1973a) Syst. Zool. 22, 549-553.

Boulter, D. (1973b) Pure and Applied Chem. 34, 539.

Boulter, D. (1974) Current Advan. Plt. Sci. 8, 1-16.

Boulter, D. & Peacock, D. (1975) Proceedings of the Eighth International Conference on Numerical Taxonomy (Estabrook, G.F. Ed.). 176-189 W.H. Freeman & Co., San Francisco.

Boulter, D., Haslett, B.G., Peacock, D., Ramshaw, J.A.M. & Scawen, M.D. (1977) In 'Plant Biochemistry II' International Review of Biochemistry Vol. 13 (Northcote, D.H., Ed.), 1-40. University Park Press, Baltimore.

Boulter, D., Ramshaw, J.A.M., Thompson, E.W., Richardson, M. & Brown, R.H. (1972) Proc. Roy. Soc. Lond. B 181, 441-455.

Bower, F.O. (1935) Primitive Land Plants Macmillan & Co. London.

Brill, A.S. & Bryce, G.F. (1968) J. Chem. Phys. 48, 4398-4404.

Church, A.H. (1919) Thalassiophyta and the Subaerial Transmigration. Botanical Memoirs No. 3. Oxford University Press.

Colman, P.M., Freeman, H.C., Guss, J.M., Murata, M., Norris, V.A., Ramshaw, J.A.M. & Venkatappa, M.P. (1978) Nature (London) 272, 319-324.

Crestfield, A.M., Moore, S. & Stein, W.H. (1963) J. Biol. Chem. 238, 622-627.

Cronquist, A. (1960) Bot. Rev. 26, 426-482.

Cronquist, A. (1968) The Evolution and Classification of Flowering Plants. Thos. Nelson & Sons Ltd., London.

Cronquist, A. (1976) Brittonia 28, 1-27.

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

Dayhoff, M.O. (1972) Atlas of Protein Sequence and Structure Vol. 5. National Biomedical Research Foundation, Silver Springs, Maryland.

Dickerson, R.E., Takano, T., Eisenberg, D., Kallai, O.B. Samson, L., Cooper, A. & Margoliash, E. (1971) J. Biol. Chem. 246, 1511-1535.

Easley, C.W. (1965) Biochim. Biophys. Acta. 107, 386-388.

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

Elmore, D.T. & Toselan, P.A. (1956) J. Chem. Soc. 188-196.

Felsenstein, J. (1975) In discussion during Proceedings of the Eighth International Conference on Numerical Taxonomy (Estabrook, G.F., Ed.). W.H. Freeman & Co., San Francisco.

Fitch, W.M. (1970) Syst. Zool. 20, 406-416.

Fitch, W.M. & Margoliash, E. (1967) Science 155, 279-284.

Fitch, W.M. & Markowitch, E. (1970) Biochem. Genetics 4, 579-593.

Freeman, H.C., Norris, V.A., Ramshaw, J.A.M. & Wright, P.E. (1978) FEBS Lett. 86, 131-135.

Gorman, D.S. & Levine, R.P. (1966) Plant Physiol. 41, 1637-1642.

Gray, W.R. (1967) Methods Enzymol. 11, 139-151.

Gray, W.R. & Hartley, B.S. (1963a) Biochem. J. 89, 59P.

Gray, W.R. & Hartley, B.S. (1963b) Biochem. J. 89, 379-380.

Graziani, M.T., Agro, A.F., Rotillio, G. Barra, D. & Mondovi, B. (1974) Biochemistry 13, 804-809.

Hare, J.W., Solomon, E.I. & Gray, H.B. (1976) J. Amer. Chem. Soc. 98, 3205-3209.

Haslett, B.G. & Boulter, D. (1976) Biochem. J. 153, 33-38.

Haslett, B.G., Gleaves, T. & Boulter, D. (1977) Phytochemistry 16, 363-365.

Heilmann, J., Barrollier, J. & Watzke, C. (1957) Hoppe-Seyler's Z-Physiol. Chem. 309, 219.

Hughes, N.F. (1976) Palaeobiology of Angiosperm Origins. Cambridge University Press.

Ilse, D. & Edman, D.E. (1963) Aust. J. Chem. 16, 411-416.

Inglis, A., Nicholls, P. & Roxburgh, C. (1971) Aust. J. Biol. Sci. 24, 1247-1250.

Jeffrey, E.C. (1902) Phil. Trans. Roy. Soc. Lond. B 195, 119-146.

Jeppsson, J.O. & Sjöquist, J. (1967) Analyt. Biochem. 18, 264-269.

Jones, T.G., Rorabacher, D.B. & Ochrymowicz, L.A. (1975) J. Amer. Chem. Soc. 97, 7485-7486.

Kasper, C.B. (I970) In 'Molec. Biol. Biochem. and Biophys.' (Needleman, S.B., Ed.) 8, 137-184. Chapman and Hall.

Katoh, S. (1960) Nature (London) 186, 533-534.

Katoh, S. & Takamiya, A. (1964) J. Biochem. (Tokyo) 55, 378-387.

Katoh, S., Shiratori, I., Takamiya, A. (1962) J. Biochem. (Tokyo) 51, 32-40.

Kelly, J. & Ambler, R.P. (1974) Biochem. J. 143, 681-690.

Laursen, R.A. (1975) In 'Solid Phase Methods in Protein Sequence Analysis' Proceedings of First International Conference (Laursen, R.A. Ed.) Pierce Chemical Company, Rockford.

Levin, D.A. & Wilson, A.C. (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 2086-2090.

Lightbody, J.J. & Krogmann, D.W. (1967) Biochim. Biophys. Acta 131, 508-515.

Liu, T.-Y. & Cheng, Y.H. (1971) J. Biol. Chem. 246, 2843-2848.

Loomis, W.D. & Battaile, J. (1966) Phytochemistry 5, 423-438.

Lyddiatt, A. (1975) Ph.D. Thesis, University of Durham.

Lyddiatt, A., Peacock, D. & Boulter, D. (1978) J. Mol. Evol. 11, 35-45.

Malcolm, D.B. (1978) Nature 275, 90-91.

Malkin, R. & Malmström, B.G. (1970) Adv. Enzymol. 33, 177-244.

Markley, J.K., Ulrich, E.L., Berg, S.P. & Krogmann, D.W. (1975) Biochemistry 14, 4428-4433.

Margoliash, E. & Fitch, W.M. (1968) Ann. N.Y. Acad. Sci. 151, 359.

Matsubara, H. & Sasaki, R.M. (1969) Biochem. Biophys. Res. Comms. 35, 175-181.

Matsubara, T., Hase, T., Wakabayashi, S. & Wada, K. (1977) Evolution of Protein Molecules. Proceedings of the Symposium on Evolution of Protein Molecules (Matsurbara, H and Yamanaka, T. Eds.) 209, Japan Sci. Soc. Press, Tokyo.

Mayer, G., Linsley, E.G. & Usinger, R.L. (1953) Methods and Principles of Systematic Zoology McGraw-Hill, New York.

Meatyard, B. (1974) Ph.D. Thesis, University of Durham.

Mendez, E. & Lai, C.Y. (1975) Anal. Bioch. 65, 281-292.

Milne, P.R. & Wells, J.R.G. (1970) J. Biol. Chem. 245, 1566-1574.

Moore, G.R. & Williams, R.J.P. (1976) Coord. Chem. Rev. 18, 125-197.

Moore, G.W. & Goodman, M. (1977) J. Mol. Evol. 9, 121-130.

Moore, S. & Stein, W.H. (1963) Methods Enzymol. 6, 819-831.

Needleman, S.B. & Wunsch, C.D. (1970) J. Mol. Biol. 48, 443-453.

Newcomb, E.H. (1963) Ann. Rev. Plant Physiol. 14, 43-64.

Offord, R.E. (1966) Nature (London) 211, 591-593.

Peacock, D. & Boulter, D. (1975) J. Mol. Biol. 95, 513-527.

Peisach, J., Levine, W.G. & Blumberg, W.E. (1967) J. Biol. Chem. 242, 2847-2858.

Percy, M.E. & Buchwald, B.M. (1972) Anal. Biochem. 45, 60-67.

Pirie, N.W. (1959) Ann. Rev. Plant Physiol. 10, 33-52.

Pisano, J.J., Bronlert, T.J. & Brewer, H.B. (1972) Anal. Biochem. 45, 43-60.

Plesnicar, M. & Bendall, D.S. (1970) Biochim. Biophys. Acta 216, 192-199.

Ramachandran, J. & Witkop, B. (1967) Methods Enzymol. 11, 283-299.

Ramshaw, J.A.M. (1972) Ph.D. Thesis, University of Durham.

Ramshaw, J.A.M., Thompson, E.W. & Boulter, D. (1970) Biochem. J. 119, 535-539.

Ramshaw, J.A.M., Brown, R.H., Scawen, M.D. & Boulter, D. (1973) Biochim. Biophys. Acta 303, 269-273.

Ramshaw, J.A.M., Richardson, D.L., Meatyard, B.T., Brown, R.H., Richardson, M., Thompson, E.W. & Boulter, D. (1972) New Phytol. 71, 773-779.

Raven, J.A. (1977) Advances in Botanical Research 5, 153-219.

Romer, A.S. (1945) Vertebrate Palaentology 2nd Ed., University of Chicago Press.

Ryden, L. & Lundgren, J.O. (1976) Nature (London) 261, 344-346.

Ryle, A.P., Sanger, F., Smith, L.F., Kitai, R. (1955) Biochem. J. 60, 541-556.

Scawen, M.D., Hewitt, E.J. & James, D.M. (1975) Phytochem. 1225-1233.

Scott, D.H. (1909) Studies in Fossil Botany, 2nd Ed. Vol. 2. London .

Scott, D.H. (1920) Studies in Fossil Botany, 3rd Ed. A.C. Black Ltd., London.

Simpson, G.G. (1961) Principles of Animal Taxonomy. Columbia University Press, New York.

Smith, D.L. (1964) Biol. Rev. 39, 137-159.

Smith, G.M. (1955) Cryptogamic Botany, Vol II. McGraw-Hill, London.

Sokal, R.R. & Sneath, P.H.A. (1963) Principles of Numerical Taxonomy. Freeman & Co. San Francisco.

Spiker, S. (1975) Biochim. Biophys. Acta. 400, 461-467.

Sporne, K.R. (1967) The Morphology of Gymnosperms Hutchinson & Co. Ltd., London.

Sporne, K.R. (1974) The Morphology of Pteridophytes Hutchinson & Co. Ltd., London.

Stahmann, M.A. (1963) Ann. Rev. Plant Physiol. 14, 137-158.

Strasburger, E. (1978) Lehrbuch der Botanit 31 Auflage (D. von Denffler, F. Ehrendorfer, K. Magdefrau & H. Zieglar, Eds.) Fischer Verlag Press, Stuttgart, New York.

Swain, T. (1974) Comprehensive Biochemistry (Florkin, M. & Stotz, E.H., Eds.) 29A, Elsevier, Amsterdam. 125-302.

Takhtajan, A. (1969) Flowering Plants - Origin and Dispersal Oliver & Boyd, Edinburgh.

Takruri, I. (1979) Ph.D. Thesis, University of Durham.

Thorne, R.F. (1968) Aliso 6, 57-66.

Trebst, A. (1974) Ann. Rev. Plant Physiol. 25, 423-458.

Udenfriend, S., Stein, S., Böhlen, P., Dairman, W., Leimgruber, W., Weigele, M. (1972) Science 178, 871-872.

