Molecuiar evolution of cytochrome c from invertebrates

Lyddiatt, Andrew

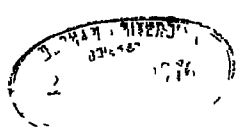
MOLECULAR EVOLUTION OF CYTOCHROME c

FROM INVERTEBRATES

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

By

Andrew Lyddiatt



January 1975.

Department of Botany.

CONTENTS

	Page No.
Summary.	(i)
Acknowledgements.	(iii)
Abbreviations.	(iv)
INTRODUCTION.	1
MATERIALS AND METHODS.	
I. General.	12
II. Cytochrome c purification methods.	29
III. Methods of protein sequence determination.	37
IV. Calculations based on amino acid sequence data.	55
RESULTS.	
1. Purification of *Locusta* cytochrome c.	58
2. Purification of *Macrobrachium* cytochrome c.	60
3. Purification of *Eisenia* cytochrome c.	64
4. Purification of *Asterias* cytochrome c.	70
5. Purification of *Loligo* cytochrome c.	75
6. Extraction of cytochrome c from Solaster.	77
7. Extraction of cytochrome c from *Aphrodite*.	77
8. The amino acid sequence of *Locusta* cytochrome c.	78
9. The amino acid sequence of *Macrobrachium* cytochrome c.	97
10. The amino acid sequence of *Eisenia* cytochrome c.	118
11. The amino acid sequence of *Asterias* cytochrome c.	133

12. Preliminary sequence investigation of *Arenicola* cytochrome *c*. 150

13. The amino acid composition of *Arenicola* cytochrome *c*. 152

14. The amino acid composition of *Nereis* cytochrome *c*. 152

15. The amino acid composition of *Loligo* cytochrome *c*. 152

16. Calculations based on a matrix of amino acid differences. 153

17. Molecular phylogenies constructed using ancestral sequence methods. 154

18. Molecular phylogenies constructed using numerical methods. 158

19. A summary of the results of phylogenetic analysis. 160

DISCUSSION.

I. Methods of extraction and purification of invertebrate cytochrome *c*. 161

II. The methods of protein sequence determination. 173

III. Comparisons of cytochrome *c* sequences. 181

IV. Structure and function in invertebrate cytochrome *c*. 188

V. Aspects of molecular evolution. 198

VI. Assumptions and errors in computed phylogenies. 215

VII. Classical and molecular phylogenies. 217

BIBLIOGRAPHY. 231

APPENDICES.

SUMMARY

Cytochrome c has been extracted and purified from five species of invertebrates, Locusta gregaria, Macrobrachium malcomsonii, Eisenia foetida, Asterias rubens and Loligo opalescens, and pilot extractions were carried out using Solaster papposus and Aphrodite aculeata. The complete amino acid sequences of Locusta, Macrobrachium, Eisenia and Asterias have been determined, and a preliminary sequence investigation was carried out on the cytochrome c from Arenicola marina. The amino acid compositions were determined for Arenicola, Nereis virens and Loligo. The sequence data obtained in this study showed that the cytochromes c from invertebrates are clearly homologous with other eukaryotic mitochondrial cytochromes c. A comparison of the sequences of invertebrate cytochrome c with those of vertebrates, including horse-heart and bonito, for which the three-dimensional structure has been determined, demonstrated the conservative nature of the majority of the amino acid changes and indicated that the tertiary structures were essentially similar. Invertebrate cytochromes c showed high variability at certain positions in particular regions of the molecule, and these differed from the regions of high variability characteristic for the vertebrates and those characteristic for the plants and fungi.

Sequence comparisons of forty species, including ten

invertebrates, were used to estimate the times of divergence of the major invertebrate phyla from the vertebrate line of descent. Molecular phylogenies were constructed by the application of an ancestral sequence method and a numerical matrix method using the invertebrate sequence data. Selected fungal, algal and lower vertebrate sequences were used to fix the invertebrates in the phylogenies with regard to the major kingdoms of organisms. The molecular phylogenies constructed using the ancestral sequence method agreed, within the estimated limits of error, with the broad view of classical phylogeny. However, they lacked the resolution required to establish definite conclusions regarding the relationships of the major invertebrate phyla. The numerical matrix method constructed phylogenies less in agreement with the classical view. It was concluded that the small size of the data set was responsible for the estimated errors in the phylogenies, the poor resolution in certain areas of particular phylogenies and the problems encountered with the computation of sequences of unequal length.

ACKNOWLEDGEMENTS

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

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

I wish to thank Mr. G. Bainbridge and Mr. J. Gilroy for technical assistance, and Miss M. Berryman for typing this thesis.

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

ABBREVIATIONS

The abbreviations used in this thesis are as recommended in "Instructions to Authors", Biochemical J. 131, 1 (1973), with the following additions:-

$CySO_3$: Cysteic acid.
MeS : Methionine sulphone.
C3+ : Cytochrome c (oxidised).
C2+ : Cytochrome c (reduced).

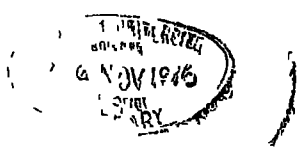
Amino acids are referred to using the recommended abbreviations, and the single letter code shown below:-

Amino Acid.	Abbreviations.	Code Letter.
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asp	D
Aspartic acid	Asn	N
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
ε-N-trimethyllysine	Me_3Lys	J
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

INTRODUCTION

Systems of classification have been devised in a number of ways, but taxonomists have never agreed as to whether the main task of a classificatory scheme is to group organisms possessing a maximum similarity of classification characters, or to erect a scheme that reflects the evolutionary relationships of the organisms. The qualities of these phenetic and phyletic classifications have been widely discussed (see Sokal & Sneath, 1963; Cronquist, 1968), but whichever view is taken, the classification of a given organism must depend on the quantity and quality of the information available to the taxonomist.

At present, the only sure way of establishing a well-documented phylogenetic relationship within a group of organisms, is the study of an adequate fossil record. This approach has been extensively applied to the sub-phylum Vertebrata of the animal kingdom, for which lines of descent and divergence have been determined (Romer, 1966; Colbert, 1969). The divergence of the major invertebrate phyla dates from a time for which an adequate fossil record is absent, in the Cambrian period and earlier. Thus, current views on invertebrate phylogeny depend as much on the morphology and embryology of extant species as on palaeontological evidence. A similar problem exists with the flowering plants because by the time the flowering plant fossils become abundant in the



Upper Cretaceous, many of the present-day orders were already represented (Walton, 1953).

Biochemical information may be used for the construction of phylogenetic schemes, but care must be taken to select suitable information in order that any derived taxonomy does not have similar limitations to schemes based on morphological data, where distortions due to convergence and variable rates of evolution are possible. Initially, biochemical profiles of organisms were constructed using presence or absence data for particular micromolecules, but more recently macromolecules have been widely investigated to provide information on, what may be broadly termed, comparative biochemistry and physiology.

When DNA was shown to be the genetic material it was clear that particular base pair sequences in DNA determined, through the production of specific messenger RNA, the amino acid sequences of particular proteins, and that all these sequences contained within their structures a record of the evolutionary history of the organism concerned. Zuckerkandl & Pauling (1965) recognised that the total chemical content of any species was a document of its evolutionary history, and they classified molecules on the relative amounts of information they contained. The first class, the semantides, consist of molecules which carry the genetic information or a transcript of that information either in the form of nucleic acids or proteins. The second class are the

episemantic molecules, which are synthesized under the control of semantides, and the third class are the asemantides whose molecules are not produced by the organism and do not, except by presence or absence, express any information about the organism. Zuckerkandl & Pauling (1965) inferred that the amount of evolutionary history preserved in a molecule would be the greater, the greater the complexity of the molecule and the smaller the parts of the molecule that have to be affected to bring about a significant change. Thus the largest information content lies within the semantides and these may be divided into the primary semantides, which are the genes of an organism, the secondary semantides, comprising messenger and transfer RNAs, and the tertiary semantides comprising the synthesized polypeptides.

Clearly, as with morphological characters, a study of biochemical characters towards the establishment of a phylogeny can have weak or strong taxonomic significance (see Davis & Heywood, 1963). Many episemantic molecules are formed by multi-enzyme controlled pathways, providing the potential for a convergence of structure and function which may imply false phylogenetic relations. Similarly, more than one biosynthetic pathway may form a particular metabolite, but the possession of that metabolite need not imply a phyletic relationship between organisms (see Bartnicki-Garcia, 1970). In the plants, many episemantides have been

used in taxonomic studies (see Swain, 1963) and in the algae, lipid metabolism and the occurrence of a variety of pigments have been used to infer phyletic relationships (Nichols, 1970; Stanier, 1974). Among the invertebrates, studies on the distribution of sterols and phosphagens have been reported, but the results are rather inconclusive (see Kerkut, 1960). In general, the study of episemantides may provide as much conflicting information as more classical investigations of morphology, palaeontology, physiology and embryology.

In providing the primary genetic link between successive generations, the semantides must contain within their structures the best account of evolutionary history at the molecular level. The characterization and analysis of primary and secondary semantides is becoming less of a technical problem (see Dayhoff, 1972; Dayhoff, 1973; Wu et al., 1974), but the main stumbling block to phylogenetic studies remains in the preparation of homologous portions of DNA from a range of species. Nucleic acid hybridization techniques are the only methods currently available for obtaining phylogenetic information from these categories of molecule (see Kohne, 1968; Gillespie, 1968; Bøvre & Szybalski, 1971), but the methods have practical difficulties, and the interpretation of results is complicated by the occurrence of inversions, translocations and repetitions of DNA base sequences during evolution.

Investigations into the tertiary semantides, or proteins, appear to offer the best approach for the phylogenetic study of macromolecules at present. The use of serology, comparative enzymology, amino acid composition and electrophoretic and chromatographic characteristics, have all been used for such comparisons (Bryson & Vogel, 1965; Boulter et al., 1966; Hawkes, 1968; Nolan & Margoliash, 1968; Vaughan, 1968a and b). Serological methods are rapid, systematically valuable, and can distinguish between antigenic substances indistinguishable by other chemical means. Serology in plant taxonomy was pioneered by Mez in the 1920s (see Chester, 1937), and the application of modern serological techniques to certain taxonomic problems has proved successful (Vaughan, 1968a and b). In the vertebrates, widespread work has been done, including phylogenetic studies on immunoglobulins (Hill et al., 1966), and lens extracts in fish (Manski et al., 1967), but in the invertebrates the work is limited, although some work has been reported on the phylogeny of immune responsiveness (Hildeman, 1974).

Comparative biochemical studies on enzymes and functional proteins provide a further approach to taxonomic problems, and have the advantage that test systems are likely to have changed little during evolution and, consequently, retain the properties of the systems in ancestral organisms. Clearly it is preferable for the components of such test systems to

occur widely in living organisms, and to this end the oxidative electron transport mechanism of aerobic organisms has received much attention, particularly cytochrome c, an electron carrier in such a system. Yamanaka and co-workers (Yamanaka, 1966) established phylogenetic relationships on the basis of the relative activities of cytochrome c purified from a variety of species with Pseudomonas and bovine cytochrome oxidases. They assumed that the closer an organism was, in evolutionary terms, to the oxidase source organism, then the more reactive its cytochrome c would be in a test system (Yamanaka & Okunuki, 1963). The method requires only small quantities of cytochrome c and Yamanaka's work is of interest because of the use of invertebrate material and the demonstration of the low yields of cytochrome c to be expected from these animals (Yamanaka et al., 1963; Yamanaka et al., 1964a and b).

At present, the greatest insights into molecular evolution could be expected to come from the primary structure determinations of tertiary semantides, but two points arise from this approach. The genetic code is degenerate in the respect that changes in base sequences of allelic stretches of a gene need not result in amino acid differences in the resulting polypeptide, and also many regions of the DNA for a particular organism are not expressed in terms of polypeptide product. However, these losses of information are not

sufficient to preclude the use of protein sequence determinations for the study of evolutionary history.

The 'Atlas of Protein Sequence and Structure' for 1966 (Dayhoff & Eck, 1966) listed 184 complete sequences of proteins and related macromolecules, and the same publication for 1972 (Dayhoff, 1972) lists 356 sequences, to which have been added 150 more in a supplement for 1973 (Dayhoff, 1973). Thus, the elucidation of protein primary structures has now become relatively straightforward with the development of techniques to overcome the majority of practical problems (see Needleman, 1970). The determination of the primary structure of the same protein from a variety of organisms can provide information on the history of the gene specifying that protein, and a phylogeny relating those species can be constructed (Fitch & Margoliash, 1967a and b; McLaughlin & Dayhoff, 1973), as has been done for certain ferredoxins, fibrinopeptides, myoglobins, haemoglobins and cytochromes c (see Dayhoff, 1972). The vast majority of proteins examined have been purified from vertebrate sources and although these studies have not led to any great evolutionary insights, the agreement of derived phylogenies with classical morphological and palaeontological evidence (Romer, 1966; Colbert, 1969) has demonstrated the value of the method.

The choice of a suitable protein for such a method is important. It should have a molecular weight of less than

20,000, in order to simplify the sequence determination, and properties of thermal and pH stability, good solubility and ease of assay are all advantageous to the purification. The protein must be evenly distributed throughout the range of species under study, performing an identical function in each case, and must be accepting mutations at a rate which provides a suitable differentiation between those species. Cytochrome c has proved to be a good choice for studies in the vertebrates, higher plants and fungi (Dayhoff, 1972), and the aim of this work was to purify and sequence cytochrome c from invertebrate sources.