Valentine, J. (1976) Ph.D. Thesis, University of Durham.

Vandekerckove, J. & Montagu, M.V. (1974) Eur. J. Biochem. 44, 279-288.

Vaughan, J.G. (1968a) Science Progress (Oxf.) 56, 205-222.

Vaughan, J.G. (1968b) In 'Chemotaxonomy and Serotaxonomy' (Hawkes, J.G., Ed.). Systematics Association, Special Vol. No. 2, 103-110.

Wallace, D.G. (1976) Biophys. Chem. 4, 123-130.

Wang, T.T. & Young, N.M. (1977) Biochim. Biophys. Res. Comm. 74, 119-125.

Watson, E.V. (1968) British Mosses and Liverworts Cambridge University Press.

Watts, R.L. (1970) 'Proteins and Plant Phylogeny'. In Phytochemical Phylogeny, 145-177 (Harborne, J.B., Ed.)

Wood, W.B., Wilson, J.H., Benbow, R.M. & Hood, L.E. (1974) Biochemistry - A Problems Approach W.A. Benjamin, Inc., Menlo Park, California.

Woods, K.R. & Wang, K.T. (1967) Biochim. Biophys. Acta 133, 369-370.

Yamanaka, T. & Okunuki, K. (1963) Biochim. Biophys. Acta 67, 379.

Yanofsky, C. (1967) Scientific American 216, 80-94.

Zuckerkandl, E. & Pauling, L. (1965a) "Evolving Genes and Proteins" (Bryson, V and Vogel, H.J., Eds.) Academic Press, New York.

Zuckerkandl, E. & Pauling, L. (1965b) J. Theoret. Biol. 8, 357-366.

APPENDIX I

Classification of species used as sources of plastocyanin

IA Bryophytes

IB Pteridophytes

IC Gymnosperms

APPENDIX IA

Classification of Bryophytes used as Sources of Plastocyanin (Watson, 1968)

Bryophyta	Hepaticae	Marchantiales	Marchantiaceae	Conocephalum conicum
				Marchantia polymorpha
	Musci			
	a) Sphagnidae	Sphagnales	Sphagnaceae	Sphagnum spp.
	b) Bryidae	Polytrichales	Polytrichaceae	Polytrichum spp.
		Eubryales	Mniaceae	Mnium hornum
		Isobryales	Fontinalaceae	Fontinalis antipyretica

APPENDIX IB

Classification of Pteridophytes used as Sources of Plastocyanin (Smith, 1955)

Lepidophyta	Lycopodinae	Selaginellales	Selaginellaceae	Selaginella spp.	
Calamophyta	Equisetinae	Equisetales	Equisetaceae	Equisetum arvense	Field Horsetail
				Equisetum sylvaticum	Wood Horsetail
Pterophyta	Filicinae				
	a) Eusporangiatae	Marattiales	Angiopteridaceae	Angiopteris palmiformis	
	b) Leptosporangiatae	Filicales	Osmundaceae	Osmunda regalis	Royal Fern
			Dicksoniaceae	Dicksonia antarctica	
			Polypodiaceae	Thelypteris erubescens	
				Pteridium aquilinum	Bracken
				Pteris cretica	
				Blechnum spicant	Hard Fern
				Phyllitis scolopendrium	Hart's Tongue
				Dryopteris filix-mas	Male Fern
				Dryopteris dilatata	Greater Buckler Fern
		Salviniales	Salviniaceae	Salvinia natans	

APPENDIX IC

Classification of Gymnosperms used as Sources of Plastocyanin (Sporne, 1965)

Gymnosperms	Cycadopsida	Cycadales	Cycadaceae	Cycas revoluta	Cycad
	Coniferopsida	Coniferales	Pinaceae	Abies grandis	Grand fir
				Picea abies	Norway Spruce
				Larix decidua	European Larch
				Pinus laricio	Corsican Pine
			Taxodiaceae	Sequoia sempervirens	Redwood
			Cupressaceae	Thuja plicata	Western Red Cedar
				Cupressus lawsonia	Lawsons Cypress
			Araucariaceae	Araucaria araucana	Monkey Puzzle
		Taxales	Taxaceae	Taxus baccata	Yew
		Ginkgoales	Ginkgoaceae	Ginkgo biloba	Maiden hair tree
	Gnetopsida	Gnetales	Ephedraceae	Ephedra spp.	

APPENDIX II

Detailed results of N-terminal plastocyanin sequences, with results of analyses of each sequencer sample.

N-terminal amino acid sequence of Equisetum sylvaticum plastocyanin with the results of analyses of each sample from the sequencer

Sequence Position	T.l.c. identification	G.l.c. identification	Identification after regeneration	Residue
1	Ala	Ala (258.4), Gly (60.0), Val (24.0), Leu/Ile (21.0), Pro (9.4)		Alanine
2	Glu*,Asp		Glu	Glutamic acid
3	Val*,Glu	Val (392.3), Gly (80.4), Ala (60.5), Pro (33.4), Leu/Ile (10.2)		Valine
4	Ile	Ile (244.7), Val (87.3), Pro (55.5), Gly (54.5), Ala (24.7)		Isoleucine
5	Met	Met* (not standardised), Leu/Ile (24.3), Val (21.1), Gly (20.1), Pro (11.1)		Methionine
6	Gly	Gly (219.2), Val (22.2), Pro (21.7), Leu (19.38), Ala (13.1)		Glycine
7	Leu*,Gly	Leu (99.3), Gly (54.5), Val (17.5), Pro (17.1), Ala (12.4)		Leucine
8	Asp		Asp*,Gly	Aspartic acid
9	Asp*,Ala		Asp*,Ala, Gly	Aspartic acid

10	Gly*,Asp	Gly (109.6), Val (20.3), Leu/Ile (14.5), Ala (10.5), Pro (9.0)		Glycine
11	Ser*,Gly Val		Ala*,Gly	Serine
12	Leu	Leu (87.6), Val (8.45), Pro (3.9)		Leucine
13	Lys*,Glu		Lys	Lysine
14	Phe	Phe (45.0), Val (8.2), Ala (4.7), Leu (4.4), Pro (1.4)		Phenylalanine
15	Leu*,Phe	Leu (80.2), Val (8.7), Pro (4.3)		Leucine
16	Pro	Pro (18.4), Leu/Ile (16.8), Val (8.7)		Proline
17	Asp*,Pro		Asp*,Pro	Aspartic acid
18	Lys*,Asp		Lys*,Asp	Lysine
19	Val*,Lys	Val (25.6), Pro (3.1), Leu/Ile (2.8)		Valine
20	Glu*,Val		Glu	Glutamic acid
21	Val*,Glu	Val (34.8), Leu/Ile (4.0), Pro (2.2)		Valine
22	Lys*,Val, Glu		Lys*	Lysine
23	Ala*,Lys	Ala (22.3), Val (10.7), Pro (3.5) Leu/Ile (3.1)		Alanine
24	Gly*,Ala Lys,Glu	Gly (23.8), Ala (8.9), Val (7.3), Leu/Ile (3.2), Pro (2.8)		Glycine

25	Glu*,Gly		Glu*,Gly	Glutamic acid
26	Lys*,Glu		Lys	Lysine
27	Ile*,Lys, Glu	Leu/Ile (16.3), Val (5.0), Pro (2.4)	Ile*,Gly	Isoleucine
28	Thr*,Ile, Lys		α-amino-butyrate	Threonine
29	Asn,Ser,Thr		Asp,α-aminobutyrate	Asparagine
30	Val	Val (25.2), Pro (5.3), Leu/Ile (4.6), Ala (3.6)		Valine
31	Asn*,Val		Asp	Asparagine
32	Phe*,Asn	Phe (10.4), Leu/Ile (4.6), Val (4.0), Pro (3.9)		Phenylalanine
33	Ala*,Asn	Ala (17.2), Val (5.7), Leu/Ile (3.0), Pro (2.2)		Alanine
34	Gly*,Ala	Gly (10.3), Ala (4.8), Val (4.2), Leu/Ile (2.5), Pro (1.1)		Glycine
35	Phe*,Gly,Ala	Phe (6.3), Val (2.5), Leu/Ile (2.2), Pro (1.9)		Phenylalanine
36	Pro*,Phe	Pro (5.5), Val (3.5), Leu/Ile (3.2)		Proline
37	Nothing		His	Histidine
38	Asn		Asp	Asparagine

39	Val,Asn	Val (11.0), Pro (5.2), Leu (4.9)	Valine
40	Val	Val (8.0), Leu/Ile (2.0), Pro (1.5)	Valine

N-terminal amino acid sequence of Equisetum arvense plastocyanin with the results of analyses of each sample from the sequencer

Sequence Position	T.l.c. identification	G.l.c. identification	Identification after regeneration	Residue
1	Ala	Ala (248.7), Pro (20.0), Gly (15.7), Val (14.9), Leu/Ile (9.0)		Alanine
2	Glu	Glu (136.5)		Glutamic acid
3	Val	Val (190.0), Ala (19.0), Pro (7.6), Leu/Ile (4.9)		Valine
4	Ile	Ile (197.5), Val (39.2), Ala (8.4)		Isoleucine
5	Met	Met* (not standardised), Leu/Ile (30.17), Val (18.7), Ala (5.1), Gly (4.9)		Methionine
6	Gly	Gly (287.5), Val (11.8), Leu/Ile (9.9), Pro (7.5), Ala (6.3)		Glycine
7	Leu	Leu (175.4), Gly (21.3), Ala (13.4), Val (5.9)		Leucine
8	Asp	Asp (80.6)		Aspartic acid
9	Asp	Asp (183.7)		Aspartic acid

10	Gly*,Asp	Gly (208.0), Leu/Ile (14.8), Val (9.8), Pro (3.8), Ala (3.2)		Glycine
11	Ser*,Thr Gly		Ala*,Gly	Serine
12	Leu	Leu (117.8), Gly (17.8), Val (8.7), Ala (1.9)		
13	Glu,Gln*		Glu*,Gly, Lys	Glutamine
14	Phe	Phe (67.8), Leu/Ile (18.7), Gly (9.1), Val (4.9), Ala (4.0)		Phenylalanine
15	Asn*,Phe		Asp*,Phe	Asparagine
16	Pro	Pro (39.5), Leu/Ile (7.4), Gly (6.4), Val (5.9), Ala (1.1)		Proline
17	Lys*,Glu, Pro		Lys*,Pro,Gly	Lysine
18	Glu*,Lys	Glu (45.9)		Glutamic acid
19	Val*,Glu	Val (60.0), Leu/Ile (11.1), Gly (10.5), Pro (10.0), Ala (3.0)		Valine
20	Val	Val (65.9), Leu/Ile (10.0), Pro (4.4)		Valine
21	Val	Val (68.9), Gly (7.0), Leu/Ile (5.5), Pro (4.4), Ala (1.7)		Valine
22	Lys*,Val		Lys*,Val	Lysine

23	Ala*,Val, Lys	Ala (31.0), Val (12.3), Pro (3.5), Leu/Ile (3.0)		Alanine
24	Gly*,Ala	Gly (39.6), Ala (16.8), Val (10.0), Pro (5.7), Leu/Ile (3.3)		Glycine
25	Glu*,Lys	Glu (27.2)		Glutamic acid
26	Glu,Lys*		Lys*,Glu	Lysine
27	Ile	Ile (19.1), Val (7.5), Gly (7.4), Ala (3.2)		Isoleucine
28	Thr*,Ser	Gly (14.3), Pro (8.0), Val (8.0), Leu/Ile (6.13), Ala (2.6)	α-amino-butyrate	Threonine
29	Phe	Phe (18.3), Leu/Ile (6.8), Val (5.1), Pro (5.1)		Phenylalanine
30	Ile*,Phe	Ileu (24.0), Val (6.7), Pro (5.6), Gly (3.5), Ala (3.3)		Isoleucine
31	Asn*,Phe, Leu/Ile		Asp	Asparagine
32	Asn		Asp	Asparagine
33	Ala*,Asn	Ala (16.5), Pro (6.5), Leu/Ile (6.1), Val (5.3), Gly (5.1)		Alanine
34	Gly*,Ala, Asn	Gly (8.5), Ala (8.1), Val (5.3), Leu/Ile (4.4), Pro (1.0)		Glycine

35	Phe*,Gly, Ala	Phe (9.6), Gly (7.0), Val (6.3), Ala (5.4), Leu/Ile (4.4), Pro (1.8)		Phenylalanine
36	Pro*,Phe	Pro (5.6), Gly (5.8), Val (4.6), Leu/Ile (3.8), Ala (2.8)	Pro*,Phe	Proline
37	Nothing		His	Histidine
38	Asn*,Gly, Ala		Asp	Asparagine
39	Val*,Asn	Val (10.7), Gly (6.5), Pro (4.1), Ala (3.7), Leu/Ile (3.4)		Valine
40	Val*,Asn	Val (17.5), Gly (4.3), Pro (4.2), Leu/Ile (3.3), Ala (2.5)		Valine

N-terminal amino acid sequence of Osmunda regalis plastocyanin with the results of analyses of each sample from the sequencer

Sequence Position	T.l.c. identification	G.l.c. identification	Identification after regeneration	Residue
1	Ala	Ala (167.6), Val (4.3), Leu/Ile (3.3)		Alanine
2	Asp		Asp	Aspartic acid
3	Val	Val (220.5), Leu/Ile (17.3)		Valine
4	Ile	Ile (219.6), Val (5.8)		Isoleucine
5	Met	Met* (not standardised), Leu/Ile (8.6), Val (7.8), Pro (3.8)		Methionine
6	Gly	Gly (262.5), Val (7.1), Leu/Ile (6.8), Ala (4.4), Pro (3.7)		Glycine
7	Gly	Gly (269.2), Ala (10.1), Leu/Ile (7.5), Pro (5.7), Val (3.9)		Glycine
8	Asp		Asp* Gly	Aspartic acid
9	Asp		Asp	Aspartic acid

10	Gly	Gly (161.8), Leu/Ile (6.0), Pro (2.6), Val (2.5), Ala (2.1)		Glycine
11	Ser*,Thr		Ala	Serine
12	Leu	Leu (74.0), Val (3.4), Ala (3.0), Pro (2.9)		Leucine
13	Ala	Ala (111.8), Leu/Ile (9.0), Val (3.3), Pro (2.1)		Alanine
14	Phe	Phe (59.5), Leu/Ile (6.5), Ala (5.8), Val (5.4), Pro (2.8)		Phenylalanine
15	Ile	Ile (105.1), Val (5.4), Ala (3.1), Pro (1.3)		Isoleucine
16	Pro	Pro (68.6), Leu/Ile (6.6), Val (4.6), Ala (2.2)		Proline
17	Asn*,Asp		Asp*,Pro	Asparagine
18	Lys*,Asn		Lys	Lysine
19	Ile*,Lys	Ile (74.0), Val (15.8)		Isoleucine
20	Val	Val (83.7), Leu/Ile (19.2), Pro (2.7)		Valine
21	Val	Val (95.8), Leu/Ile (12.7), Pro (2.7)		Valine
22	Ser*,Thr,Val		Ala*,Gly	Serine
23	Val*,Ser	Val (97.6), Leu/Ile (17.8)		Valine

24	Gly	Gly (84.6), Val (22.1), Leu/Ile (7.9), Pro (3.5)		Glycine
25	Glu		Glu*,Gly	Glutamic acid
26	Pro*,Phe	Pro (126.6), Val (14.0)		Proline
27	Ile*,Pro	Ile (69.5), Pro (17.0), Val (4.9)		Isoleucine
28	Thr*,Lys		α-amino-butyrate	Threonine
29	Phe*,Lys,Thr	Phe (26.6), Leu/Ile (16.9), Val (12.3) Pro (5.7)		Phenylalanine
30	Lys*,Phe		Lys*,Phe	Lysine
31	Asn*,Lys,Phe		Asp*,Lys	Asparagine
32	Asn*,Lys		Asp*,Lys	Asparagine
33	Ala*,Asn	Ala (38.3), Leu/Ile (6.2), Val (6.2), Pro (3.1)		Alanine
34	Gly*,Ala	Gly (28.3), Ala (19.1), Val (7.9), Leu/Ile (6.1), Pro (4.4)		Glycine
35	Phe*,Gly,Ala	Phe (24.4), Leu/Ile (6.1), Ala (5.8) Pro (4.6), Val (4.6)		Phenylalanine
36	Pro*,Phe	Pro (20.3), Leu/Ile (5.9), Val (4.9) Ala (3.1)		Proline
37	Nothing		His	Histidine

38	Asn*,Asp		Asp*,Pro	Asparagine
39	Val	Val (25.4), Leu/Ile (10.5), Pro (10.4)		Valine
40	Val*,Asn	Val (66.1), Pro (11.4), Leu/Ile (9.75)		Valine

N-terminal amino acid sequence of Pteridium aquilinum plastocyanin with the results of analyses of each sample from the sequencer

Sequence Position	T.l.c. identification	G.l.c. identification	Identification after regeneration	Residue
1	Ala	Ala (343.1), Gly (113.4), Val (10.5), Leu/Ile (10.4), Pro (5.9)		Alanine
2	Lys		Lys	Lysine
3	Val	Val (178.7), Ala (86.4), Leu/Ile (5.9), Pro (5.9)		Valine
4	Glu	Glu (243.8)		Glutamic acid
5	Val	Val (120.6), Leu/Ile (39.3), Pro (24.1), Ala (8.2)		Valine
6	Gly	Gly (129.6), Val (23.0), Pro (5.1), Leu/Ile (7.1), Phe (55.6)		Glycine
7	Asp	Asp (58.9), Glu (13.4)		Aspartic acid
8	Glu*,Leu,Lys	Glu (103.5), Asp (17.4)		Glutamic acid
9	Val	Val (93.8), Leu/Ile (15.0), Pro (40.2), Gly (28.8), Ala (37.5)		Valine
10	Gly	Gly (87.5), Leu/Ile (12.5), Pro (17.6) Val (20.3), Ala (11.5)		Glycine

11	Ser		Ala*,Gly,Phe	Serine
12	Phe	Phe (46.8), Leu/Ile (16.9), Pro (18.1) Val (15.1), Gly (25.0), Ala (18.3)		Phenylalanine
13	Lys*,Asp		Lys,Phe	Lysine
14	Phe*,Lys	Phe (77.5), Leu/Ile (12.1), Pro (11.75), Val (8.5), Gly (14.5), Ala (18.04)		Phenylalanine
15	Thr*,Asp		α-amino-butyrate*, Phe,Gly	Threonine
16	Pro	Pro (61.8), Leu/Ile (11.7), Val (7.3) Gly (21.05), Ala (7.6)		Proline
17	Asp*,Asn	Asp (115.1)		Aspartic acid
18	Thr*,Ser	Pro (36.9), Gly (17.6), Leu/Ile (13.2), Val (11.4), Ala (10.3)	α-amino-butyrate*, Gly-Asp	Threonine
19	Leu/Ile*	Ile (70.1), Pro (10.6), Val (10.5), Gly (4.0), Ala (6.15)		Isoleucine
20	Thr*/Ser	Pro (31.2), Gly (25.3), Ala (16.1), Leu/Ile (14.9), Val (9.0)	α-amino-butyrate*, Gly	Threonine
21	Val	Val (60.9), Leu/Ile (20.1), Pro (18.3), Gly (34.5), Ala (28.2)		Valine

22	Ala	Ala (56.0), Leu/Ile (8.30), Pro (7.5) Val (13.6), Gly (6.7)		Alanine
23	Ala	Ala (60.9), Leu/Ile (13.35), Pro (11.75), Val (16.3), Gly (20.0)		Alanine
24	Gly*,Ala	Gly (70.0), Leu/Ile (15.8), Pro (16.9), Val (16.9), Ala (32.175)		Glycine
25	Glu	Glu (87.0)		Glutamic acid
26	Ala	Ala (26.9), Leu/Ile (8.4), Pro (6.5), Val (6.4), Gly (5.5)		Alanine
27	Ile*/Leu	Ile (51.9), Pro (14.8), Val (18.6), Gly (15.4), Ala (31.5)		Isoleucine
28	Glu	Glu (53.9)		Glutamic acid
29	Phe	Phe (34.4), Leu/Ileu (9.3), Pro (5.9), Val (5.8), Ala (5.1)		Phenylalanine
30	Thr*,Lys		α-amino butyrate*, Gly,Phe,Leu	Threonine
31	Leu*/Ileu	Leu (41.5), Pro (8.0), Val (16.1), Ala (14.62)		Leucine
32	Val	Val (38.3), Leu/Ile (10.04), Pro (3.0), Ala (1.3)		Valine

33	Gly*,Val,Ala	Gly (22.2), Leu/Ile (6.6), Pro (4.0), Val (10.9), Ala (3.5)		Glycine
34	Glu	Glu (19.8)		Glutamic acid
35	Thr*,Ser,Lys, Glu		α-amino-butyrate*, Glu	Threonine
36	Gly*,Thr,Glu			? Glycine
37	Nothing		His	Histidine
38	Asn		Asp*,Val	Asparagine
39	Val*,Asn	Val (29.2), Leu/Ile (3.2), Pro (3.7),		Valine
40	Val	Val (39.5), Leu/Ile (3.5), Pro (2.9),		Valine
41	Phe*,Val	Phe (20.6), Leu/Ile (5.5), Pro (5.8), Val (23.25)		Pnenylalanine
42	Asp*,Pro*,Thr	Pro (18.0), Leu/Ile (5.6), Val (6.6), Gly (3.4), Ala (6.9)	Asp,Pro	Proline and Aspartic acid
43	Ile*/Leu,Asp	Ileu (28.0), Pro (6.2), Val (9.2), Gly (4.2), Ala (8.1)		Isoleucine
44	Pro,Ser,Thr	Pro (12.3), Leu/Ile (10.5), Val (6.5), Gly (4.2), Ala (7.2)		Proline

45	Ala*,Ile	Ala (22.0), Leu/Ile (15.8), Val (9.0), Pro (8.3), Gly (7.7)		Alanine
46	?Ser/Thr	Gly (11.4), Leu/Ile (8.4), Pro (6.7), Val (8.8), Ala (7.8)	Gly*,Ala	Gly
47	Ala*,Glu,Thr	Ala (27.0), Leu/Ile (7.9), Pro (5.6), Val (6.9), Gly (3.4)		Alanine
48	Pro*,Ala,Phe	Pro (21.6), Leu/Ile (7.4), Val (5.4), Gly (4.0), Ala (9.6)		Proline
49	?Gly			
50	?Ser/Thr	Pro (7.7), Leu/Ile (6.6), Val (6.8), Gly (1.3), Ala (5.3)		? Proline
51	Val	Val (16.7), Leu/Ile (7.6), Pro (6.7), Gly (5.0), Ala (7.8)		Valine
52	Ala*,Val,Glu			? Alanine
53	Ala,Glu			
54	Glu,Gly			
55	Leu*,Glu			Leucine
56	Ser*/Thr			? Serine