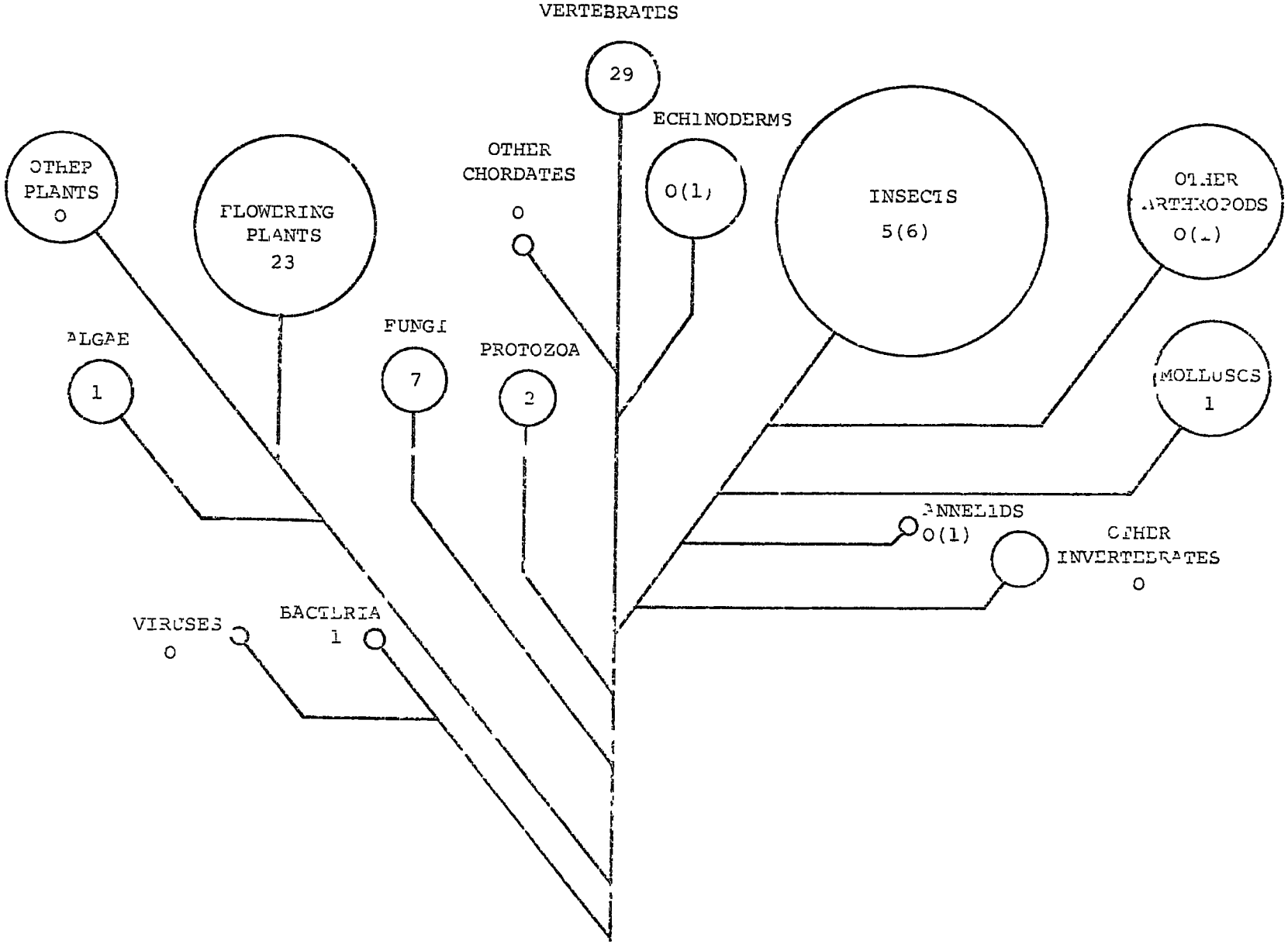
Cytochrome c consists of a single polypeptide chain of molecular weight 13,000 with a covalently bound haem group, and is found on the inner mitochondrial membranes of aerobic eukaryotes functioning as an electron carrier within the terminal oxidation chain (Bachmann et al., 1966). The protein is relatively stable during purification and the history of the refinement of purification methods, together with a history of the early work on cytochrome c, has been reviewed (Keilin, 1966; Lemberg & Barrett, 1973). The availability of pure preparations has enabled the physical and chemical characterisation of the protein from animal, plant, bacterial, protozoan and fungal sources (Margoliash & Schejter, 1966; Lemberg & Barrett, 1973), and a large number of cytochromes c have been sequenced (Dayhoff, 1972; 1973).

A major use of protein taxonomy has been the application of the method to a study of the plant kingdom where no continuous fossil record exists comparable to that for the vertebrates. Boulter and co-workers (Boulter et al., 1972; Boulter, 1973) have constructed a phylogenetic tree relating plant species on the basis of the computer analysis of their cytochrome c amino acid sequences.

The invertebrates are well represented in the fossil record, but the major phyla were established before the end of the Cambrian period so that only radiation within the invertebrate phyla is well documented. Evidence for the proposed evolutionary relationships within the invertebrates has come from the morphology and embryology of extant species as well as palaeontology, and a number of phylogenetic schemes have been proposed to account for the evidence (see Hyman, 1950; Hadzi, 1953; 1963; Marcus, 1958; Hanson, 1961; Appendices IV-XII). For example, doubts exist over the relationships within the lower metazoa, the mono- or multi-phyletic origins of the arthropods and the origin of the chordates. The object of this study was to purify and sequence the cytochromes c from representative species of invertebrates and to use the sequence data to construct a phylogeny relating the invertebrates, thus adding to the published phylogenies for the cytochromes c of vertebrates, higher plants, fungi and micro-organisms (Dayhoff, 1972;

McLaughlin & Dayhoff, 1973). Figure 1 demonstrates how the majority of cytochrome c sequences determined to date have been taken from vertebrate and higher plant sources. Thus, 75% of the published data has been obtained from groups that, in terms of numbers of species, represent less than 35% of the known species of living organisms.

A representative sample of invertebrate cytochrome c sequences, after computer analysis, would provide a molecular phylogeny relating the species from which an estimate could be made of the times of divergence of the major groups within the sample. However, the choice of starting material could not be made entirely on the basis of evolutionary interest because, assuming a minimum of 1-2 μmol of protein for a sequence determination, the majority of invertebrates would be expected to give such low cytochrome c yields that only large-scale extractions could possibly provide sufficient pure protein for sequence analysis. The cytochrome c content of an organism is related to the amount of active muscle tissue it possesses, so that only the flying insects and certain crustaceans could be expected to match the yields of cytochrome c reported for vertebrate heart muscle (see for example, Margoliash et al., 1962; Chan et al., 1967; Augusteyn et al., 1973). In addition, the nature of the organisms under study made necessary the extraction of the total organism rather than a selected tissue, such as heart muscle


VERTEBRATES
29
ECHINODERMS
0(1)
OTHER CHORDATES
0
OTHER PLANTS
0
FLOWERING PLANTS
23
INSECTS
5(6)
OTHER ARTHROPODS
0(1)
ALGAE
1
FUNGI
7
PROTOZOA
2
MOLLUSCS
1
ANNELIDS
0(1)
OTHER INVERTEBRATES
0
VIRUSES
0
BACTERIA
1

from which the majority of vertebrate cytochromes c have been purified (Lemberg & Barrett, 1973). This approach led to the release of gut proteases into the initial purification stages and was thought to account, in part, for the low yields of cytochrome c recorded during the experiments with echinoderm, mollusc and annelid materials.

A large-scale approach to the purification was adopted using similar strategies to the methods developed for cytochrome c purification from higher plants (Richardson et al., 1970; 1971a). This enabled the purification of cytochrome c from the echinoderm, Asterias rubens; the crustacean, Macrobrachium malcolmsonii; the annelid, Eisenia foetida; and the insect, Locusta migratoria, in sufficient quantities to allow the determination of the amino acid sequences. Pilot extraction experiments using the echinoderms, Holothuria forskali and Solaster papposus, the annelids Nereis virens, Arenicola marina and Aphrodite aculeata, and the mollusc Mytilus edulis, demonstrated that the quantities of starting material required and the time factors involved would preclude the use of these species in the investigation.

MATERIALS AND METHODS

I. GENERAL

1. Biological materials

Eisenia foetida (Savigny) (common brandling worm) was obtained from the Worm Farm, Bulwell, Nottingham.

Asterias rubens L. (common starfish) was collected from the estuary at Burnham-on-Crouch.

Macrobrachium malcolmsonii (H. M. Edwards) (freshwater prawn) originated from the Calcutta region, and was obtained through a London fishmonger.

Locusta migratoria L. (common locust) was obtained from the British Museum (Natural History).

Loligo opalescens (Berry 1911) (common squid) originated from Monterey in California, and was obtained through a London fishmonger.

Solaster papposus L. (common sunstar) was collected from the estuary at Burnham-on-Crouch.

Aphrodite aculeata L. (sea mouse) was collected from the estuary at Burnham-on-Crouch.

The invertebrate material was supplied through the British Museum (Natural History) by Dr. R. P. S. Jefferies.

Pseudomonas fluorescens was obtained as a lyophilised sample from Cyclo Chemicals, Los Angeles, through Cambrian Chemicals, Croydon, England.

Ox hearts were obtained from a local slaughterhouse.

2. Chemicals and Reagents

With the exception of those listed below, all chemicals and reagents were obtained from British Drug Houses Ltd., Poole, Dorset, and were of analytical reagent grade whenever available.

Helix aspersa (Müller) (garden snail) cytochrome c was a gift from Dr. R. H. Brown.

Arenicola marina L. (lugworm) and Nereis virens L. (king ragworm) cytochromes c were the gift of Dr. M. Richardson.

Bovine cytochrome c was prepared by the method described in Section II.

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

α-Chymotrypsin E.C.3.4.4.5 (three times recrystallized) 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 the Sigma Chemical Co. Ltd., St. Louis, Mo., U.S.A.

Amberlite C.G.50 (100-200 mesh Type 1)

Amberlite M.B.1

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

Sephadex G-10

Sephadex G-15

Sephadex G-75

Sephadex G-200

CM-Sephadex C-50

Blue Dextran 2000

were obtained from Pharmacia Ltd., Uppsala, Sweden.

Whatman CM-52 Cellulose

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

Biogel P-2

Biogel P-30

were obtained from BioRad Laboratories Ltd., London.

Horse-heart cytochrome c (Type VI)

Standard amino acid mixture

EDTA (free acid)

p-Dimethylaminobenzaldehyde

Protamine sulphate

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

Hydrazine sulphate

Tris-(hydroxymethyl)-methylamine

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

Hydrazine (95%+)

was obtained from Eastman Chemicals Ltd.

Arginylarginine

L-Pyrrolidone carboxylyl-L-alanine

was obtained from Cyclo Chemical Corp., Los Angeles, Cal., U.S.A.

9,10-Phenanthraquinone

Thiodiglycol (TDG)

2-Pyrrolidone

was obtained from Koch Light Labs. Ltd., Colnbrook, Bucks.

Methyl oxitol

was obtained from Shell Chemicals, U.K. ltd.

Ascorbic acid

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

Triethylamine

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

Sequencer kits, comprising Anhydrous n-heptafluorobutyric acid

Quadrol-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.

Pyridine (no-ninhydrin grade)

was obtained from Rathburn Chemicals, Walkerburn, Peebleshire.

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

3. Other Materials

Polyamide sheets were obtained from the Cheng Chin Trading Co. Ltd., Tapei, Taiwan, through BDH Chemicals Ltd.

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

20 cm x 20 cm Kieselgel 60T254 (D.C. Alufolien) TLC plates were obtained from Merck/Darnstadt.

4. Preparation of Solutions

(a) Buffer solutions

(i) 10 mM-aluminium sulphate, pH 4.5 for the extraction of cytochrome c.

5.0 g $Al_2(SO_4)_3 . 12H_2O$ per litre was adjusted to pH 4.5 with 2 M-H_2SO_4 solution.

(ii) 50 mM-sodium phosphate, pH 8.0 for dialysis and CM-50 Sephadex chromatography.

7.80 g $NaH_2PO_4 . 2H_2O$ per litre was adjusted to pH 8.0 with 2 M-NaOH solution.

(iii) 300 mM-sodium phosphate, pH 7.2 for gradient elution during CM-52 cellulose chromatography.

$Na_2HPO_4.12H_2O$		77.40 g
$NaH_2PO_4.2H_2O$		13.10 g
Water	made to	1000 ml

(iv) 200 mM-sodium phosphate, pH 7.2 for the phenylacetaldehyde reagent and elution of CM-52 cellulose.

300 mM-sodium phosphate, pH 7.2, buffer was diluted 1.5-fold and the pH adjusted as necessary.

(v) 10 mM-sodium phosphate, pH 7.2 for gradient elution during CM-52 cellulose chromatography, gel filtration with Biogel P-30 and general dialysis.

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

(vi) 10 mM-sodium phosphate, pH 11.5 for gradient elution during CM-52 cellulose chromatography.

4.40 g $Na_4P_2O_7.10H_2O$ per litre was adjusted to pH 11.5 with 6 M-NaOH solution.

(vii) 75 mM-Tris-KCl, pH 7.5 for molecular exclusion chromatography on Sephadex G-75.

Potassium chloride	7.46 g
Tris (hydroxymethyl) methylamine	6.06 g
Water	made to 1000 ml and

adjusted to pH 7.5 with 6 M-HCl.

(viii) 200 mM-sodium citrate, pH 3.0 for acetyl group determination.

Citric acid.H_2O	3.44 g
Tri-sodium citrate.$2H_2O$	1.06 g
Water	made to 100 ml

(ix) 50 mM-sodium phosphate, pH 7.3 for the extraction of L-pyrrolidone carboxylyl peptidase from Pseudomonas fluorescens and molecular exclusion chromatography on Sephadex G-200.

10 mM-2-mercaptoethanol	1.56 g (1.40 ml)
1 mM-EDTA	0.74 g
100 mM-2-pyrrolidone	17.20 g (15.30 ml)

50 mM-sodium phosphate, pH 7.3 made to 2000 ml and the pH adjusted as necessary.

The 50 mM-sodium phosphate was prepared from 12.9 g $Na_2HPO_4.12H_2O$ and 2.18 g $NaH_2PO_4.2H_2O$ made to 1000 ml with distilled water.

(x) 50 mM-sodium phosphate, pH 7.3 for dialysis of L-pyrrolidonecarboxylyl peptidase preparations prior to enzyme assays and proteolytic digestions.

The buffer was prepared as in (ix) without the 0.1 M-2-pyrrolidone.

(xi) 200 mM-N-ethyl morpholine, pH 8.5 for proteolytic digestions of peptides and proteins.

N-ethyl morpholine	25 ml
Acetic acid	1 ml
Water	made to 1000 ml

and the pH adjusted as necessary.

(xii) 200 mM-ammonium hydrogen carbonate, pH 8.5 for proteolytic digestions of peptides and proteins.

17.0 g NH_4HCO_3 was dissolved in 1000 ml of distilled water, and the pH adjusted as necessary.

(xiii) 200 mM-citrate buffer, pH 5.5 for the ninhydrin analysis of amino acids and proteins.

Citric acid.H_2O	21.0 g
NaOH	8.0 g
$SnCl_2.H_2O$	0.8 g
Water	made to 500 ml

The solution was mixed with one containing 20 g ninhydrin in 500 ml methyl cellosolve immediately before use.

(b) Electrophoresis Solutions

(i) pH 6.5

Pyridine	250 ml
Acetic acid	10 ml
Water	2250 ml

The pH was adjusted as necessary before use.

(ii) pH 1.9

Glacial acetic acid	200 ml
Formic acid (98-100%)	50 ml
Water	2250 ml

The pH was adjusted as necessary before use.

(iii) Electrophoresis standards solution

Arginylarginine was dissolved in 1 M-sodium hydrogen carbonate to give a 0.1 M solution and this was treated with an equal volume of 0.2 M-dansyl chloride in acetone. After incubating at 37°C for 1 h, the solution was diluted 1000-fold and ethanolic dansyl-arginine added to give a concentration of 0.1 mM.

(c) Paper chromatography solutions

(i) BAWP solvent

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

The solvent was freshly prepared for each experiment.

(ii) Chromatography marker solution

A 0.1 mM-solution of dansyl-arginine ethanol was used.

(d) Polyamide sheet chromatography solutions

(i) Solvent I : 1.5% (v/v) formic acid (Woods & Wang, 1967)

(ii) Solvent II : Toluene-acetate acid (9:1 by volume)

(iii) Solvent III: Butyl acetate-methanol-acetic acid (90:60:1 by volume)

(iv) Dansyl-standard solutions

Dansyl-arginine, dansyl-glutamic acid, dansyl-

glycine, dansyl-isoleucine, dansyl-phenylalanine, dansyl-proline and dansyl-serine were dissolved in 95% (v/v) ethanol at a concentration of 0.1 mg/ml.

(e) Thin-layer chromatography solutions

(i) Solvent I : N-heptane-propionic acid-1,2-dichloroethane (58:17:25 by volume)

(ii) Solvent II : N-heptane-butanol-formic acid (50:30:9 by volume)

(f) Electrophoretic and chromatographic location reagents

Guide strips were cut in the direction of the separation on the electrophoretogram or chromatogram, representing about 5-10% of the width, i.e. about 5 to 10% of the material on the paper was used for location purposes.

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

A solution of

Cadmium acetate	100 mg
Acetic acid	5 ml
water	10 ml
Acetone	100 ml

was used to prepare a fresh 1% (w/v) ninhydrin solution through which the paper was passed. After drying in a draught at room temperature the paper was heated under observation at 80°C. Positive reactions were shown as characteristically coloured regions against a white background; some regions developing colour more slowly than others.

(ii) Phenanthraquinone reagent (Yamada & Itano, 1966)

Solution A : 5 mg, 9,10-phenanthraquinone in 25 ml 95% (v/v) ethanol

Solution B : 2.5 g sodium hydroxide dissolved in 70% (v/v) ethanol

The solutions A and B were mixed immediately before use. Guide strips were dipped through the mixture and dried in a draught at room temperature. Positive reactions were observed as fluorescent regions under u.v. light (350 nm), indicating the presence of arginine.

(iii) Phenanthraquinone-ninhydrin reagent

A solution of	Cadmium acetate	100 mg
	Acetic acid	20 ml
	Water	5 ml
	Acetone	80 ml

was used to prepare a fresh 1% (w/v) ninhydrin solution through which the paper was passed. Development and the recording of positive reactions was as in (i).

(iv) Ehrlich reagent

A solution of 2% (v/v) HCl in acetone was used to prepare a fresh 2% (w/v) solution of p-dimethylaminobenzaldehyde through which the paper was passed. The paper was dried in a draught at room temperature and positive reactions, indicating the presence of tryptophan, showed as transient purple regions against a white background.

(v) Pauly reagent

The following solutions were freshly prepared and chilled separately in an ice bath.

Solution A : 1 g sulphanilic acid dissolved in 100 ml of 10% (v/v) hydrochloric acid

Solution B : 100 ml of 5% (w/v) sodium nitrite

Solution C : 100 ml of 10% (w/v) sodium carbonate

Equal volumes of A and B were mixed and lightly sprayed onto the paper. The paper was then lightly sprayed with solution C until the colours formed. Histidine-containing peptides were positively identified as orange-red zones on a yellow background.

(vi) Phenylacetaldehyde reagent

Solution A : Phenylacetaldehyde 0.18 g

Ninhydrin 0.09 g

95% (v/v) ethanol made to 100 ml

Solution B . 0.2 M sodium phosphate, pH 7.2

The paper was sprayed with solution A and dried in a draught at room temperature. The paper was then sprayed with solution B and heated at 60°C for 15 minutes. Positive reactions were observed on the paper under u.v. light (350 nm) as brightly fluorescent regions.

The reagents were used in the order and combinations recommended by Easley (1965). In addition, it was found that reagents (ii), (iii) and (iv) could be used in order and

reagent (vi) was suitable for experiments requiring increased sensitivity.

(g) Amino acid autoanalyser solutions

(i) 4 M-sodium acetate buffer, pH 5.5

Anhydrous sodium acetate	655.85 g
Acetic acid	200 ml
Deionised water (50°C) made to	2000 ml

The pH of the solution was adjusted as necessary.

(ii) Ninhydrin solution

Methyl cellosolve	3125 ml
Sodium acetate buffer	1250 ml
Deionised water	625 ml

The solution was deoxygenated by bubbling through oxygen-free nitrogen before adding 60 g ninhydrin and stirring for 2 hours under nitrogen. 1.667 g of stannous chloride was then dissolved in the solution, which was stored in the dark under nitrogen.

(iii) BRIJ 35 solution

50 g BRIJ was dissolved in 150 ml deionised water (50°C).

(iv) pH 2.2 buffer

Citric acid	21.0 g
Sodium hydroxide	8.4 g
Conc. hydrochloric acid	16 ml
Thiodiglycol	5 ml

Sodium octanoate	0.1 g
Water	made to 1000 ml

(v) pH 3.25 buffer

Citric acid	105.0 g
Sodium hydroxide	41.25 g
Conc. hydrochloric acid	53.25 ml
Thiodiglycol	2.5 ml
BRIJ 35	13.5 ml
Sodium octanoate	0.5 g
Water	made to 5000 ml

The pH was adjusted to pH 3.25 with conc. hydrochloric acid.

(vi) pH 4.25 buffer

Citric acid	42.0 g
Sodium hydroxide	16.5 g
Conc. hydrochloric acid	9.4 g
BRIJ 35	5.4 ml
Sodium octanoate	0.2 ml
Water	made to 2000 ml

(vii) pH 6.65 buffer

Sodium citrate	490.0 g
Conc. hydrochloric acid	7.2 ml
BRIJ 35	13.5 ml
Sodium octanoate	0.5 g
Water	made to 5000 ml

(viii) pH 3.2 loading buffer

pH 3.2 buffer	70 ml
Methanol	30 ml

(ix) Standard amino acid mixture

Standard amino acid mixture (2.5 µmole/ml)	100 µl
Nor-leucine (2.5 µmole/ml)	100 µl
Cysteic acid (2.5 µmole/ml)	100 µl
pH 2.2 buffer	made to 5 ml

5. Preparation L-pyrrolidone carboxylyl peptidase

The method was based on a small scale version of Doolittle & Armentrout (1968) and Armentrout & Doolittle (1969).

(a) Preparation

1 g of lyophilized Pseudomonas fluorescens cells were sonified in 4 ml of 50 mM-sodium phosphate buffer, pH 7.3 containing 10 mM-2-mercaptoethanol, 1 mM-EDTA and 100 mM-2-pyrrolidone using a Dawe Soniprobe (Type 1130A, Dawe Instruments, London) equipped with an ice jacket. The homogenate was centrifuged at 30,000 g for 30 min at 2°C, the supernatant was recentrifuged and both pellets discarded. Protamine sulphate was added to give a 0.14% solution (w/v) which was stirred on ice for 30 min and then centrifuged at 15,000 g for 30 min at 2°C. The supernatant was slowly taken to 42% saturation with ammonium sulphate, using the nomogram of Dixon (1953), and stirred for 1 h on ice before centrifugation at 39,000 g

for 30 min at 2°C. The pellet was resuspended in 0.5 ml of the 50 mM-phosphate buffer at pH 7.3 and loaded on to a G-200 Sephadex column (1 cm x 30 cm), freshly packed and equilibrated in the phosphate buffer. The column was developed at a flow rate of 40 ml/h and the eluate monitored, using an Isco Model 222 recording ultraviolet analyser set for 280 nm. The 'crude enzyme' peak eluted behind the larger 280-absorption of the main protein peak and well in front of the ammonium salt peak. The enzyme peak was isolated by incubating samples of the eluate with L-pyrrolidone carboxylyl-l-alanine (pyr-ala) and analysis by the dansyl chloride or ninhydrin methods (see assay details).

The G-200 enzyme peak was suitable for use in sequence analysis and was precipitated using a 60% saturation of ammonium sulphate. Centrifugation at 39,000 g for 30 min gave a slurry that could be stored at -20°C for up to 3 months.

(b) Enzyme assays

The enzyme slurry was dissolved in about 1 ml of 50 mM-sodium phosphate buffer, pH 7.3, containing 10 mM-2-mercaptoethanol and 1 mM-EDTA, and dialysed against 5 litres of the same buffer for 3 h at 2°C. 50 µl of the enzyme preparation, or 50 µl taken from the G-200 eluate fractions, was mixed with 25 µl of 0.5% (w/v) pyr-ala in a stoppered glass centrifuge tube. Incubation was for 1 h at 37°C and

terminated by the addition of 2 ml 95% (v/v) ethanol. After standing for 30 min the sample was centrifuged at 1000 g for 10 min and the supernatant retained.

Free alanine was estimated semi-quantitatively using 50 µl aliquots of the supernatant using the dansyl chloride method of Gray & Hartley (1963a) (see Materials and Methods III.7(a)). Substrate and enzyme control experiments were conducted and the dansyl derivatives identified using polyamide sheet chromatography (see Materials and Methods III.7(b)).

Free alanine was estimated quantitatively using a modified method of Hirs (1956) and enzyme activity was estimated from the appropriate ninhydrin-alanine calibration curve. One unit of enzyme activity was defined as that amount of enzyme producing 1 µmole of free alanine per minute. Relative protein concentration was measured as the optical density of 280 nm and the specific activity of the enzyme was expressed as 1000 times the amount of alanine produced per minute (nmol) under the assay conditions divided by the $O.D._{280}$ of the enzyme solution (Armentrout & Doolittle, 1969).

II. CYTOCHROME c PURIFICATION METHODS

A generalised purification scheme is shown in Figure 2.

1. Cytochrome c assay

Cytochrome c was assayed qualitatively in solution by the addition of a trace of ascorbic acid and the observation of the α-band absorption at 550 nm using a low dispersion, direct vision hand spectroscope (R. & J. Beck Ltd., London).

Quantification of cytochrome c was by a spectrophotometric method using silica cells of 1 cm light path on a Unicam SP800A or SP1800, or a Perkin Elmer 402 recording spectrophotometer. Assuming a molecular weight of 13,000 and the mammalian cytochrome c extinction coefficient of 27.7 $mM^{-1}cm^{-1}$, cytochrome c concentrations were estimated from the α-absorption at 550 nm (Margoliash & Frohwirt, 1959).

The purity of samples was estimated spectrophotometrically from the ratios of the 280 nm, Soret (γ, 410 nm) and β-band (550 nm) absorption for the protein in the oxidised and reduced states, reduction being effected by the addition of a trace of ascorbic acid or sodium dithionite. The ratios were compared with those of cytochromes c of known structure and purity (Keilin, 1966; Margoliash & Schejter, 1966).

2. Cytochrome c extractions

All stages of extraction were carried out at 0-4°C.

Animals were extracted in separate batches of 3-30 kg, each batch being rapidly thawed at 4°C and carefully washed

FIGURE 2.

A generalised scheme for the purification of Invertebrate cytochrome c.

Frozen material rapidly thawed at 4°C.

Blended with 10 mM-aluminium sulphate and ice.

Stirred at pH 4.5 for 2 h.

Filtered through muslin/basket centrifuge.

Adjusted to pH 8.0 - re-filtered as necessary.

Adsorbed on Amberlite CG-50 resin.

Eluted with 2 M-NaCl.

Dialysed.