N-terminal amino acid sequence of Blechnum spicant plastocyanin with the results of analyses of each sample from the sequencer

Sequence Position	T.l.c. identification	G.l.c. identification	Identification after regeneration	Residue
1	Ala	Ala (294.3), Gly (15.8), Ile/Leu (9.5), Val (8.0), Pro (5.7)		Alanine
2	Lys*,Glu		Lys	Lysine
3	Val	Val (258.5), Pro (7.7), Leu/Ile (7.4)		Valine
4	Glu		Glu	Glutamic acid
5	Val	Val (207.4), Leu/Ile (9.5), Pro (4.9)		Valine
6	Gly	Gly (208.8), Val (20.1), Ala (7.1), Pro (5.1), Leu/Ile (4.3)		Glycine
7	Asp	Asp (143.2)		Aspartic acid
8	Glu	Glu (189.1)		Glutamic acid
9	Val	Val (166.1), Ala (9.6), Gly (7.5), Pro (5.5), Leu/Ile (4.6)		Valine

10	Gly	Gly (179.5), Val (25.5), Ala (8.0), Pro (5.8), Leu/Ile (4.86)		Glycine
11	Asn*, Ser/Thr		Asp*, Gly	Asparagine
12	Phe	Phe (61.1), Gly (10.4), Leu/Ile (5.3) Pro (3.8), Ala (3.5), Val (2.7)		Phenylalanine
13	Lys*, Glu		Lys	Lysine
14	Phe	Phe (53.3), Leu/Ile (4.4), Pro (4.0)		Phenylalanine
15	Tyr*, Phe, Glu		Tyr	Tyrosine
16	Pro*, Phe	Pro (69.3), Leu/Ile (4.6)		Proline
17	Glu*, Lys	Glu (48.5)		Glutamic acid
18	Thr*, Asp		α-amino-butyrate	Threonine
19	Ile	Ile (51.6), Pro (11.4), Ala (7.1), Val (6.7), Gly (5.3)		Isoleucine
20	Thr*, Asp		α-amino-butyrate	Threonine
21	Val*, Thr	Val (45.0), Leu/Ile (9.1), Pro (6.3), Gly (4.2), Ala (4.1)		Valine
22	Ala*, Val	Ala (38.3), Val (15.5), Pro (8.8), Leu/Ile (5.3), Gly (5.1)		Alanine
23	Ala	Ala (55.3), Val (15.7), Pro (8.0), Leu/Ile (7.6)		Alanine

24	Gly*,Ala	Gly (35.3), Ala (18.3), Pro (5.8), Leu/Ile (5.0), Val (3.7)		Glycine
25	Glu*,Lys	Glu (18.7)		Glutamic acid
26	Ala*,Glu	Ala (28.8), Gly (4.9), Leu/Ile (3.9), Pro (2.9), Val (2.9)		Alanine
27	Val*,Ala	Val (35.6), Ala (21.9), Gly (9.4), Leu/Ile (8.6), Pro (6.1)		Valine
28	Glu*,Lys	Glu (21.8)		Glutamic acid
29	Phe*,Glu	Phe (15.3)		Phenylalanine
30	Thr*,Ser		α-amino-butyrate	Threonine
31	Leu*,Ile	Leu (21.8), Ala (8.4), Pro (8.1)		Leucine
32	Val*,Leu	Val (17.3), Leu/Ile (9.6), Pro (4.7), Gly = 0		Valine
33	Gly*,Val	Gly (7.2), Val (15.3), Leu/Ile (10.2), Ala (6.8), Pro (4.8)		Glycine
34	Glu	Glu*, Val (8.4), Leu/Ile (8.4), Ala (5.0), Pro (3.1), Gly = 0		Glutamic acid
35	Thr*,Glu	Val (9.2), Leu/Ile (7.9), Ala (5.3) Pro (5.2), Gly = 0.	α-amino-butyrate, Pro	Threonine

36	Gly*,Glu,Pro	Gly (4.3), Pro (3.9), Val (6.7), Leu/Ile (5.4), Ala (5.4)	Pro*,Gly,Ala, α-amino-butyrate	Proline? Glycine?
37	Nothing	Organic phase: Gly = 0, Pro (1.4), Val (7.5), Leu/Ile (5.0), Ala (5.8)	His	Histidine
38	Asn		Asp	Asparagine
39	Ile*,Asn	Ile (15.0), Ala (10.0), Val (7.5), Pro (4.3)		Isoleucine
40	Val	Val (16.1), Leu/Ile (10.7), Ala (7.1), Pro (3.8)		Valine

N-terminal amino acid sequence of Dryopteris dilatata plastocyanin with the results of analyses of each sample from the sequencer

Sequence Position	T.l.c. identification	G.l.c. identification	Identification after regeneration	Residue
1	Ala	Ala (295.8)		Alanine
2	Lys		Lys	Lysine
3	Val	Val (166.4), Ala (26.7), Pro (15.3), Leu/Ile (12.8)		Valine
4	Glu		Glu*,Gly	Glutamic acid
5	Val	Val (192.7), Leu/Ile (41.8), Pro (19.5), Ala (15.6), Gly = 0		Valine
6	Gly	Gly (76.8), Val (97.0), Pro (11.4), Leu/Ile (7.5)		Glycine
7	Asp		Asp	Aspartic acid
8	Glu		Glu	Glutamic acid
9	Val	Val (172.3), Gly (16.8), Ala (16.7) Pro (16.4), Leu/Ile (14.6)		Valine

10	Gly	Gly (108.9), Val (144.7), Pro (33.2), Ala (32.6), Leu/Ile (29.6)		Glycine
11	Asn		Asp	Asparagine
12	Phe	Phe (107.3), Val (14.2), Gly (11.4), Leu/Ile (6.7), Pro (4.2)		Phenylalanine
13	Lys		Lys	Lysine
14	Phe	Phe (79.8), Val (18.5), Ala (9.3), Leu/Ile (7.7), Pro (6.0), Gly (5.9)		Phenylalanine
15	Tyr		Tyr	Tyrosine
16	Pro	Pro (77.3), Val (13.5), Leu/Ile (11.6), Ala (10.7), Gly (9.3)		Proline
17	Glu		Glu*,Pro	Glutamic acid
18	Ala	Ala (127.6), Val (14.7), Pro (14.6), Leu/Ile (10.9), Gly (9.8)		Alanine
19	Ile	Ile (104.0), Ala (35.5), Val (5.2)		Isoleucine
20	Thr*,Ser		α-amino-butyrate	Threonine
21	Val	Val (117.7), Ala (14.2), Leu/Ile (13.4) Pro (10,1), Gly (7.4)		Valine

22	Ala*,Val	Ala (255.6), Val (67.6), Leu/Ile (9.8), Pro (9.15)		Alanine
23	Ala	Ala (155.2), Val (9.7), Pro (7.9), Leu/Ile (4.9), Gly (1.1)		Alanine
24	Gly*,Ala	Gly (20.5), Ala (2.50), Leu/Ile (8.8), Val (3.5), Pro (3.3)		Glycine
25	Glu		Glu*,Asp	Glutamic acid
26	Ser*,Thr,Ala		Ala*, α-amino-butyrate	Serine
27	Ile*,Leu		Ile	Isoleucine
28	Glu		Glu*,Gly	Glutamic acid
29	Phe	Phe (66.1), Ala (7.6), Leu/Ile (5.7), Val (3.8), Pro (1.3)		Phenylalanine
30	Thr*,Ser		α-amino-butyrate	Threonine
31	Leu*,Ile	Leu (50.5), Ala (12.3), Pro (7.7), Val (7.2), Gly (4.0)		Leucine
32	Val	Val (85.3), Leu/Ile (22.8), Ala (10.25), Pro (6.4), Gly = 0		Valine

33	Gly*,Val	Gly (11.6), Val (35.5), Ala (11.9), Leu/Ile (11.7), Pro (5.3)		Glycine
34	Glu		Glu,Gly	Glutamic acid
35	Thr*,Ser		α-amino-butyrate	Threonine
36	Gly		Gly	Glycine
37	Nothing		His	Histidine
38	Asn		Asp	Asparagine
39	Ile*,Leu		Ile*,Asp	Isoleucine
40	Val		Val*,Ile	Valine

N-terminal amino acid sequence of Dryopteris filix-mas plastocyanin with the results of analyses of each sample from the sequencer

Sequence Position	T.l.c. identification	G.l.c. identification	Identification after regeneration	Residue
1	Ala	Ala (332.8), Val (37.1), Leu/Ile (11.9), Pro (3.2)		Alanine
2	Lys*,Ala		Lys	Lysine
3	Val*,Lys	Val (260.0), Ala (22.5), Leu/Ile (12.1), Pro (10.3)		Valine
4	Glu*,Val	Glu		Glutamic acid
5	Val*,Glu	Val (222.9), Pro (10.7), Ala (9.7), Leu/Ile (5.9)		Valine
6	Gly*,Val	Gly (297.2), Val (48.6), Pro (13.2), Leu/Ile (8.3)		Glycine
7	Asp	Asp		Aspartic acid
8	Glu*,Asp	Glu		Glutamic acid
9	Val	Val (151.3), Pro (8.7), Leu/Ile (8.0)		Valine

10	Gly*,Val	Gly (251.6), Val (49.0), Ala (11.6), Leu/Ile (11.2), Pro (7.9)		Glycine
11	Asn		Asp*,Gly	Asparagine
12	Phe	Phe (129.2), Gly (27.1), Ala (17.5), Val (17.5), Leu/Ile (16.1), Pro (9.4)		Phenylalanine
13	Lys*,Glu		Lys	Lysine
14	Phe	Phe (97.7), Ala (8.9), Leu/Ile (7.2), Val (6.3), Pro (4.9)		Phenylalanine
15	Tyr*,Phe,Lys, Glu		Tyr	Tyrosine
16	Pro	Pro (56.5), Leu/Ile (12.3), Ala (9.5), Gly (7.9), Val (5.9)		Proline
17	Asp	Asp		Aspartic acid
18	Ser*,Asp		Ala	Serine
19	Ile	Ile (58.9), Val (9.8), Ala (8.0), Pro (6.2)		Isoleucine
20	Thr*,Ser		α-amino butyrate	Threonine
21	Val	Val (59.4), Leu/Ile (16.0), Ala (9.0), Pro (6.6)		Valine

22	Serine		Ala	Serine
23	Ala	Ala (44.8), Val (10.9), Leu/Ile (6.2), Pro (4.0)		Alanine
24	Gly*,Ala	Gly (53.0), Ala (24.7), Leu/Ile (6.1), Val (5.8), Pro (3.1)		Glycine
25	Glu	Glu		Glutamic acid
26	Ala*,Glu	Ala (45.3), Gly (15.0), Leu/Ile (5.4), Pro (4.4), Val (4.4)		Alanine
27	Val*,Ala	Val (35.2), Ala (26.7), Leu/Ile (10.3), Pro (5.1)		Valine
28	Glu	Glu		Glutamic acid
29	Phe	Phe (33.8), Val (11.6), Ala (8.5), Leu/Ile (8.2), Pro (4.0)		Phenylalanine
30	Thr*,Asp		α-amino-butyrate	Threonine
31	Leu	Leu (21.1), Ala (7.4), Pro (4.1), Val (4.1)		Leucine
32	Val*,Leu	Val (20.4), Leu/Ile (9.5), Ala (4.3) Pro (2.7), Gly = 0		Valine

33	Gly	Gly (12.1), Val (13.1), Leu/Ile (7.1), Ala (6.8), Pro (4.9)		Glycine
34	Glu		Glu	Glutamic acid
35	Thr*,Ser,Glu		α-amino-butyrate	Threonine
36	Pro,Gly			? Proline ? Glycine
37	Nothing		His	Histidine
38	Asn		Asp	Asparagine
39	Ile*,Leu	Leu/Ile (12.3), Val (7.0)	Ile,Asp	Isoleucine
40	Val*,Ile	Val (12.4), Leu/Ile (6.7)		Valine

N-terminal amino acid sequence of Cycas revoluta plastocyanin with the results of analyses of each sample from the sequencer

Sequence Position	T.l.c. identification	G.l.c. identification	Identification after regeneration	Residue
1	Ile	Ile (171.1), Ala (15.2), Gly (14.9), Pro (6.8)		Isoleucine
2	Glu	Glu		Glutamic acid
3	Val	Val (136.1), Leu/Ile (9.1)		Valine
4	Leu	Leu (138.0), Val (14.4), Ala (7.1)		Leucine
5	Leu	Leu (265.1), Val (11.5)		Leucine
6	Gly	Gly (98.5), Leu/Ile (27.6), Val (10.3), Ala (4.9)		Glycine
7	Gly	Gly (82.0), Leu/Ile (15.8), Val (8.9), Ala (4.5)		Glycine
8	Asn		Asp*,Gly	Asparagine
9	Gly	Gly (62.9), Leu/Ile (6.0), Val (5.8)		Glycine

10	Gly	Gly (44.0), Val (3.8), Leu/Ile (3.5)		Glycine
11	Glu	Glu		Glutamic acid
12	Leu*,Glu	Leu (111.8), Val (6.2), Ala (5.5)		Leucine
13	Ala	Ala (87.7), Leu/Ile (32.8), Val (6.8), Gly (2.5)		Alanine
14	Phe*,Ala	Phe (75.2), Ala (40.6), Leu/Ile (15.1), Val (6.6)		Phenylalanine
15	Ile*,Phe	Ile (65.9), Ala (9.0), Val (4.1), Gly (1.4), Pro = 0		Isoleucine
16	Pro	Pro (17.6), Leu/Ile (33.3), Gly (13.7), Val (10.7), Ala (6.9)		Proline
17	Asp	Asp		Aspartic acid
18	Lys		Lys*,Gly,Pro	Lysine
19	Phe	Phe (58.8), Leu/Ile (13.4), Ala (8.9), Val (8.0), Pro (6.3), Gly (5.1)		Phenylalanine
20	?Glu		Glu*,Phe	Glutamic acid
21	Val	Val (100.8), Leu/Ile (33.6), Ala (14.4)		Valine

22	Ala	Ala (77.5), Val (52.0), Leu/Ile (12.5), Pro (4.2)		Alanine
23	Pro*,Ala	Pro (49.6), Ala (43.3), Val (24.2), Leu/Ile (16.1)		Proline
24	Gly*,Ala	Gly (23.4), Ala (18.9), Pro (14.2), Val (13.6), Leu/Ile (12.2)		Glycine
25	Asp,Glu	Glu*,Asp		Glutamic acid
26	?Glu		Glu*,Gly,Ala	Glutamic acid
27	Ile	Ile (60.5), Val (8.2)		Isoleucine
28	Val	Val (58.7), Leu/Ile (33.3), Ala (5.9)		Valine
29	Phe	Phe (38.9), Leu/Ile (17.0), Val (14.2), Ala (10.6)		Phenylalanine
30	Lys		Lys*,Ala,Gly	Lysine
31	Asn		Asp*,Lys	Asparagine
32	Asn		Asp*,Lys,Pro	Asparagine
33	?Ala	Ala (24.1), Gly (12.0), Val (8.9), Leu/Ile (8.3)		Alanine

34	?Ala	Ala (32.6), Gly (10.1), Leu/Ile (6.8), Val (2.9)		Alanine
35	?Phe	Phe (16.6), Ala (16.9), Val (4.1), Leu/Ile (3.8), Pro (1.1)		Phenylalanine
36		Pro (4.0), Ala (10.9), Phe (8.5), Val (6.5), Pro (4.0)	Pro*,Ala,Gly, Glu,Asp	Proline
37	Nothing		His,Ala,Gly, Glu,Asp	Histidine
38	Asn		Asp*,Gly	Asparagine
39	?Val	Val (11.8), Leu/Ile (3.6)		Valine
40	?Val	Val (19.0), Leu/Ile (3.8)		Valine

N-terminal amino acid sequence of Araucaria araucana plastocyanin with the results of analyses of each sample from the sequencer

Sequence Position	T.l.c. identification	G.l.c. identification	Identification after regeneration	Residue
1	Ala	Ala (450.0)		Alanine
2	Glu	Glu		Glutamic acid
3	Val	Val (294.0), Ala (26.5)		Valine
4	Leu	Leu (304.5), Ala (8.1), Val (7.2)		Leucine
5	Met	Met* (not standardised), Leu/Ile (14.4)		Methionine
6	Gly	Gly (348.0)		Glycine
7	Gly	Gly (312.0)		Glycine
8	Asn		Asp	Asparagine
9	Gly	Gly (253.5), Ala (22.1), Val (6.7), Pro (6.45), Leu/Ile (5.85)		Glycine
10	Gly	Gly (336.8), Pro (8.8) Val (8.1), Leu/Ile (4.4)		Glycine

11	Glu		Glu*,Gly	Glutamic acid
12	Leu	Leu (185.8)		Leucine
13	Ala	Ala (236.8), Leu/Ile (12.5)		Alanine
14	Phe	Phe (158.7), Gly (33.1), Leu/Ile (8.1), Ala (3.2)		Phenylalanine
15	Ile	Ileu (140.8), Gly (36.0)		Isoleucine
16	Pro	Pro (175.7), Leu/Ile (14.6), Gly (12.2)		Proline
17	Ser		Ala*, α-amino-butyrate	Serine
18	Glu		Glu*,Lys	Glutamic acid
19	Phe	Phe (101.8), Gly (17.3), Pro (10.1), Leu/Ile (8.4), Val (7.7)		Phenylalanine
20	Ser		Ala	Serine
21	Val	Val (156.9), Leu/Ile (6.8), Pro (6.7)		Valine
22	Ala	Ala (167.1), Val (16.7), Pro (7.0), Leu/Ile (5.6)		Alanine

23	Pro	Pro (111.2), Ala (17.4), Gly (12.0) Leu/Ile (10.1), Val (7.2)		Proline
24	Gly	Gly (147.3), Pro (15.2), Val (8.0), Ala (6.7), Leu/Ile (6.2)		Glycine
25	Glu		Glu,* Gly	Glutamic acid
26	Thr		α-amino butyrate	Threonine
27	Ile	Ile (97.4), Pro (10.7)		Isoleucine
28	Thr		α-amino butyrate	Threonine
29	Phe	Phe (69.5), Val (5.6), Pro (5.5) Gly (4.5), Leu/Ile (4.0), Ala (2.0)		Phenylalanine
30	Lys		Lys	Lysine
31	Asn		Asp	Asparagine
32	Asn,* Asp		Asp	Asparagine
33	Ala	Ala (100.2), Gly (9.0), Val (7.5), Leu/Ile (7.2), Pro (5.7)		Alanine
34	Gly	Gly (72.9), Ala (19.5), Val (8.7), Pro (7.0), Leu/Ile (5.6)		Glycine

35	Phe	Phe (38.4), Gly (18.9), Pro (10.3), Val (7.9), Leu/Ile (6.2), Ala (6.2)		Phenylalanine
36	Pro	Pro (35.6), Leu/Ile (6.4), Val (4.8)		Proline
37	Nothing		His	Histidine
38	Asn		Asp	Asparagine
39	Val	Val (51.4), Pro (10.8), Leu/Ile (6.2)		Valine
40	Ile*,Val	Ile (45.2), Val (27.8), Pro (7.3)		Isoleucine

N-terminal amino acid sequence of Taxus baccata plastocyanin with the results of analyses of each sample from the sequencer

Sequence Position	T.l.c. identification	G.l.c. identification	Identification after regeneration	Residue
1	Leu	Leu (243.3)		Leucine
2	Glu	Glu		Glutamic acid
3	Val	Val (124.1), Leu/Ile (39.2), Ala (20.6)		Valine
4	Leu	Leu (122.6), Val (25.6)		Leucine
5	Met	Met* (not standardised), Val (40.1), Leu/Ile (77.3), Gly (12.9), Ala (5.5)		Methionine
6	Gly	Gly (87.4), Leu/Ile (43.5), Val (33.6)		Glycine
7	Gly	Gly (115.5), Leu/Ile (17.2), Val (6.5), Pro (4.7), Ala (4.0)		Glycine
8	Asn		Asp	Asparagine
9	Gly	Gly (130.2), Leu/Ile (10.3), Val (9.7)		Glycine

10	Gly	Gly (127.7), Val (8.9), Leu/Ile (8.5), Pro (5.7)		Glycine
11	Glu		Glu*,Gly	Glutamic acid
12	Leu	Leu (51.0), Gly (30.7), Val (8.9)		Leucine
13	Val	Val (54.2), Leu/Ile (5.0)		Valine
14	Phe*,Leu	Phe (42.2), Leu/Ile (27.1), Gly (19.7), Val (15.6), Pro (4.9)		Phenylalanine
15	Ile	Ile (60.3), Val (23.5), Ala (5.3)		Isoleucine
16	Pro	Pro (56.9), Gly (19.8), Leu/Ile (17.5), Val (14.4), Ala (4.9)		Proline
17	Ser		Ala*,Pro	Serine
18	Glu		Glu*,Gly,Pro	Glutamic acid
19	Phe	Phe (29.6), Leu/Ile (12.5), Pro (7.3), Val (7.2), Ala (4.5)		Phenylalanine
20	Gln		Glu,Gly	Glutamine
21	Leu	Leu (31.0), Val (11.3), Pro (9.4), Ala (6.5)		Leucine
22	Thr		α-amino-butyrate	Threonine

23	Ala	Ala (20.6), Leu/Ile (12.8), Val (4.1), Pro (2.8)		Alanine
24	Gly	Gly (25.7), Leu/Ile (9.7), Pro (7.8), Ala (6.3), Val (4.7)		Glycine
25	Asp		Asp	Aspartic acid
26	Thr		α-amino-butyrate	Threonine
27	Ile	Ile (15.9), Gly (8.8), Pro (3.7), Val (3.5), Ala (3.3)		Isoleucine
28	Val	Val (15.2), Leu/Ile (7.4), Gly (7.2), Pro (4.2), Ala (3.1)		Valine
29	Phe	Phe (12.2), Leu/Ile (8.3), Val (4.5)		Phenylalanine
30	Lys		Lys	Lysine
31	Asn		Asp	Asparagine
32	Asn		Asp	Asparagine
33	Ala	Ala (12.3), Leu/Ile (6.9), Val (6.0), Pro (3.8)		Alanine
34	Gly	Gly (11.5), Ala (5.5), Leu/Ile (5.2), Val (4.5), Pro (4.3)		Glycine