Concentrated on CM-Sephadex resin.

Eluted with 0.5 M-NaCl.

Ammonium sulphate fractionation.

Gel-filtration on Biogel P-30.

Chromatography on CM-52 cellulose.

Desalted on G-10 Sephadex/Amberlite MB-1.

Pure cytochrome lyophilised.

with running tapwater. One volume of animals was blended with 2 vol of chilled 10 mM-aluminium sulphate (Margoliash & Walasek, 1967) and 1 vol of ice crystals in a 5 litre Waring blendor set at maximum speed for 5 min. The pH of the total homogenate (30-300 litres) was adjusted to pH 4.5 with 2 M-sulphuric acid and the extract left to stir for 2 h at 2-4°C before being stirred through two layers of muslin. The resulting cake was re-extracted whilst the crude filtrate was passed through a Terylene bag (Type 1481F, Samuel Hall Ltd., Rochdale, Lancs.) fitted in a 21" perforated basket centrifuge (Type 86, Thomas Broadbent and Sons Ltd., Huddersfield). The filtrate was adjusted to pH 8.0 with 2 M-Tris and any further precipitation removed by basket centrifugation or filtration under vacuum through Whatman No. 6 paper on 27 cm Büchner funnels.

3. Ion-exchange chromatography on Amberlite CG-50 resin

(a) Regeneration and preparation of resin.

The resin was prepared for use in the ammonium form. New or used resin was stirred overnight in 4-5 vol of acetone and washed on a sintered glass funnel with 20 vol of distilled water, before conversion to the sodium form by vigorous stirring in 4-5 vol of 2 M-NaOH at 80°C for 5-6 h. After extensive washing with 20 vol of distilled water, the resin slurry was maintained at pH 1.0 for 3-4 h with conc. H_2SO_4 before washing with a further 20 vol. of distilled water. The resin was

suspended in 5 vol of distilled water and stirred at pH 10 with concentrated ammonia solution for 2 h before a final washing with 20 vol of distilled water. The regenerated resin was kept as a slurry at 2-4°C.

(b) Adsorption.

Ion-exchange chromatography was carried out in a number of ways in order to cope with the variable quality of the pH 8.0 filtrate. Clean filtrates were easily handled but those extracts which filtered poorly were treated batchwise, whilst those showing signs of bacterial contamination were treated speedily using pressure methods.

(i) Funnel method

An 18 cm x 10 cm sintered glass funnel was packed with Amberlite CG-50 and the pH 8.0 filtrate was pumped through the resin at flow rates of 15-20 l/h at 2-4°C. The resin capacity was judged to be 20-30 litres.

(ii) Column method

Filtrate at pH 8.0 was passed through 6 cm x 15 cm columns of Amberlite CG-50 at 2-4°C at flow rates of 500 ml-1500 ml/h. Column adsorption capacity was judged to be 15-20 litres of filtrate.

(iii) Columns under pressure

These were run as in (ii) but under a negative pressure of 5-10 mm Hg. The columns were repacked frequently

to maintain the flow rates.

(iv) Batch method

1 vol of Amberlite CG-50 was stirred with 50 vol of pH 8.0 filtrate for 2 h at 2-4°C. The resin was allowed to settle out and was reclaimed by decantation, the treatment being repeated on the supernatant. In situations where the resin would not settle out, it was reclaimed by the passage of the suspension through an M.S.E. continuous action rotor fitted to an M.S.E. High Speed 18 refrigerated centrifuge operating at 18,000 g and a flow rate of 30 l/h.

(c) Elution.

The resin was exhaustively washed with distilled water until the washings became clear and then eluted batchwise. The resin was suspended in 2 vol of 2 M-NaCl and stirred at 0-4°C for one hour maintained at pH 8.0 with 2 M-NaOH. The resin was drained on a sintered glass funnel and washed with 2 M-NaCl until the washings, after reduction with a trace of ascorbic acid, showed no 550 nm absorption using a direct vision, low dispersion hand spectroscope.

4. Chromatography on CM-Sephadex

The eluate from (3) (5-15 l) was adjusted to pH 8.0 and dialysed against running tap water for 8 h and 60 l 50 mM-sodium phosphate buffer at pH 8.0 and 2-4°C for a further 8 h. The dialysis residue was centrifuged as necessary and

passed through a 6 cm x 10 cm column of CM-50 Sephadex equilibrated in 50 mM-sodium phosphate, pH 8.0 at 2-4°C, with a flow rate of 200-400 ml/h. The column was washed with 1 litre of dialysis buffer and the cytochrome eluted as a single fraction (100-1000 ml) with 0.5 M-NaCl in 50 mM-sodium-phosphate buffer at pH 8.0.

After a dialysis step, the chromatography step was repeated using a 2 cm x 8 cm column of CM-50 Sephadex and a flow rate of 100-200 ml/h. Elution with 0.5 M-NaCl in 50 mM-sodium phosphate, pH 8.0, gave a volume of 10-40 ml.

5. Ammonium sulphate fractionation

The cytochrome c (10-40 ml) was reduced with a trace of ascorbic acid and maintained at pH 8.0 and 2-4°C. Finely ground ammonium sulphate was slowly added to give a 50% saturated solution, the degree of saturation being determined from the nomogram of Dixon (1953) ignoring the presence of any salt originally in the solution. The solution was stirred at pH 8.0 and 2-4°C for 2 h, or overnight where possible, and then centrifuged at 30,000 g for 20 min. The pellet was discarded. The salt concentration was slowly increased in 10% steps up to 80% with the appropriate centrifugations. Pellets were examined for precipitated cytochrome c and washed with a minimal volume of distilled water as necessary. The increases in saturation were stopped when precipitation ceased and the supernatant containing the cytochrome c was

retained.

6. Concentration step

The supernatant was diluted twice and exhaustively dialysed against 10 mM-sodium phosphate, pH 7.2, and concentrated on a 1 cm x 5 cm column of CM-52 cellulose equilibrated in the same buffer. The sample was eluted with 200 mM-sodium phosphate, pH 7.2, to give a volume of 2-5 ml.

7. Gel filtration on Biogel P-30

Biogel P-30 resin was equilibrated in 10 mM-sodium phosphate buffer, pH 7.2, and 2-4°C and packed into a column (3 cm x 90 cm or 2 cm x 90 cm) with a hydrostatic head of about 20 cm. The sample was loaded in a volume of 2-5 ml and the column developed at a flow rate of 40-60 ml/h.

8. CM-52 Cellulose chromatography

The column (1 cm x 15 cm, 2 cm x 15 cm or 1 cm x 30 cm) was packed from a slurry of CM-52 Cellulose equilibrated in 10 mM-sodium phosphate, pH 7.2, by pumping through buffer using an LKB Varioperspex peristaltic pump (12000) at a flow rate of 20-40 ml/h. The cytochrome c solution was diluted to an appropriate ionic strength and pumped on to the column followed by 15 ml of 5 mM-potassium ferricyanide dissolved in starting buffer. The layer of adsorbed cytochrome c on top of the column was stirred carefully to prevent "panning" and the column washed free of ferricyanide with a further

20 ml of 10 mM-sodium phosphate, pH 7.2. The column was eluted under a linear ionic gradient from 10 mM to 300 mM-sodium phosphate at pH 7.2, or under a linear pH gradient from pH 7.2 to pH 11.5 using 10 mM-sodium phosphate buffers. The gradients were constructed using a device as described by Bock & Ling (1954).

9. Concentration step

The cytochrome c solution was adjusted with regard to ionic strength or pH and pumped onto a 1 cm x 15 cm column of CM-52 cellulose equilibrated in 10 mM-sodium phosphate buffer, pH 7.2. The sample was eluted with 200 mM-sodium phosphate, pH 7.2, to give a volume of 2-5 ml.

10. Gel filtration on Sephadex G-75

The column (3 cm x 100 cm) was packed using a slurry of G-75 Sephadex equilibrated in 75 mM-Tris-KCl buffer at pH 7.5. The sample was pumped on in a volume of 2-5 ml and the column developed in an ascending direction at a flow rate of 30 ml/h.

11. Desalting by gel filtration

Pure cytochrome c preparations were concentrated by freeze-drying over NaOH and conc. H_2SO_4. The lyophilised sample was taken up in a minimal volume of distilled water and desalted in 1 ml batches on a 1 cm x 10 cm column of G-10 Sephadex or Biogel P-2 equilibrated in distilled water. The salt-free cytochrome c was lyophilised and stored at -20°C.

12. Desalting by an ion-exchange method

Pure cytochrome c preparations were concentrated by freeze-drying over NaOH and conc. H_2SO_4. The lyophilised sample was taken up in a minimal volume of distilled water and desalted in 1 ml batches on a 1 cm x 15 cm column of Amberlite MB-1 resin equilibrated in distilled water. Salt-free cytochrome c samples were lyophilised and stored at -20°C.

III. METHODS OF PROTEIN SEQUENCE DETERMINATION

1. Denaturation of cytochrome c

The method was based on that of Margoliash et al. (1962). The salt-free cytochrome c was dissolved in distilled water (5 mg/200 µl) and oxidised by the addition of 1 µl saturated potassium ferricyanide solution. The solution was made to 95% ethanol (v/v) by the addition of absolute ethanol and left to stand at room temperature for 4-5 h, the precipitate being resuspended hourly. The precipitate was separated by centrifugation, washed three times in 95% ethanol and after a final centrifugation dried in vacuo.

2. Proteolytic digestion

Denatured cytochrome c (0.5 µmol) was suspended in 1 ml water and adjusted to pH 8.0 with 50 mM-ammonia solution. The protein solution (1-2 ml) was equilibrated at pH 8.0 and 37°C on a Radiometer TTTlc Autotitrator fitted with a temperature compensator. 25 mM-sodium hydroxide was used to maintain the pH at 8.0 and the solution was kept under a stream of oxygen-free nitrogen.

(a) Chymotryptic digestion.

The chymotrypsin was dissolved in distilled water and added to the equilibrated cytochrome c solution to give a concentration of 2% (w/w) at zero time. A further 2% was added at 60-90 min and the digestion terminated at 60-120 min by adjusting the pH to 4.5 with 50-100 µl of acetic acid.

The digest was then frozen and lyophilised.

(b) Tryptic digestion.

The digest conditions and enzyme concentrations were as for the chymotryptic digestion, except that the trypsin was dissolved in 1 mM-hydrochloric acid.

Rates of digestion were recorded in both cases on a Radiometer SBR2C recorder connected to the autotitrator.

3. Cyanogen bromide cleavage

Lyophilised cytochrome c was dissolved in 200-400 µl of 70% (v/v) formic acid. A 150 molar excess of cyanogen bromide was added in 200-400 µl of 70% (v/v) formic acid and incubation was for 24 h at 26°C in the dark. The reaction was terminated by lyophilisation and the products were separated by passage through a 1 cm x 100 cm column of G-50 Sephadex equilibrated and eluted with 70% (v/v) formic acid.

4. Peptide purification

(a) Electrophoresis.

Peptides were separated by high-voltage electrophoresis at pH 6.5 on a flat-plate apparatus (106 cm x 15 cm; the Locarte Co., London SW3, U.K.) using a pyridine-acetic acid-water buffer (25:1:225 by vol.) and Whatman 3MM paper (15 cm). The applied voltage was 9 Kv to give a current of 30-50 mA for 90-120 min at 7 p.s.i. pressure. Peptides requiring further purification were treated to electrophoresis at pH 1.9 using an acetic acid-

formic acid-water buffer (4:1:45 by vol.). The applied voltage was 9 Kv to give a current of 50-60 mA for 60-90 min at 7 p.s.i. pressure.

(b) Paper chromatography.

Peptides were separated on Whatman 3MM chromatography paper (55 cm x 46 cm). The mixtures were loaded at 8 cm streaks at the origin and separated in a descending manner using the BAWP solvent system (butanol-acetic acid-water-pyridine, 15:3:12:10 by vol.) and Gallenkamp chromatography frames and bags. The samples were chromatographed for 18 h at 24°C, and the papers dried at 45°C. Where necessary, dried papers were rechromatographed for a further 18 h.

(c) Peptide location.

Peptides were located using 10% strips of the electrophoretogram or chromatogram. The detection reagents were used in the orders and combinations recommended by Easley (1965), except it was found that the phenanthraquinone reagent could be used before the modified ninhydrin and Ehrlich reagents. The phenylacetaldehyde reagent was used where increased sensitivity was required.

(d) Peptide mobilities.

(i) Electrophoresis

Peptide mobilities at pH 6.5 were measured in positive and negative directions from a true neutral point

and expressed relative to the true neutral co dansyl-arginyl-arginine distance. Zones were measured from their mid-point. The true neutral position was a point estimated to be 4/11 of the dansyl-arginylarginine to the 1-dimethyl-aminonaphthalene-5-sulphonic acid distance measured from the dansyl-arginyl-arginine.

At pH 1.9, the mobilities were measured from the 1-dimethylaminonaphthalene-5-sulphonic acid and expressed relative to the distance of the dansyl-arginine standard.

(ii) Chromatography

Peptide mobilities were measured for the origin and expressed relative to the distance moved by the dansyl-arginine standard.

(e) Peptide elutions.

After localisation of the peptides as described, the chromatograms and electrophoretograms were cut into the appropriate strips and eluted for 4 h into Pyrex screw-cap tubes (1 cm x 5 cm) using 20% (v/v) pyridine to give a volume of 200-500 µl. The strips were dried at room temperature and re-eluted for a further 4 h. The eluted samples were lyophilised and stored at -20°C.

(f) Heme peptide purification.

Heme peptides were treated normally. No precipitation at pH 4.5 or differential elution with water at 2°C and 20%

(v/v) pyridine at room temperature, was possible (see Ramshaw, 1972). Paper chromatography proved to be the most effective final purification step, tending to minimize the streaking of the heme zone.