35	Phe	Phe (7.0), Ala (6.0), Leu/Ile (5.7), Gly (5.7), Val (4.9), Pro (2.2)		Phenylalanine
36	Nothing		His	Histidine
38	Asn		Asp	Asparagine
39	Val	Val (5.8), Leu/Ile (4.1)		Valine
40	Val	Val (4.4), Leu/Ile (4.0)		Valine

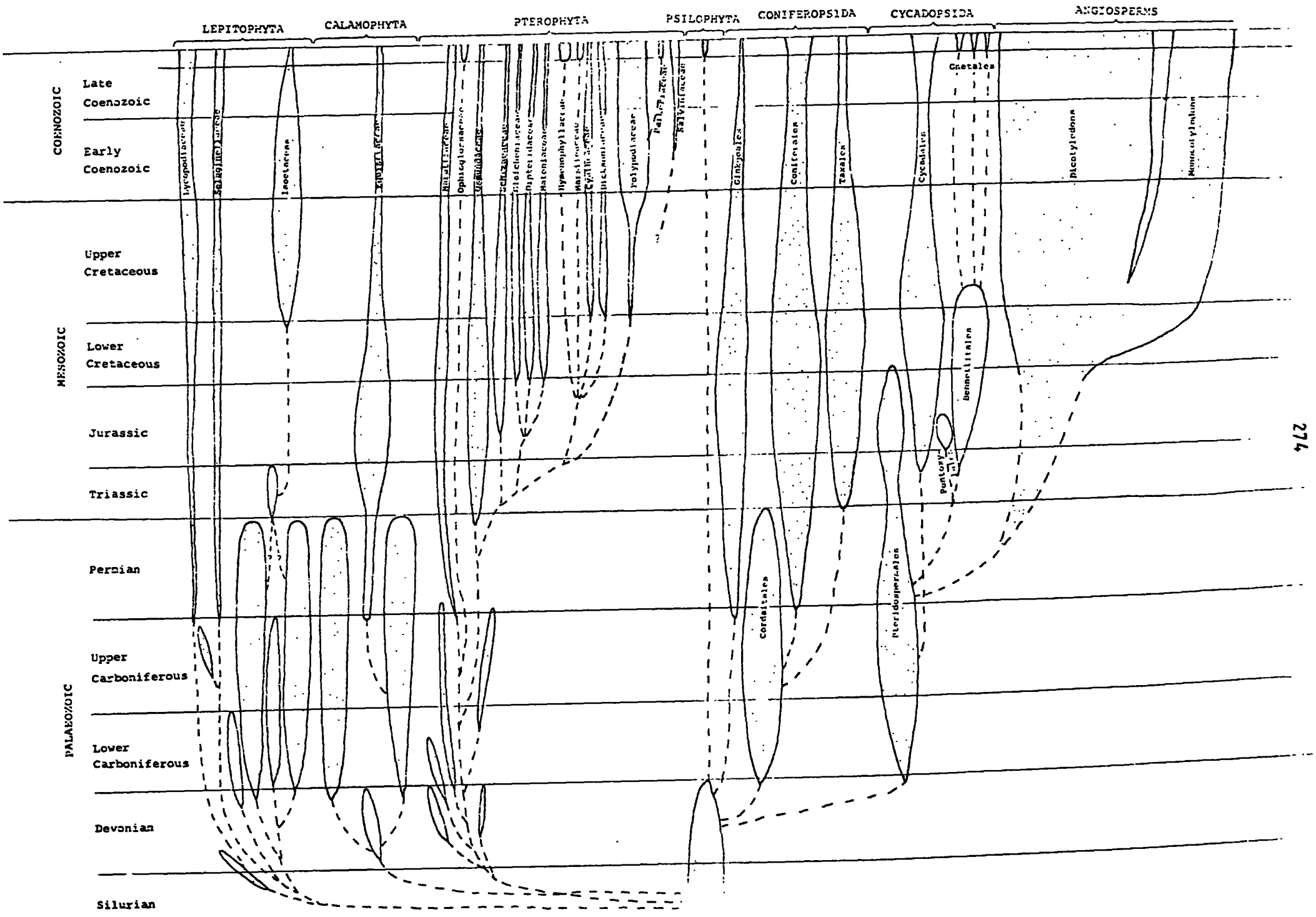
APPENDIX III

Evolutionary schemes based on fossil evidence

APPENDIX III A

The geological range and suggested interrelationships among pteridophytes, gymnosperms and angiosperms, modified from Smith, 1955 and Strasburger *et al.*, 1978.

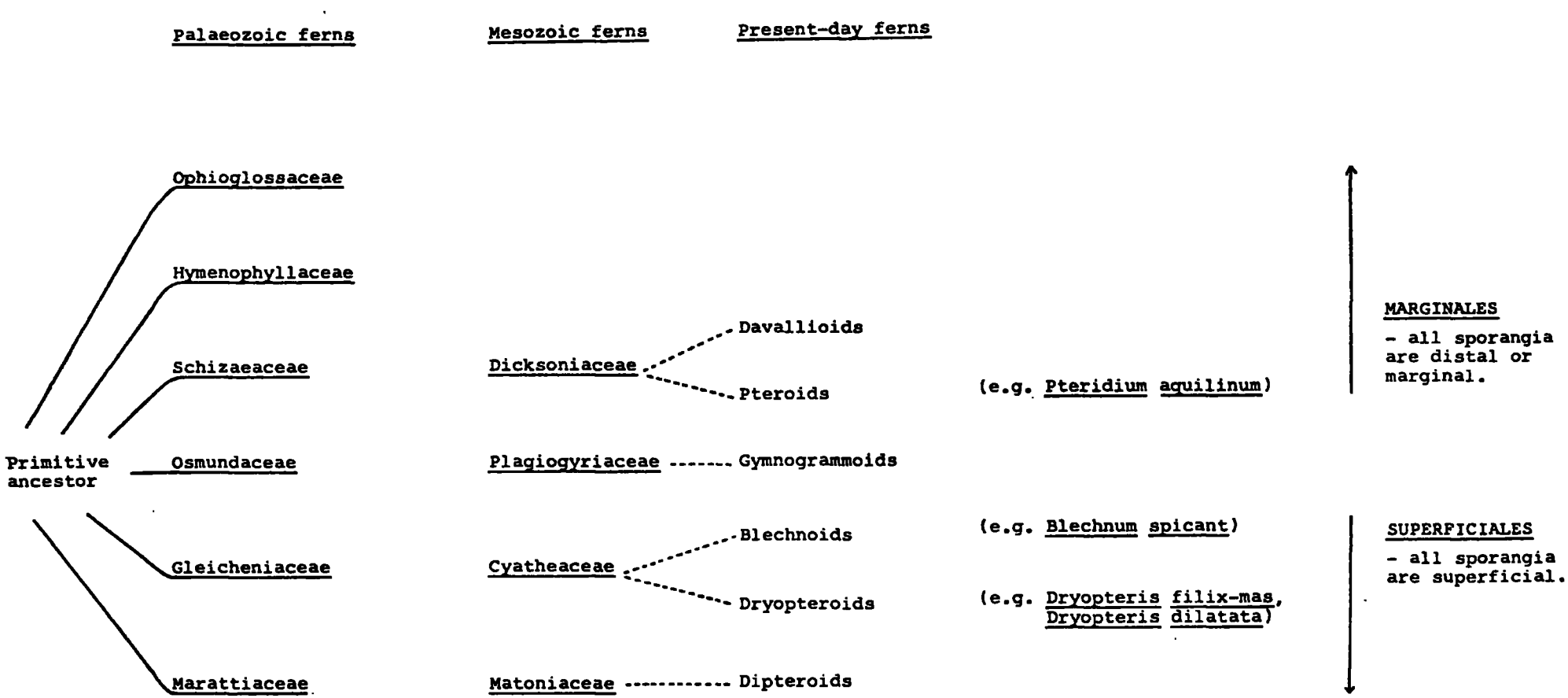
ANGIOSPERMS
CYCADOPSIDA
CONIFEROPSIDA
PSILOPHYTA
PTEROPHYTA
CALAMOPHYTA
LEPTOPHYTA
Monocotyledons
Dicotyledons
Gnetales
Bennettitales
Cycadales
Pteridosperms
Taxales
Coniferales
Cordaitales
Ginkgoales
Psilophytales
Polypodiaceae
Hymenophyllaceae
Matoniaceae
Dipteridaceae
Dicksoniaceae
Cyatheaceae
Schizaeaceae
Gleicheniaceae
Marattiaceae
Ophioglossaceae
Osmundaceae
Filicales
Equisetales
Lycopodiales
Isoetales
Selaginellales
Lycopodiales
Late Coenozoic
Early Coenozoic
Upper Cretaceous
Lower Cretaceous
Jurassic
Triassic
Permian
Upper Carboniferous
Lower Carboniferous
Devonian
Silurian
CENOZOIC
MESOZOIC
PALAEOZOIC


LEPITOPHYTA
CALAMOPHYTA
PTEROPHYTA
PSILOPHYTA
CONIFEROPSIDA
CYCADOPSIDA
ANGIOSPERMS
CENOZOIC
Late Coenozoic
Early Coenozoic
MESOZOIC
Upper Cretaceous
Lower Cretaceous
Jurassic
Triassic
PALAEOZOIC
Permian
Upper Carboniferous
Lower Carboniferous
Devonian
Silurian
Lycopodiaceae
Equisetaceae
Ophioglossaceae
Polypodiaceae
Salviniaceae
Ginkgoales
Coniferales
Taxales
Cycadales
Gnetales
Dicotyledons
Monocotyledons
Cordaitales
Pteridospermales
Bennettitales

APPENDIX III B

Scheme of Evolution of Pterophyta, based on position of sporangia (Bower, 1935).

Lines are left disconnected to avoid any idea of direct descent, while conveying some approximate suggestion of probable relations.


Palaeozoic ferns
Mesozoic ferns
Present-day ferns
Ophioglossaceae
Hymenophyllaceae
Schizaeaceae
Primitive ancestor
Osmundaceae
Gleicheniaceae
Marattiaceae
Dicksoniaceae
Davallioids
Pteroids
Plagiogyriaceae
Gymnogrammoids
Cyatheaceae
Blechnoids
Dryopteroids
Matoniaceae
Dipteroids
(e.g. Pteridium aquilinum)
(e.g. Blechnum spicant)
(e.g. Dryopteris filix-mas, Dryopteris dilatata)
MARGINALES
- all sporangia are distal or marginal.
SUPERFICIALES
- all sporangia are superficial.

APPENDIX III C

Geological Histroy of the Gymnosperms (Sporne, 1965)

The table indicates the approximate distribution of gymnosperms during successive geological periods.