5. Quantitative amino acid composition of proteins and peptides

The amino acid compositions of proteins and peptides were determined quantitatively using a Locarte automatic amino acid analyser.

Protein (50-200 µg), or peptide (0.025-0.25 µmol), was hydrolysed with 0.5 ml constant boiling 5.7 M-HCl in evacuated Pyrex tubes (Moore & Stein, 1963). Protein samples were hydrolysed for 24, 48 and 72 h in order to obtain zero time values for serine and threonine and maximum values for valine, isoleucine and leucine (Moore & Stein, 1963). Peptide samples were hydrolysed for 24 h. No special arrangements were made for the recovery of tryptophan (see Matsubara & Sasaki, 1969). After hydrolysis the ampoules were dried *in vacuo* over NaOH to remove the HCl and stored at -20°C until analysed.

6. Semi-quantitative amino acid and composition of peptides

A 10-50 nmol peptide aliquot was dried *in vacuo* over NaOH and conc. H_2SO_4 in a Durham tube (30 mm x 4 mm; A. Gallenkamp Ltd., London E.C.2), previously cleaned by acid washing or baking at 550°C for 12 h (Gray & Smith, 1970).

50 µl of constant boiling 5.7 M-HCl was added to the sample and the tube sealed under a partial vacuum. The tubes were heated for 18 h at 110°C and the acid was removed by drying *in vacuo* over NaOH at 60°C. The free amino acids were labelled by the dansyl method of Gray & Hartley (1963a), omitting the final hydrolysis step and the dansyl amino acids were identified by chromatography on polyamide sheets (see Materials and Methods III.7(a),(b)).

7. Peptide sequencing methods

Peptides were sequenced by the N-terminal dansyl-Edman procedure of Gray & Hartley (1963b) using 5-20% of the peptide starting material (20-500 nmol) to identify the N-terminal amino acid at each step of the Edman degradation. C-terminal sequences were determined from the results of variously timed digestions with carboxypeptidase-A. Amino acids thus liberated were identified as their dansyl derivatives using the dansyl procedure (Gray & Hartley, 1963a) without acid hydrolysis. Similarly, the free amino acid encountered following the final Edman degradation for a particular peptide, and those amino acids "carried over" from previous Edman degradation steps, were identified as their dansyl derivatives, employing the dansyl procedure without the final hydrolysis step.

(a) The dansyl method (Gray & Hartley, 1963a).

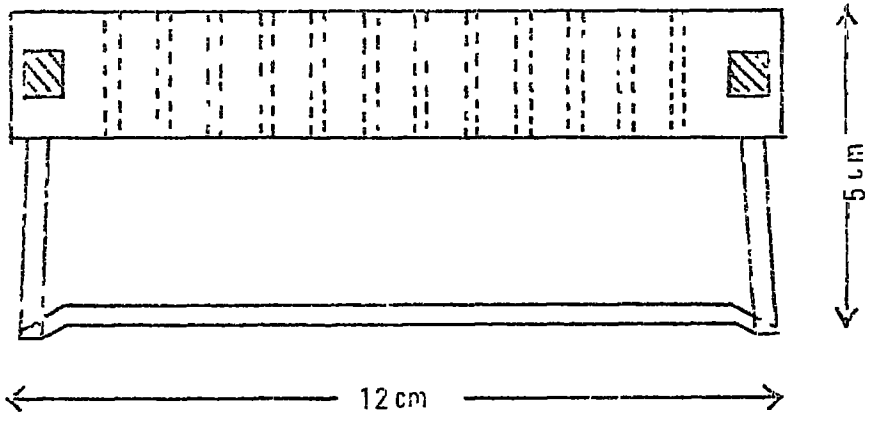
5-100 nmol of peptide sample was dried *in vacuo* over

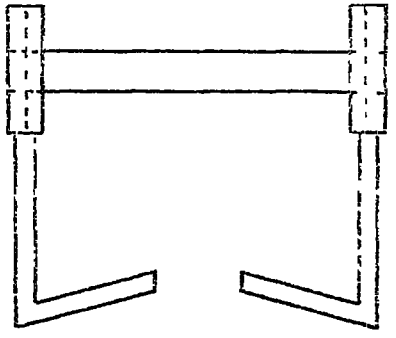
NaOH and conc. H_2SO_4 in a clean Durham tube (tubes were silylised by steeping them in 5% (v/v) dimethyldichlorosilane in toluene for one hour followed by extensive washing in dry methanol). The peptide residue was dissolved in 10 µl of 0.2 M-sodium bicarbonate and dried over NaOH and conc. H_2SO_4. Equal volumes of water and a solution containing 5 mg dansyl chloride per ml acetone were mixed and 10 µl of the solution was added to the dry sample. The tube was sealed with Parafilm and incubated at 45°C for 1 h, the reaction being terminated by drying *in vacuo* over NaOH and conc. H_2SO_4. Alternatively, the peptide sample was dried down in a Durham tube and mixed with 10 µl of a 1.1 mixture of a solution containing 5 mg dansyl-chloride per ml acetone and a solution of 0.1 M-N-ethylamine. The incubation and termination of the reaction was then carried out as above. After the dansylation step, 50 µl of constant boiling 5.7 M-HCl was added to the residue in the Durham tube and sealed under a partial vacuum. The tube was hydrolysed at 105-110°C for 4-18 h and the hydrolysate dried *in vacuo* over NaOH at 60°C.

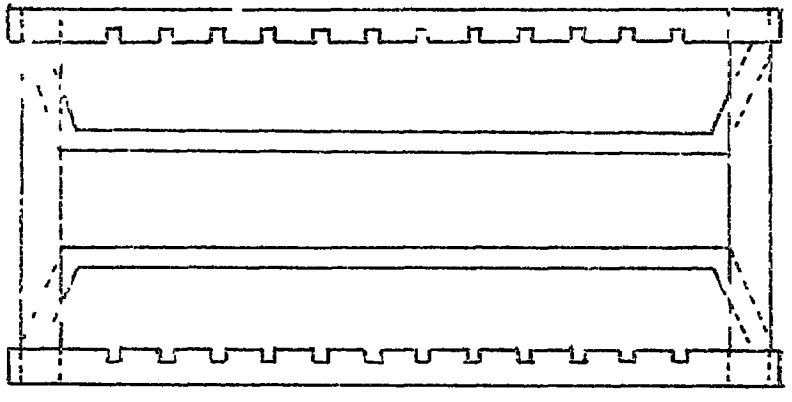
(b) Chromatography of dansyl derivatives.

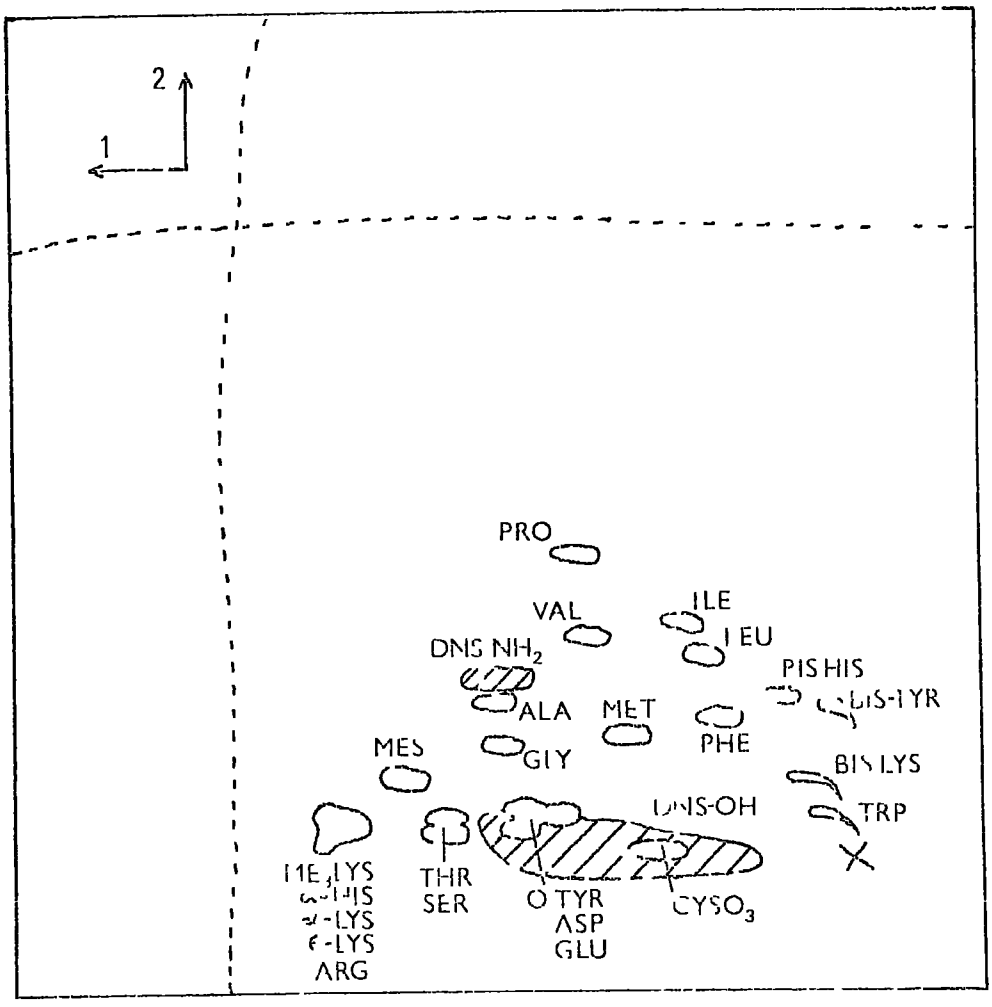
The chromatography of dansyl derivatives was carried out on two different sizes of polyamide sheets (Woods & Wang, 1967). 15 cm x 15 cm sheets were supported during chromatography in frames of the type described by Smith (1958), whilst the 5 cm x 5 cm sheets were supported in specially

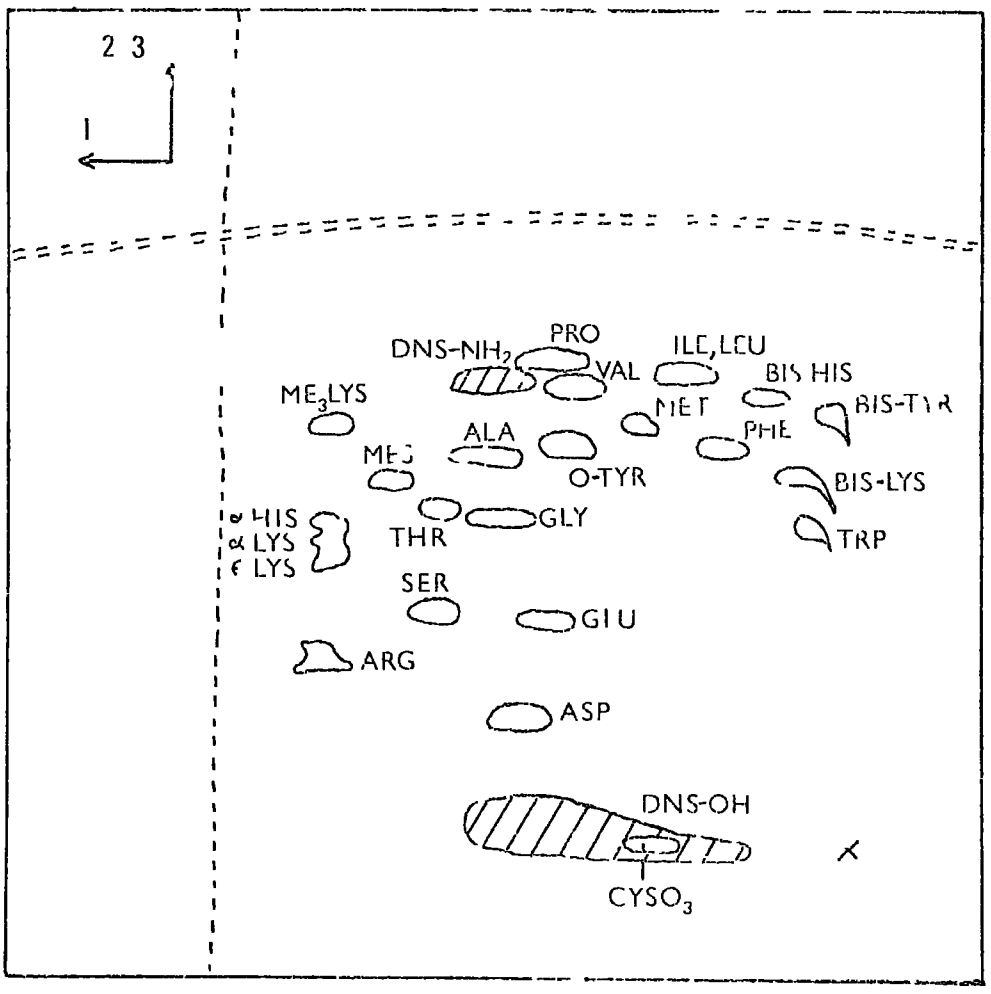
designed 12-seat racks (Figure 3). The dansylated sample was dissolved in 5 ml of 50% (v/v) pyridine and spotted onto an origin, common to both sides of the sheet, in a 4:1 ratio with the majority of the sample on the front. A 0.2-1 µl aliquot of the dansyl-amino acid standard solution (see Materials and Methods II.4(d)) was spotted onto the back of the sheet. Spotting was best carried out beneath a u.v. lamp and a hot draught, using an Eppendorf micro-pipette. When dry, the larger chromatograms were developed in two dimensions using first the solvent I for 50 min and secondly solvent II for 45 min, the sheets being dried in a hot draught after each run. The results of the two dimensional separation were recorded by illuminating the polyamide sheets under a u.v. lamp (350 nm) before the sheets were run in solvent III for 45 min in the second dimension. After drying in a hot draught, the sheets were examined for any further resolution of the dansyl spots (see Figures 4 and 5). The small sheets (5 cm x 5 cm) were chromatographed in a similar fashion for 10 min in each dimension, and it was found that repeating the run with solvent II before the use of solvent III improved the resolution of dansyl-derivatives. Co-chromatography of the samples on the front of the sheet and the standard mixture plus the sample trace on the back was used to identify the dansyl-derivatives. Polyamide sheets were washed after each experiment in an acetone-1 M-ammonia solution (1:1, by vol.)


5 cm
12 cm






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


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

for 1 h. The sheets could be re-used up to fifty times.

(c) The Edman degradation method.

150 µl of 5% (v/v) redistilled phenylisothiocyanate was added to 0.1-0.5 µmol of peptide dissolved in 150 µl of 20% (v/v) pyridine in an Edman tube (1 cm x 5 cm). The tube was flushed for 15 sec with oxygen-free nitrogen (Ilse & Edman, 1963), capped and incubated at 45°C for 1 h before excess reagent and volatile reaction by-products were removed by *in vacuo* drying over NaOH and P_2O_5 at 60°C. When the sample was completely dry, the tube was flushed again with nitrogen (Percy & Buchwald, 1972) and 200 µl of anhydrous trifluoro-acetic acid was added (Elmore & Toseland, 1956). The tube was sealed with Parafilm and incubated at 45°C for 30 min before excess reagent was removed by drying *in vacuo* over NaOH at 60°C. The degraded peptide was dissolved in 200 µl of distilled water and extracted once with 3 ml of butylacetate (Gray, 1967) and after decantation of the organic phase, the aqueous phase was dried *in vacuo* over NaOH and conc. H_2SO_4 at 60°C. The degraded peptide was dissolved in 150 µl 20% pyridine for further manipulation.

(d) Proteolytic digestion.

(i) Chymotrypsin and trypsin

Peptides were digested with 5% (w/w) α-chymotrypsin and 5% (w/w) trypsin dissolved in 0.2 M-NH_4HCO_3 buffer at pH 8.5 and 37°C for 30-60 min. The reaction was

terminated by freezing and lyophilisation.

Alternatively, the digestions were carried out in 0.2 M-N-ethyl morpholine-acetic acid buffer at pH 8.5.

(ii) Carboxypeptidase-A

Peptide samples were digested in Durham tubes with 0.2-0.5 mg carboxypeptidase-A/mol in 0.2 M-NH_4HCO_3 buffer at pH 8.4 and 37°C for times designed to maximize information. The reaction was terminated by drying *in vacuo* over NaOH. Alternatively, 5 µl of a carboxypeptidase-A suspension (80-120 µg) was washed three times in distilled water, suspended in 150 µl 0.2 M-$NaHCO_3$ and dissolved using 100-150 µl 0.1 M-NaOH, the solution being made neutral with 100-150 µl 0.1 M-HCl. The solution was made up to 1.5 ml with 0.2 M-N-ethylmorpholine-acetic acid buffer at pH 8.5 and 20 µl aliquots were used for each digestion.

(iii) Pyrrolidone carboxylyl peptidase

The G-200 enzyme 60% ammonia sulphate pellet was dissolved in 1 ml of 50 mM-sodium phosphate buffer, pH 7.3 containing 10 mM-2-mercaptoethanol and 1 mM-EDTA and dialysed against the same buffer (5 l) for 3 h at 2-4°C. The enzyme for a particular preparation (see Materials and Methods I.5(a), (b)) had a specific activity of 300-600 and an O.D.280 of about 2.0. Digestion was with 5 units of enzyme on 0.1-0.5µmol of peptide at 37°C for 3 h in a total volume of 500 µl 50 mM-sodium phosphate at pH 7.3. The reaction was terminated by

the addition of 2 ml of absolute ethanol and after standing at room temperature for 30 min, the protein precipitate was centrifuged at 1000 g for 10 min. The supernatant was retained and dried *in vacuo* over NaOH. The degraded peptide was then subjected to dansyl-Edman analysis.

(e) Removal of heme moiety from heme peptides prior to sequence analysis.

(i) Performic oxidation method (Nolan & Margoliash, 1966)

The heme peptide (0.1-0.5 μmol) was dissolved in 200 μl of 90% (v/v) formic acid at 0°C and 5 μl portions of 30% (w/v) hydrogen peroxide were added at hourly intervals to performic oxidise the thio-ether link. The reaction continued at 0°C until the red colour in the solution had been discharged. The reaction was terminated by freezing after dilution, and drying *in vacuo* over NaOH.

(ii) Mercuric chloride method (Ambler, 1963)

The heme peptide (0.1-0.5 μmol) was dissolved in 500 μl 5% (v/v) formic acid and 5 mg of mercuric chloride was added. The suspension was stirred at 37°C for 6 h before the reaction was terminated by the addition of 500 μl acetone. The solution was extracted five times with 2 ml portions of ether, made peroxide-free by shaking with a 5% (w/v) solution of ferrous sulphate. The aqueous phase was dried *in vacuo* and dissolved in 50 μl of 98-100% (v/v) formic acid to which was added 100 μl ice-cold performic acid preformed in the

manner of Hirs (1956) (0.5 ml 30% (v/v) H_2O_2 + 4.5 ml 88% (v/v) formic acid at 50°C for 3 min). The oxidation was continued for 1 h at 0°C before the solution was frozen after dilution and dried *in vacuo* over NaOH. The residue was extracted with 500 µl of 0.1 M-ammonia solution and the extract dried *in vacuo* before being dissolved in 150 µl of 20% (v/v) pyridine for sequence analysis.

8. Protein sequencing methods

(a) Manual sequencing methods.

(i) N-terminal analysis

Proteins were analysed by a modified method of Gray (1972). 25-100 µmol of protein was dried *in vacuo* over NaOH in a glass centrifuge tube and to this was added 50 µl of 1% (w/v) SDS solution. The tube was stoppered and placed in a heating block at 100°C for 5 min. When cool, 50 µl of N-ethyl morpholine was added, together with 75 µl of a solution of dansyl chloride in anhydrous dimethyl-formamide (25 mg/ml). The tube was stoppered and incubated at 45°C for 1 h before the labelled protein was precipitated by standing with 500 µl acetone for 40 min. The sample was centrifuged at 1000 *g* for 10 min and the pellet washed three times with 80% (v/v) acetone. After drying *in vacuo* over NaOH, the sample was hydrolysed for 18 h at 105-110°C with 5.7 M constant boiling HCl and the dansyl derivatives identified by polyamide sheet chromatography (see Materials and Methods

III.7(b)).

Alternatively, 25-100 nmol of protein was denatured by standing in 50 µl of 90% (v/v) ethanol for 4 h and then labelled, using the normal dansyl method (see Materials and Methods III.7(a)). After the hydrolysis step, the sample was dried *in vacuo* over NaOH at 60°C and transferred to an Edman tube (Pyrex screw-cap 1 cm x 5 cm) with 1 ml of dilute HCl at pH 3.5. The solution was extracted with 3 x 250 µl aliquots of diethyl ether and the ether extract, containing the N-terminal dansyl-amino acid, was dried down and identified by polyamide sheet chromatography. The N-terminal analysis of proteins was also carried out using a modified method of Gros & Labouesse (1969). 10-100 nmol of protein was dissolved in 200 µl of water to which was added 250 mg of ammonia-free urea and the volume was made up to 500 µl with water. 150 µl of 400 mM-sodium phosphate buffer at pH 8.2, 250 µl of dimethyl formamide and 100 µl of 0.2 M-dansyl chloride in acetonitrile were added and the solution incubated at 20°C for 60 min. The labelled protein was precipitated by the addition of chilled 10% (v/v) trichloroacetic acid, followed by centrifugation at 1200 *g* for 15 min. The precipitate was washed with 2M-hydrochloric acid and hydrolysed for 4-16 h at 110°C in 500 µl of 5.7 M constant boiling HCl. Dansyl derivatives were identified by polyamide sheet chromatography (see Materials and Methods III.7(b)).

(ii) C-terminal analysis

25-100 nmol of protein was denatured by heating at 100°C for 5 min in 50 µl of 1% (w/v) SDS solution. The protein was precipitated in 500 µl of acetone, washed three times with 80% (v/v) acetone, and dried *in vacuo* over NaOH. C-terminal sequences were determined from digestions with carboxypeptidase-A using the peptide method (see Materials and Methods III.7(d)(ii)).

(iii) Edman degradation

50-150 nmol of protein was degraded sequentially by the Edman procedure (Gray & Hartley, 1963b) following denaturation by the method of Gray (1972). Amino acids were identified as their dansyl-amino acids using polyamide sheet chromatography.

Alternatively, a modified method of Weiner *et al.* (1972) was applied. 50 nmol of cytochrome *c* was dissolved in 200 µl of 0.5 M-$NaHCO_3$ and adjusted to pH 9.8 with 2 M-NaOH. 75 µl of 10% (w/v) SDS was added, together with 10 µl of redistilled phenylisothiocyanate. The solution was incubated at 45°C for 1 h under nitrogen with mixing at 15 min intervals. 2.5 ml of acetone was added and the precipitate centrifuged at 1000 *g* for 10 min and washed with a further 2 ml of acetone. The protein precipitated was spread around the wall of the reaction tube and dried *in vacuo* over NaOH at 60°C. 200 µl of anhydrous trifluoroacetic acid was added and the sample was

incubated under nitrogen for 30 min. The trifluoroacetic acid was then removed by *in vacuo* drying over NaOH at 60°C. The residue was redissolved in the coupling buffer, together with 75 µl of 10% (w/v) SDS at 50°C. A 10% aliquot was taken for dansyl-analysis and the cycle was then repeated. Half a volume of a 5 mg/ml solution of dansyl-chloride in acetone was added to the 10% aliquot in a Durham tube and the solution was incubated at 45°C for 30 min. Precipitation and hydrolysis of the labelled protein was by the method of Gros & Labouesse (1972) (see Materials and Methods III.8(a)(i)). Dansyl-derivatives were identified by polyamide sheet chromatography.

(b) Automatic sequencing methods.

A Beckman 890c sequencer operating on a quadrol fast protein double cleavage program was used as recommended in the Beckman (1972) operation manual. The amino acid released at each degradation step was collected as the phenylthiazolinone derivative in a refrigerated fraction collector, and blown to dryness under a stream of nitrogen. This was converted to the more stable phenylthiohydantoin (PTH) derivative by reacting the dried fraction with 0.2 ml of 1 M-hydrochlorine acid containing ethanethiol (1 µl/ml) for 10 min at 80°C under nitrogen. The PTH derivatives were then extracted twice using 0.7 ml portions of peroxide-free ethylacetate. Both the organic phase, which would contain the majority of the PTH-amino acids, and the aqueous phase, which would contain the

PTH-derivatives of arginine, histidine and cysteic acid, were dried down and retained.

PTH amino acid derivatives were identified by gas chromatography on 10% SP400 AW Chromosorb W (100-200 mesh) using a Varian Aerograph 1400 gas chromatograph (Beckman Manual, 1972; Pisano et al., 1972). They were also identified by thin-layer chromatography on silica plates incorporating a fluorescent indicator and located by their fluorescent quenching under 260 nm light (Jeppsson & Sjöquist, 1967). Alternatively, the hydriodic acid procedure of Inglis et al. (1971) was used to regenerate the parent amino acid from the PTH derivative which was then identified by conversion to the dansyl derivative and subsequent chromatography on polyamide sheets (see Materials and Methods III.7(a),(b)).

9. Acetyl group determination

Acetyl groups were determined as their 1-acetyl-2-dansyl hydrazine derivatives (Schmer & Kreil, 1969). The peptide sample (50-150 nmol) was dried *in vacuo* in a Durham tube before 25 µl of 0.1 M-HCl was added and redried *in vacuo* over NaOH at 60°C. 20 µl 95% + (v/v) hydrazine was added and the tube sealed to give a partial vacuum before heating at 105°C for 18 h. The sample was dried *in vacuo* over NaOH at 60°C before being dissolved in 0.2 M-sodium citrate buffer at pH 3.0. 5 µl of a solution containing 5 mg dansyl-chloride per ml acetone was added, the tube capped with Parafilm and

incubated at 45°C for 18 h. The sample was dried and compared during chromatography on polyamide sheets with standard 1-acetyl-2-dansyl hydrazine derivatives.