Be - Bennettitales

Ca - Caytoniaceae

Con - Coniferales

Cor - Cordaitales

Cy - Cycadales

Gi - Ginkgoales

Gn - Gnetales

Le - Lebachiaceae

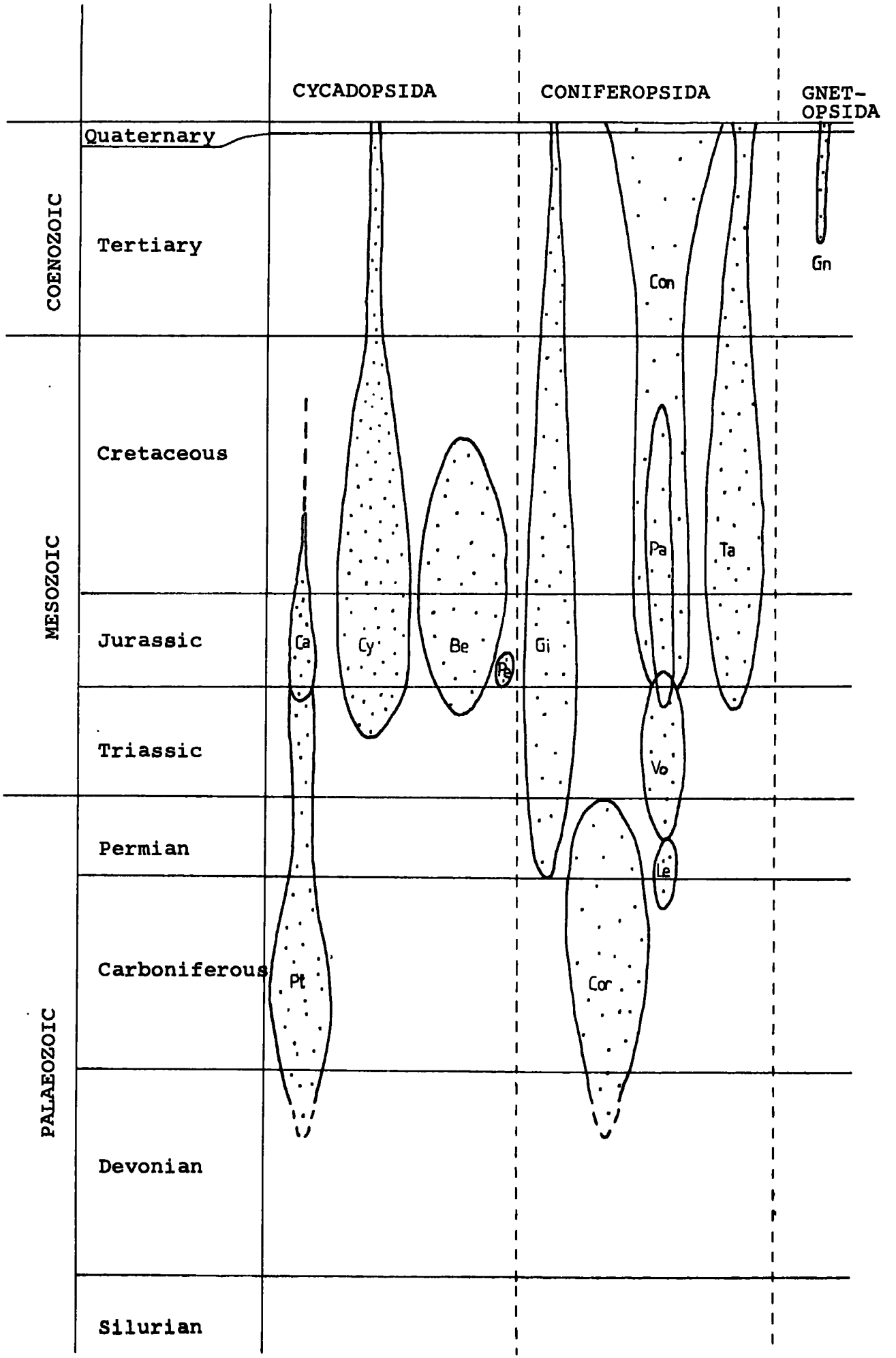
Pa - Palissyaceae

Pe - Pentoxylales

Pt - Pteridospermales

Ta - Taxales

Vo - Voltziaceae


CYCADOPSIDA
CONIFEROPSIDA
GNET-
OPSIDA
Quaternary
COENOZOIC
Tertiary
Gn
Con
Cretaceous
Pa
Ta
MESOZOIC
Jurassic
Ca
Cy
Be
Pe
Gi
Triassic
Vo
Permian
Le
Carboniferous
Pt
Cor
PALAEOZOIC
Devonian
Silurian

APPENDIX IV

Sequence Information

IVA Complete sequences of plastocyanin from higher plants, bracken and algae.

IVB Variability of each residue position for plastocyanin data set.

IVC Invariance of amino acids within the major plant groups.

APPENDIX IVA

Complete sequences of plastocyanin from higher plants, bracken and algae

Numbering is with higher plant N-terminus as 1.

Sequence information for higher plants and algae was obtained from Boulter *et al.*, (1977).

Invariant residues are outlined

\+ insert -Pro-Ala- here (Aitken, 1975)

* insert -Pro- here (Aitken, 1975)

\- no residue

```
                -2 -1  1              5                  10                  15                  20                  25                  30
Anabaena:       Glu-Thr-Tyr-Thr-Val-Lys-Leu-Gly-Ser-Asp-Lys-Gly-Leu-Leu-Val-Phe-Glu-Pro-Ala-Lys-Leu-Thr-Ile-Lys-Pro-Gly-Asp-Thr-Val-Glu-Phe-Leu-Asn-Asn-Lys-Val-
Chlorella:         Asp-Val-Thr-Val-Lys-Leu-Gly-Ala-Asp-Ser-Gly-Ala-Leu-Val-Phe-Glu-Pro-Ser-Ser-Val-Thr-Ile-Lys-Ala-Gly-Glu-Thr-Val-Thr-Trp-Val-Asn-Asn-Ala-Gly-
Rumex:             Ile-Glu-Ile-Lys-Leu-Gly-Gly-Asp-Asp-Gly-Ala-Leu-Ala-Phe-Val-Pro-Gly-Ser-Phe-Thr-Val-Ala-Ala-Gly-Glu-Lys-Ile-Val-Phe-Lys-Asn-Asn-Ala-Gly-
Sambucus:          Val-Glu-Ile-Leu-Leu-Gly-Gly-Glu-Asp-Gly-Ser-Leu-Ala-Phe-Ile/Val-Pro-Ser/Gly-Asn-Phe-Ser-Val-Pro-Ser-Gly-Glu-Lys-Ile-Thr-Phe-Lys-Asn-Asn-Ala-Gly-
Spinacia:          Val-Glu-Val-Leu-Leu-Gly-Gly-Gly-Asp-Gly-Ser-Leu-Ala-Phe-Leu-Pro-Gly-Asp-Phe-Ser-Val-Ala-Ser-Gly-Glu-Glu-Ile-Val-Phe-Lys-Asn-Asn-Ala-Gly-
Cucurbita:         Ile-Glu-Val-Leu-Leu-Gly-Gly-Asp-Asp-Gly-Ser-Leu-Ala-Phe-Ile-Pro-Asn-Asp-Phe-Ser-Val-Ala-Ala-Gly-Glu-Lys-Ile-Val-Phe-Lys-Asn-Asn-Ala-Gly-
Capsella:          Ile-Glu-Val-Leu-Leu-Gly-Gly-Gly-Asp-Gly-Ser-Leu-Ala-Phe-Val-Pro-Asn-Asp-Phe-Ser-Ile-Ala-Lys-Gly-Glu-Lys-Ile-Val-Phe-Lys-Asn-Asn-Ala-Gly-
Solanum:           Leu-Asp-Val-Leu-Leu-Gly-Gly-Asp-Asp-Gly-Ser-Leu-Ala-Phe-Ile-Pro-Gly-Asn-Phe-Ser-Val-Ser-Ala-Gly-Glu-Lys-Ile-Thr-Phe-Lys-Asn-Asn-Ala-Gly-
Vicia:             Val-Glu-Val-Leu-Leu-Gly-Ala-Ser-Asp-Gly-Gly-Leu-Ala-Phe-Val-Pro-Asn-Ser-Phe-Glu-Val-Ser-Ala-Gly-Asp-Thr-Ile-Val-Phe-Lys-Asn-Asn-Ala-Gly-
Phaseolus:         Leu-Glu-Val-Leu-Leu-Gly-Ser-Gly-Asp-Gly-Ser-Leu-Val-Phe-Val-Pro-Ser-Glu-Phe-Ser-Val-Pro-Ser-Gly-Glu-Lys-Ile-Val-Phe-Lys-Asn-Asn-Ala-Gly-
Mercurialis:       Leu-Asp-Val-Leu-Leu-Gly-Ser-Asp-Asp-Gly-Glu-Leu-Ala-Phe-Val-Pro-Asn-Asn-Phe-Ser-Val-Pro-Ser-Gly-Glu-Lys-Ile-Thr-Phe-Lys-Asn-Asn-Ala-Gly-
Lactuca:           Ala-Glu-Val-Leu-Leu-Gly-Ser-Ser-Asp-Gly-Gly-Leu-Val-Phe-Glu-Pro-Ser-Thr-Phe-Ser-Val-Ala-Ser-Gly-Glu-Lys-Ile-Val-Phe-Lys-Asn-Asn-Ala-Gly-
Pteridium:         Ala-Lys-Val-Glu-Val-Gly-Asp-Glu-Val-Gly-Ser-Phe-Lys-Phe-Thr-Pro-Asp-Thr-Ile-Thr-Val-Ala-Ala-Gly-Glu-Ala-Ile-Glu-Phe-Thr-Leu-Val-Gly-Glu-

 35                  40                  45                  50                  55                  60                  65                  70
Pro-Pro-His-Asn-Val-Val-Phe-Asp-Ala-Ala-Leu-Asn-Pro-Ala-Lys-Ser-Ala-Asp-Leu-Ala-Lys-Ser-Leu-Ser-His-Lys-Gln-Leu-Leu-Met-Ser-Pro-Gly-Gln-Ser-Thr-Ser-Thr-Thr-Phe+
Phe-Pro-His-Asn-Ile-Val-Phe-Asp-Glu-Asp-Glu-Val-Pro-Ser-Gly-Ala-Asn-Ala-Glu-Ala-Leu-Ser- — - — -His-Glu-Asp-Tyr-Leu-Asn-Ala-Pro-Gly-Glu-Ser-Tyr-Ser-Ala-Lys-Phe-
Phe-Pro-His-Asn-Ile-Val-Phe-Asp-Glu-Asp-Glu-Val-Pro-Ala-Gly-Val-Asp-Ala-Ser-Lys-Ile-Ser-Met-Ser-Glu-Glu-Asp-Leu-Leu-Asn-Ala-Pro-Gly-Glu-Thr-Tyr-Ala-Val-Thr-Leu-
Phe-Pro-His-Asn-Val-Val-Phe-Asp-Glu-Asp-Glu-Val-Pro-Ser-Gly-Val-Asp-Ser-Ala-Lys-Ile-Ser-Met-Ser-Glu-Asp-Asp-Leu-Leu-Asn-Ala-Pro-Gly-Glu-Thr-Tyr-Ser-Val-Thr-Leu-
Phe-Pro-His-Asn-Val-Val-Phe-Asp-Glu-Asp-Glu-Ile-Pro-Ser-Gly-Val-Asp-Ala-Ala-Lys-Ile-Ser-Met-Ser-Glu-Glu-Asp-Leu-Leu-Asn-Ala-Pro-Gly-Glu-Thr-Tyr-Lys-Val-Thr-Leu-
Phe-Pro-His-Asn-Val-Val-Phe-Asp-Glu-Asp-Glu-Ile-Pro-Ser-Gly-Val-Asp-Ala-Gly-Lys-Ile-Ser-Met-Asn-Glu-Glu-Asp-Leu-Leu-Asn-Ala-Pro-Gly-Glu-Val-Tyr-Lys-Val-Asn-Leu-
Phe-Pro-His-Asn-Val-Val-Phe-Asp-Glu-Asp-Glu-Ile-Pro-Ser-Gly-Val-Asp-Ala-Ser-Lys-Ile-Ser-Met-Asp-Glu-Asn-Asp-Leu-Leu-Asn-Ala-Ala-Gly-Glu-Thr-Tyr-Glu-Val-Ala-Leu-
Phe-Pro-His-Asn-Val-Val-Phe-Asp-Glu-Asp-Glu-Ile-Pro-Ala-Gly-Val-Asp-Ala-Ser-Lys-Ile-Ser-Met-Ala-Glu-Glu-Asp-Leu-Leu-Asn-Ala-Ala-Gly-Glu-Thr-Tyr-Ser-Val-Thr-Leu-
Phe-Pro-His-Asn-Val-Val-Phe-Asp-Glu-Asp-Glu-Ile-Pro-Ser-Gly-Val-Asp-Ala-Ala-Lys-Ile-Ser-Met-Pro-Glu-Glu-Asp-Leu-Leu-Asn-Ala-Pro-Gly-Glu-Thr-Tyr-Ser-Val-Lys-Leu-
Phe-Pro-His-Asn-Val-Val-Phe-Asp-Glu-Asp-Glu-Ile-Pro-Ala-Gly-Val-Asp-Ala-Val-Lys-Ile-Ser-Met-Pro-Glu-Glu-Glu-Leu-Leu-Asn-Ala-Pro-Gly-Glu-Thr-Tyr-Val-Val-Thr-Leu-
Phe-Pro-His-Asn-Val-Val-Phe-Asp-Glu-Asp-Glu-Ile-Pro-Ser-Gly-Val-Asp-Ala-Ser-Lys-Ile-Ser-Met-Asp-Glu-Ala-Asp-Leu-Leu-Asn-Ala-Pro-Gly-Glu-Thr-Tyr-Ala-Val-Thr-Leu-
Phe-Pro-His-Asn-Val-Val-Phe-Asp-Glu-Asp-Glu-Ile-Pro-Ala-Gly-Val-Asp-Ala-Ser-Lys-Ile-Ser-Met-Ser-Glu-Glu-Asp-Leu-Leu-Asn-Ala-Pro-Gly-Glu-Thr-Tyr-Ala-Val-Thr-Leu-
Thr-Gly-His-Asn-Val-Val-Phe-Pro/Asp-Ile-Pro-Ala-Gly-Ala-Pro-Gly-Pro-Val-Ala-Ser-Glu-Leu-Ser-Met-Asp-Gln-Asp-Asp-Leu-Leu-Ser-Gln-Asn-Glu-Asp/Glu-Asp-Phe-Thr-Ala-Lys-Val-

 76          *       80                  85                  90                  95
Asp-Ala-Ala-Gly-Glu-Tyr-Thr-Phe-Tyr-Cys-Glu-Pro-His-Arg-Gly-Ala-Gly-Met-Val-Gly-Lys-Ile-Thr-Val-Ala-Gly
Asp-Thr-Ala-Gly-Thr-Tyr-Gly-Tyr-Phe-Cys-Glu-Pro-His-Gln-Gly-Ala-Gly-Met-Lys-Gly-Thr-Ile-Thr-Val-Gln
Ser-Glu-Lys-Gly-Thr-Tyr-Ser-Phe-Tyr-Cys-Ser-Pro-His-Gln-Gly-Ala-Gly-Met-Val-Gly-Lys-Val-Thr-Val-Gln
Thr-Glu-Ser-Gly-Thr-Tyr-Lys-Phe-Tyr-Cys-Ser-Pro-His-Gln-Gly-Ala-Gly-Met-Val-Gly-Lys-Val-Thr-Val-Asn
Thr-Glu-Lys-Glu-Thr-Tyr-Lys-Phe-Tyr-Cys-Ser-Pro-His-Gln-Gly-Ala-Gly-Met-Val-Gly-Lys-Val-Thr-Val-Asn
Thr-Glu-Lys-Gly-Ser-Tyr-Ser-Phe-Tyr-Cys-Ser-Pro-His-Gln-Gly-Ala-Gly-Met-Val-Gly-Lys-Val-Thr-Val-Asn
Thr-Glu-Ala-Gly-Thr-Tyr-Ser-Phe-Tyr-Cys-Ala-Pro-His-Gln-Gly-Ala-Gly-Met-Val-Gly-Lys-Val-Thr-Val-Asn
Ser-Glu-Lys-Gly-Thr-Tyr-Thr-Phe-Tyr-Cys-Ala-Pro-His-Gln-Gly-Ala-Gly-Met-Val-Gly-Lys-Val-Thr-Val-Asn
Asp-Ala-Lys-Gly-Thr-Tyr-Lys-Phe-Tyr-Cys-Ser-Pro-His-Gln-Gly-Ala-Gly-Met-Val-Gly-Gln-Val-Thr-Val-Asn
Asp-Thr-Lys-Gly-Thr-Tyr-Ser-Phe-Tyr-Cys-Ser-Pro-His-Gln-Gly-Ala-Gly-Met-Val-Gly-Lys-Val-Thr-Val-Asn
Thr-Glu-Lys-Gly-Ser-Tyr-Ser-Phe-Tyr-Cys-Ser-Pro-His-Gln-Gly-Ala-Gly-Met-Val-Gly-Lys-Val-Thr-Val-Asn
Thr-Glu-Lys-Gly-Thr-Tyr-Ser-Phe-Tyr-Cys-Ala-Pro-His-Gln-Gly-Ala-Gly-Met-Val-Gly-Lys-Val-Thr-Val-Asn
Ser-Thr-Pro-Gly-Thr-Tyr-Thr-Tyr-Tyr-Cys-Thr-Pro-His-Gln-Gly-Ala-Gly-Met-Lys-Gly-Thr-Lys-Ile-Val-Trp
```