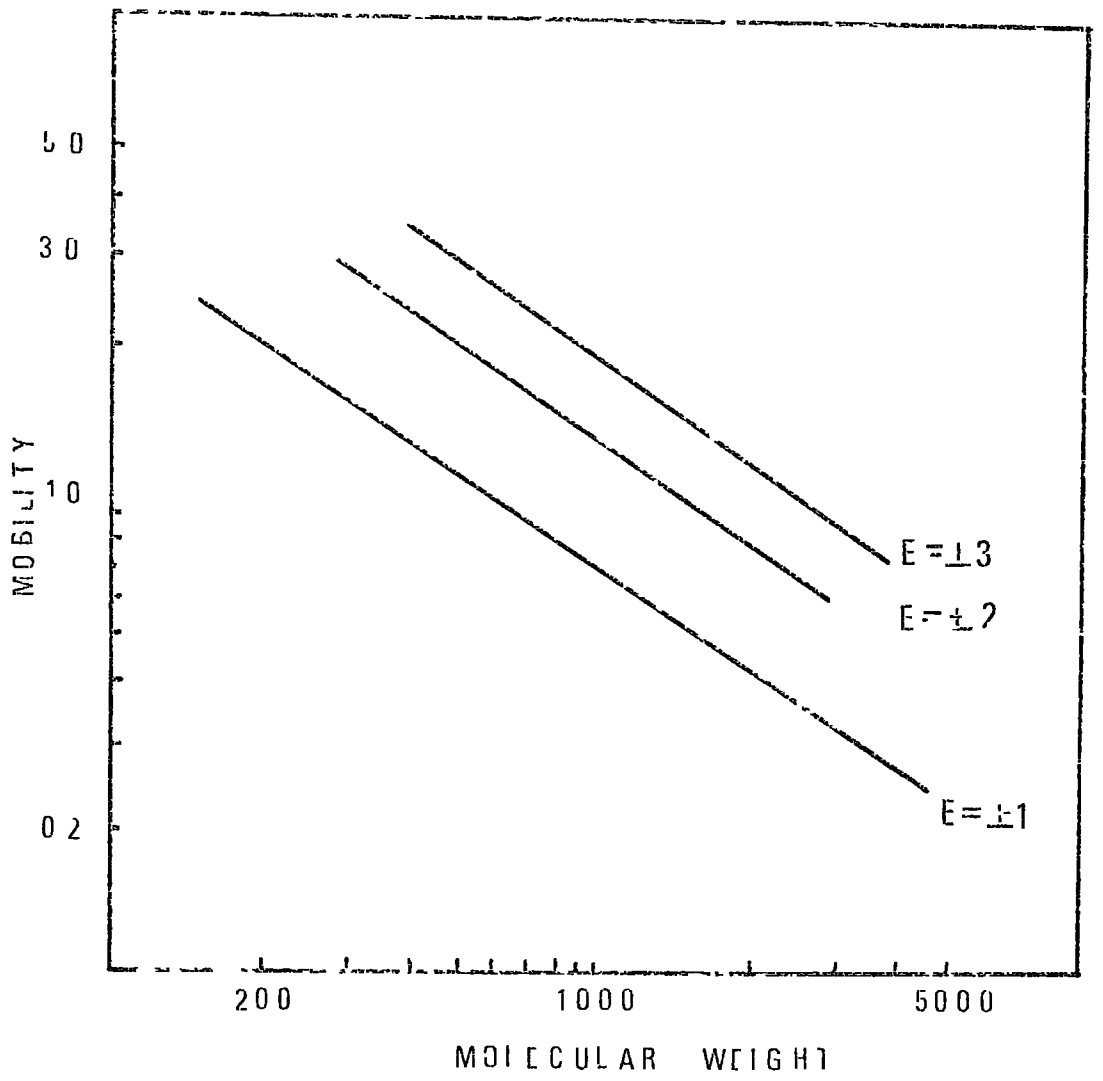
10. Amide determinations

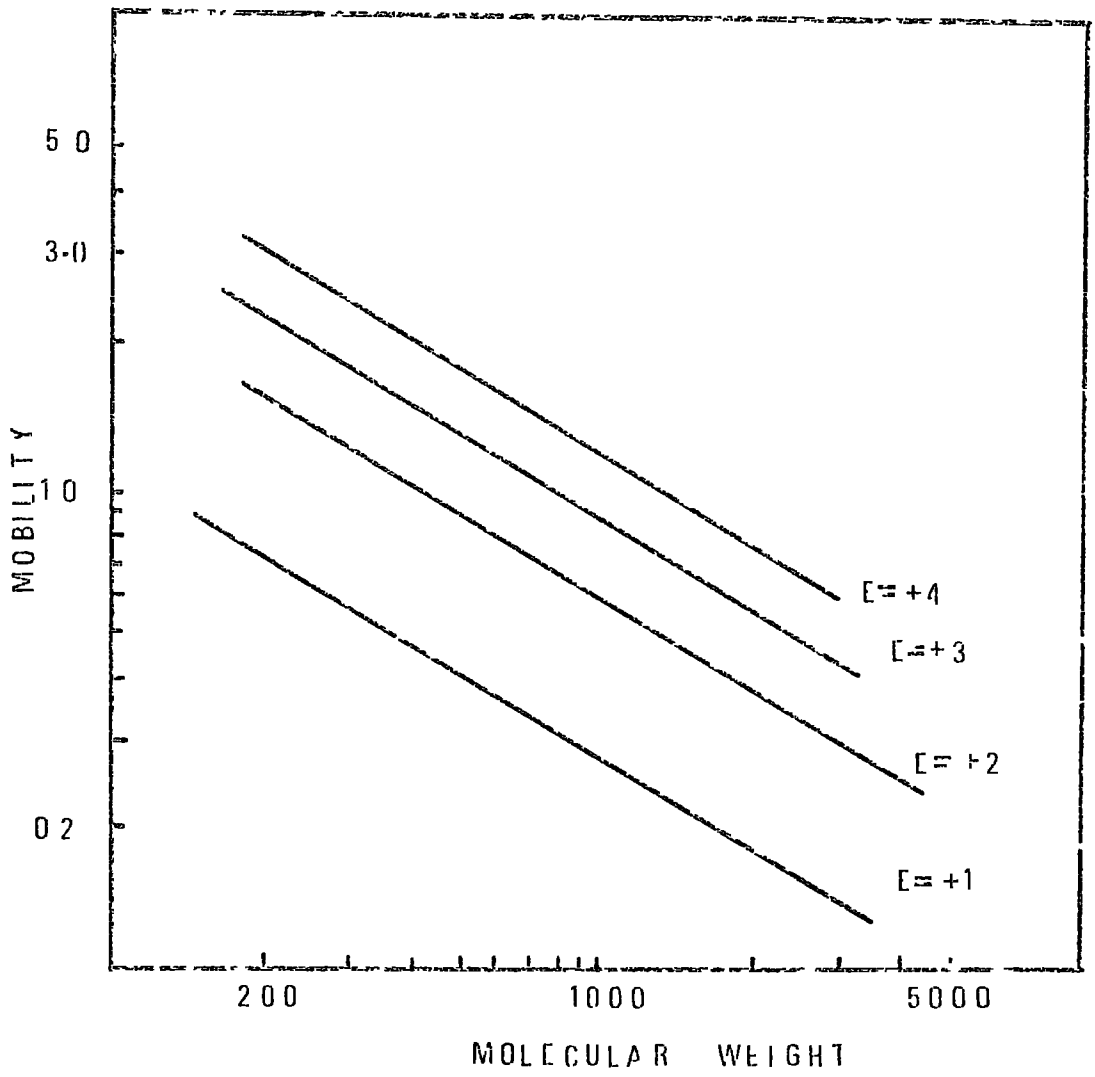
Amide residues were determined from sequence data and pH 6.5 electrophoretic mobilities using the method of Offord (1966). A graph was constructed from peptide data giving peptide charges from mobilities expressed relative to the dansyl markers used (see Figures 6 and 7).

Alternatively, amide residues were determined from the pH 6.5 mobilities of the intact peptide and of samples of the peptide taken at the appropriate stages of degradation. In this case, allowances were made for the partial or complete blockage of the ε-amino functions of lysine by phenylthiocarbamyl groups due to exposure to PITC during the Edman degradation steps.

11. Nomenclature used to describe sequence analysis data

Peptides are numbered on the basis of their occurrence in the complete sequence starting at the N-terminus of the protein. Peptides prefixed by C refer to the chymotryptic peptides and those by T to the tryptic peptides. Peptides resulting from partial cleavage within a major peptide are given a letter subscript to the major peptide. Peptides derived by further cleavage are given a subscript to the parent peptide


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


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

with regard to the nature of the cleavage, as above, and their order in the parent peptide. All residue numbering is given in the appropriate figures. Arrows (⟶) indicate positions confirmed by dansyl-Edman analysis, and arrows (⟵) indicate positions confirmed by carboxypeptidase-A digestion followed by dansyl-analysis. Residues underlined (X) are those confirmed as peptide C-terminals by dansylation without hydrolysis following the final Edman degradation step. Residues shown in brackets are placed from evidence other than dansyl-Edman analysis.

IV. CALCULATIONS BASED ON AMINO ACID SEQUENCE DATA

1. Amino acid difference matrix construction

Sequence alignments were made relative to the cysteinyl residues to which the heme group is attached. Differences due to deletions and blanks due to non-determinations were normally considered as single changes.

2. Phylogenetic tree construction; ancestral sequence method

Phylogenetic trees were constructed by an ancestral sequence method based on that of Dayhoff & Eck (1966) and Dayhoff (1969; 1972), using a program written by A. A. Young. Computing Laboratory, University of Durham, in PL1. Jobs were executed on the Cambridge IBM 370 computer system.

The computing strategy was divided into three main procedures (Boulter et al., 1972; Dayhoff, 1972). The construction commences with any three sequences which can only be related by a single tree. A fourth sequence was then added to the tree at all possible points, and the best topology was selected for further tree construction. Additional sequences were added successively to the best tree obtained at the previous step, until a final tree is obtained.

The second computer procedure evaluated the topologies obtained during the building and shuffling procedures. The "ancestral sequence" was inferred for each node and each position along the sequences was considered in turn, three

lists being made of the residues found in the position along the three branches associated with that node. If, for a given node and a given position, only one residue was found occurring on more branch lists than any other, then it was selected as the nodal amino acid. Less clear-cut nodal positions were left blank. When all the nodes were assigned an amino acid or a blank for every position in the sequence, the situations at the blank positions were reassessed. Thus, if a blank had at least two of its three adjacent neighbours (either node or sequence) the same, then this was assigned to that nodal position. This process was repeated until no further additions could be made, whereupon the nodal sequences were checked so that an amino acid at the node not identical to at least two of its adjacent neighbours, was changed to a blank. This process gave a definite assignment of the ancestral sequence residues whenever one choice was clearly preferable, but left blanks where reasonable doubt existed. The tree was then evaluated by counting the number of amino acid changes along every branch of the tree at each position in the sequences to give a total relating the whole tree. In terms of the overall evaluation, the choice of amino acid for a position left blank in the assignment of ancestral sequences was immaterial because, unless two parallel changes existed at adjacent nodes, the total number of changes relating the final phylogeny remained constant. When the branch-lengths were evaluated

for the final tree, the minimum number of changes counted around a blank, or a series of blanks, was divided equally among all the independent branches.

The third procedure evaluated the branches in alternative positions, branch by branch. This was necessary, because once fixed to a tree during the building strategy, a sequence would not change its relative position. No account was made of the remaining sequences still to be added, when each new sequence was fixed during building, so that the final tree may have been only a close approximation to the "best" possible tree relating the data. The procedure evaluated the position of all the branches selecting better alternatives as they arose.

3. Phylogenetic tree construction; numerical matrix method

Phylogenetic trees were constructed using a numerical matrix method of Fitch & Margoliash (1967a) modified by the "additive hypothesis" of Moore *et al.* (1973).

RESULTS

1. The Purification of Locusta cytochrome c

3 kg of frozen locusts were homogenised with ice in 10 mM-aluminium sulphate solution and stirred for 2 h at pH 4.5 and 4°C. The homogenate filtered easily using a 21" basket centrifuge and no significant precipitation was observed after raising the pH to 8.0 using 2 M-Tris. The 40 l of filtrate was passed through a sintered glass funnel (18 cm x 10 cm) filled with Amberlite CG-50 resin regenerated in the NH_4^+ form. Application of a reduced pressure, using a water-pump, gave a flow rate of 15 l/h. The resin capacity was judged to be 20 l. The filtrate from this step was then passed overnight through 6 cm x 10 cm columns of Amberlite CG-50 resin before being discarded. Both batches of resin were washed with distilled water and eluted batchwise with 2 M-NaCl maintained at pH 8.0 with 2 M-NaOH. Cytochrome c was assayed for in the eluate using a direct vision hand spectroscope, following reduction with a trace of ascorbic acid. A small quantity of cytochrome c was observed in the column eluate and was combined with the main fraction to give a total volume of 8 l of partially purified cytochrome c. This was dialysed against running tap water for 8 h and then against 60 l of 50 mM-sodium phosphate buffer at pH 8.0 and 4°C for a further 8 h. No filtration step was required before the concentration of the cytochrome on a 6 cm x 10 cm

column of CM-50 Sephadex equilibrated in 50 mM-sodium phosphate buffer at pH 8.0. After loading, the column was washed with a litre of buffer before elution of the cytochrome c using 500 mM-NaCl in 50 mM-sodium phosphate buffer at pH 8.0. 500 ml of eluate was dialysed for 8 h against 60 l of 50 mM-sodium phosphate buffer at pH 8.0 before further concentration using a 2 cm x 5 cm column of CM-50 Sephadex. The protein was eluted as before and represented a 50-fold concentration of the partially purified cytochrome c. The ammonium sulphate fractionation was carried out at pH 8.0 and 4°C. No precipitation of the cytochrome c was observed at saturations up to 80%, at which point the fractionation was terminated, but quantities of a white precipitate were discarded following centrifugation. The cytochrome c in 80 ml of 80% saturated ammonium sulphate was diluted twice and exhaustively dialysed against changes of 10 mM-sodium phosphate buffer at pH 7.2 before concentration on a 1 cm x 5 cm column of CM-52 cellulose and elution in a volume of 4 ml using 200 mM-sodium phosphate buffer at pH 7.2. The cytochrome c, which had a $E^{C3+}_{410}/E^{C3+}_{280}$ spectral ratio of 2.6, was chromatographed on Biogel P.30 (Figure 8) and fractions having $E^{C3+}_{410}/E^{C3+}_{280}$ values greater than 3.0 were pooled. The eluate was pumped through a 1 cm x 10 cm column of CM-52 cellulose equilibrated in 10 mM-sodium phosphate buffer at pH 7.2. The cytochrome c adsorbed to the top of the resin bed and was eluted using a linear ionic

gradient (Figure 9). The eluate was split into parts A and B. Pool A had a $E^{C3+}_{410}/E^{C3+}_{280}$ value of 3.4 and was presumed to contain deamidated fractions together with a certain amount of native cytochrome c as a result of "column overload". The pool was diluted appropriately and re-run using a 1 cm x 30 cm column of CM-52 cellulose and an identical loading and elution scheme to that outlined above (Figure 10). Pool B had an $E^{C3+}_{410}/E^{C3+}_{280}$ value of 4.0 and was considered to be pure. Samples of Locusta cytochrome c were desalted by passage through a 1 cm x 15 cm column of Amberlite MB-1 resin equilibrated and eluted in distilled water. Desalted samples were lyophilised and stored at -20°C. Details of the yields and $E^{C3+}_{410}/E^{C3+}_{280}$ purity ratios of the cytochrome c during the various stages of purification are shown in Table 1 and the final spectral ratios of the pure cytochrome c are given in Table 2. The elution profiles of the chromatographic fractionations are shown in Figures 8-10.