APPENDIX IVB

Variability of each residue position for plastocyanin data set

Underlined residues are unique to the algal sequences

Numbering is done as for higher plant sequences

\+ insert -Pro-Ala- here (Aitken, 1975)

* insert -Pro- here (Aitken, 1975)

-2	-1	1								10									20					
		Ala	Asp	Ile	Glu	Leu	Gly	Ala	Asp	Asp	Gly	Ala	Leu	Ala	Phe	Glu	Pro	Ala	Asp	Phe	Glu	Ile	Ala	Ala	Gly	Asp
Glu	Ala	Phe	Glu	Leu	Lys	Met		Asp	Glu	Glu		Glu	Phe	Val		Ile		Asp	Glu	Ile	Asn	Leu	Asp	Lys		Glu
	Asp	Ile	Leu	Val	Leu	Val		Gly	Gly	Gly		Gly		Lys		Leu		Gly	Lys	Leu	Gln	Val	Lys	Pro		
	Thr	Leu	Thr		Asn			Ser	Asn	Lys		Leu		Gln		Ser		Lys	Asn	Val	Ser		Asn	Ser		
		Met	Lys		Ile				Ser	Asn		Asn				Val		Asn	Gln		Thr		Pro	Val		
		Gln								Ser		Ser				Asn		Gln	Ser		Val		Ser			
		Val										Thr				Thr		Ser	Thr				Thr			
		Tyr										Val				Tyr		Glu	Ala							

				30									40										50			
Glu	Ile	Glu	Phe	Ile	Asn	Asn	Ala	Ala	Phe	Pro	His	Asn	Ile	Ile	Phe	Asp	Ala	Ala	Ala	Ile	Pro	Ala	Gly	Ala	Ala	Ala
Lys	Val	Ile	Trp	Lys	Leu	Phe	Lys	Gly	Pro	Gly			Val	Val		Pro	Glu	Asp	Glu	Asn		Ala	Lys	Ser	Asp	Asp
Ser		Thr	Asn	Leu		Val	Ser	Val	Tyr								Ile	Pro	Leu	Val		Ser		Val	Asn	Ser
Thr		Val		Val			Gly	Glu	Lys											Gly		Pro		Pro	Val	
Pro				Thr					Thr																	
Ala																										

						60									70										
Ala	Ala	Ile	Ser	-	-	Glu	Ala	Asp	Leu	Leu	Met	Ala	Ala	Gly	Glu	Ser	Thr	Ala	Ala	Ala	Phe+	Asp	Ala*	Ala	Gly	Glu
Asp	Lys	Lys		Leu	Ala	His	Asp	Glu	Tyr		Asn	Ser	Pro	Glu	Gln	Thr	Tyr	Glu	Thr	Lys	Leu	Ser	Glu	Lys		Ser
Glu	Glu	Leu		Met	Asp	Gln	Glu	Leu			Ser	Gln	Asn		Asp	Val	Phe	Lys	Val	Asn	Val	Thr	Thr	Ser		Thr
Gly					Asn		Lys	Gln								Asp		Ser		Thr				Pro		
Leu					Pro		Asn											Val								
Ser					Ser													Thr								

80									90								100	
Tyr	Gly	Phe	Phe	Cys	Ala	Pro	His	Ala	Gly	Ala	Gly	Leu	Lys	-	Lys	Ile	Thr	Val	Ala	-
	Lys	Val	Tyr		Glu			Gln				Met	Val	Gly	Gln	Val	Ile		Asn	Gly
	Ser	Tyr			Ser			Arg							Thr				Gln	
	Tyr				Thr														Trp	

APPENDIX IVC

Invariance of amino acids within the major plant groups

A gap indicates that more than one amino acid occurs at that position within the members of that group.

Residues invariant between all groups are outlined.

The sequence of *Osmunda regalis* is shown complete and not combined with members of the polypodiaceae

	1				5				10				15				20				25				30				35				40
Green Algae			Val		Leu	Gly		Asp		Gly		Leu		Phe		Pro	Ser		Val				Ala	Gly	Glu					Asn	Asn	Ala	Gly	Phe	Pro	His	Asn	Ile	Val
Equisetum	Ala	Glu	Val	Ile	Met	Gly	Leu	Asp	Asp	Gly	Ser	Leu		Phe		Pro			Val		Val	Lys	Ala	Gly	Glu	Lys	Ile	Thr			Asn		Ala	Gly	Phe	Pro	His	Asn	Val	Val
Osmunda	Ala	Asp	Val	Ile	Met	Gly	Gly	Asp	Asp	Gly	Ser	Leu	Ala	Phe	Ile	Pro	Asn	Lys	Ile	Val	Val	Ser	Val	Gly	Glu	Pro	Ile	Thr	Phe	Lys	Asn	Asn	Ala	Gly	Phe	Pro	His	Asn	Val	Val
Higher Ferns	Ala	Lys	Val	Glu	Val	Gly	Asp	Glu	Val	Gly		Phe	Lys	Phe		Pro			Ile	Thr	Val		Ala	Gly	Glu			Glu	Phe	Thr	Leu	Val	Gly	Glu	Thr		His	Asn		Val
Gymnosperms		Glu	Val	Leu		Gly	Gly	Asn	Gly	Gly	Glu	Leu		Phe	Ile	Pro			Phe					Gly			Ile		Phe	Lys	Asn	Asn	Ala		Phe	Pro	His	Asn	Val	
Higher Plants					Leu	Gly			Asp	Gly		Leu		Phe		Pro			Phe					Gly			Ile		Phe	Lys	Asn	Asn	Ala	Gly	Phe	Pro	His	Asn		Val