2. Purification of Macrobrachium cytochrome c

The cytochrome c from Macrobrachium was purified from two batches of starting material.

Batch I

13 kg of frozen abdomens of Macrobrachium were rapidly thawed at 4°C and homogenated with ice in 10 mM-aluminium sulphate. The homogenate was adjusted to pH 4.5 and stirred for 2 h at 4°C. Filtration using a 21" basket centrifuge

TABLE 1.

The Purification of Locusta Cytochrome c

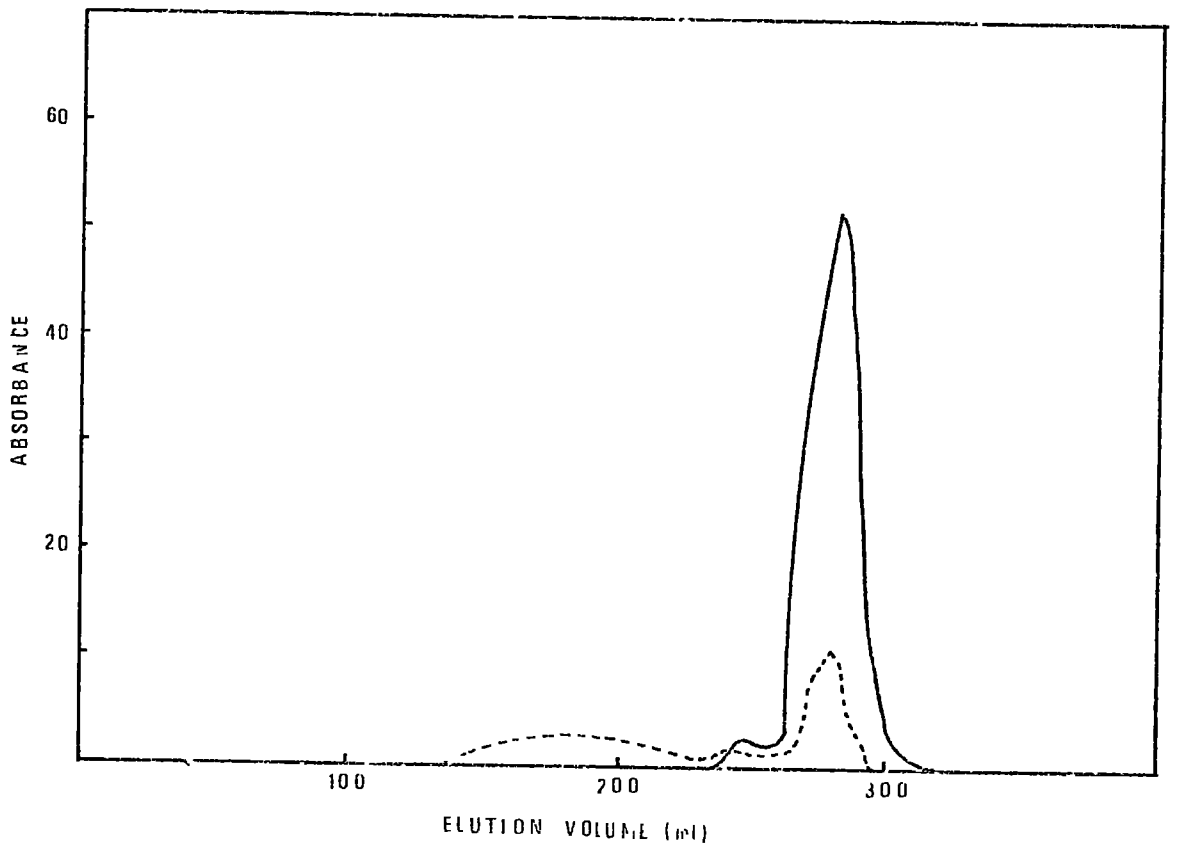
Step	Yield mg cytochrome/ kg starting material	Purity Ratio $E^{C3+}_{410}/E^{C3+}_{280}$
Amberlite CG-50	-	0.6
CM-50 Sephadex	41	1.2
Ammonium sulphate fractionation	36	2.6
Biogel P-30	26	3.7
CM-52 Cellulose. Ionic Gradient	20	4.0
Amberlite MB-1	20	4.0

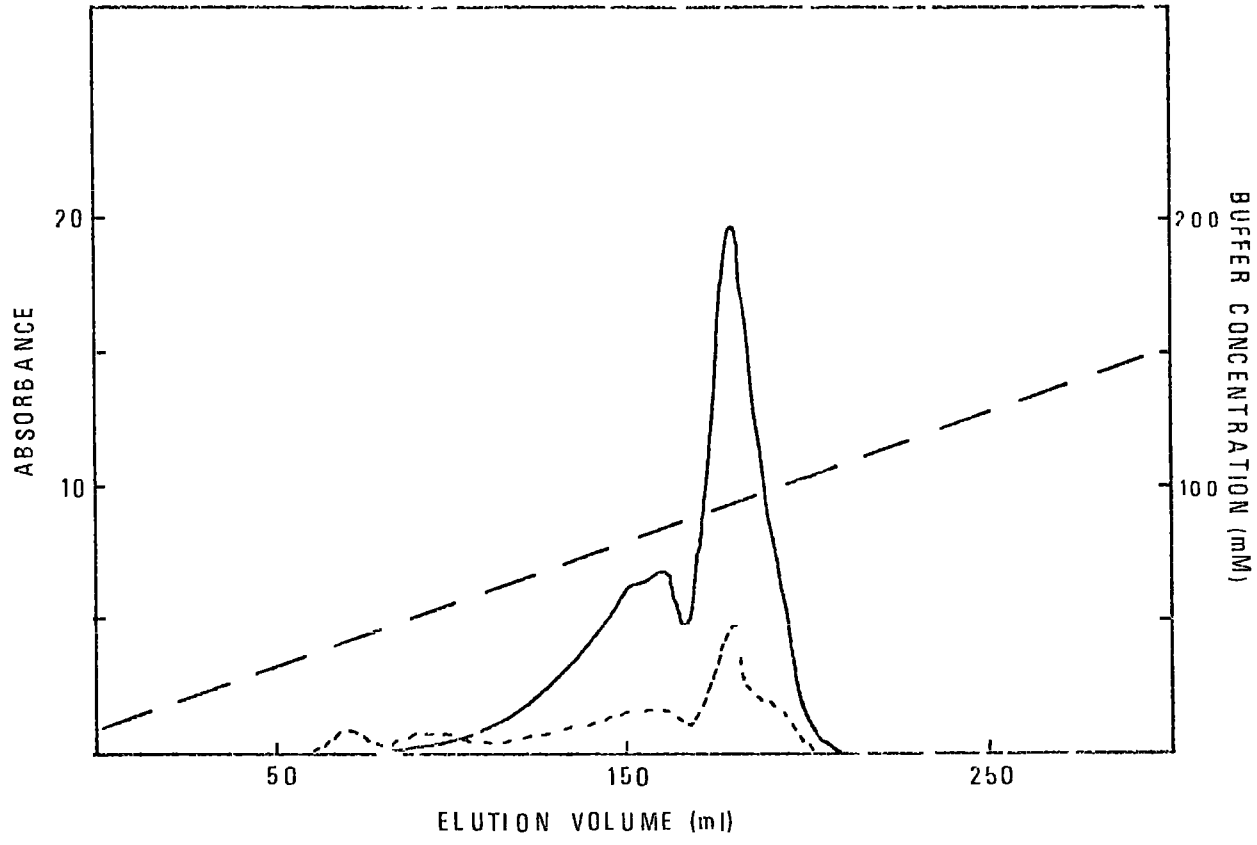
TABLE 2.

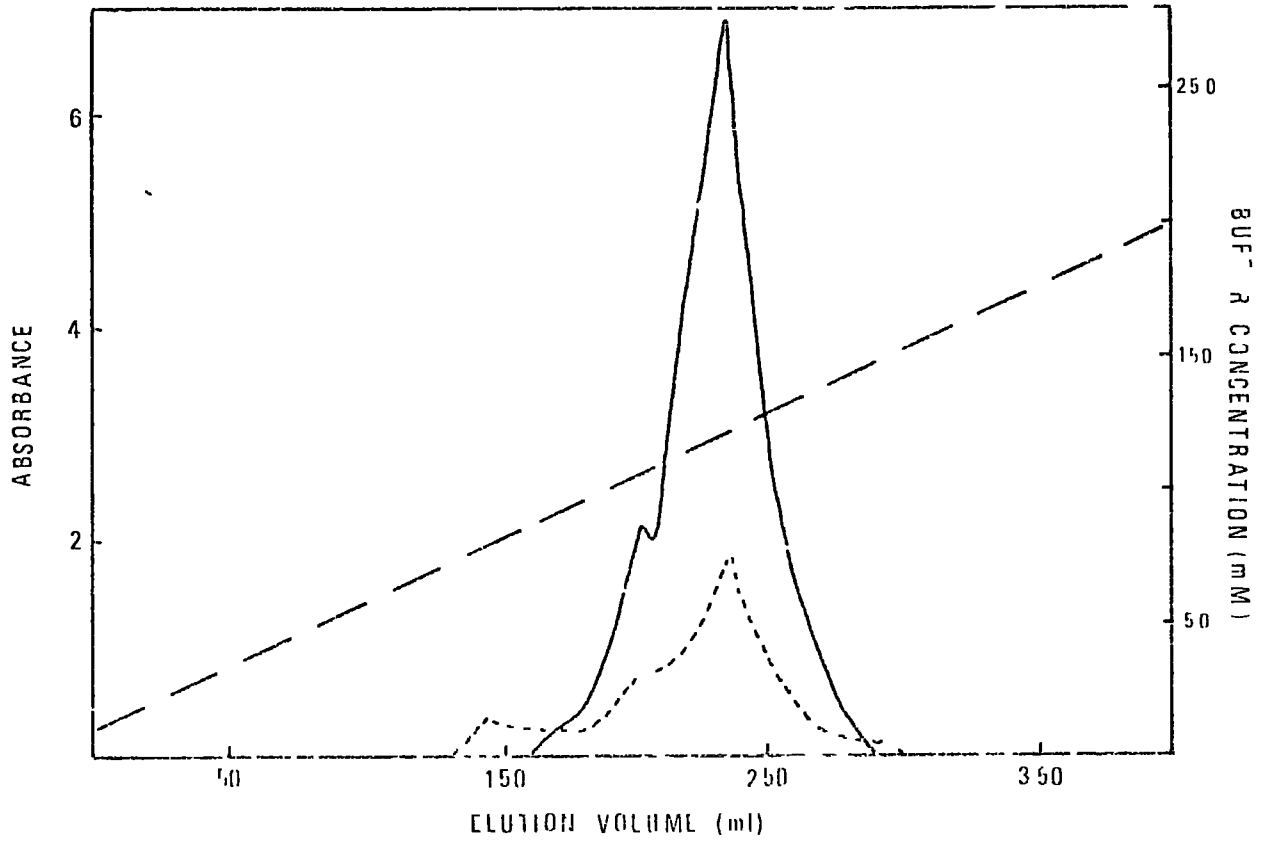
The Spectral Ratios of Purified *Locusta* Cytochrome *c*

	Ratio
$E^{C2+}_{550}/E^{C3+}_{280}$	0.9
$E^{C3+}_{410}/E^{C3+}_{280}$	3.9
$E^{C2+}_{418}/E^{C3+}_{410}$	1.2
$E^{C2+}_{418}/E^{C2+}_{550}$	5.3

The final mass was 60.5 mg salt-free


60
40
20
ABSORBANCE
100
200
300
ELUTION VOLUME (ml)


ABSORBANCE
20
10
50
150
250
ELUTION VOLUME (ml)
200
100
BUFFER CONCENTRATION (mM)


ABSORBANCE
6
4
2
50
150
250
350
ELUTION VOLUME (ml)
250
150
50

was slow because of the slimy nature of the extract, but the pH 4.5 filtrate was very clear. Raising the pH to 8.0 with 2 M-Tris caused the formation of a white precipitate which could only be partially removed by filtration on Büchner funnels using Whatman 54 filter paper. 40 l of filtrate was passed through a sintered glass funnel (18 cm x 10 cm) containing Amberlite CG-50 resin regenerated in the NH_4^+ form. The flow rate was 15 l/h and the resin capacity was fixed at 20 l. The filtrate from this step was then passed overnight through 6 cm x 10 cm Amberlite CG-50 columns before being discarded. Both batches of resin were washed and eluted by the method described for the Locusta preparation. Cytochrome c was assayed for in the eluate using a direct vision hand spectroscope following reduction with a trace amount of ascorbic acid. Significant amounts were observed in the funnel eluate, but none in that from the columns, and this was discarded. Six litres of eluate was dialysed under the conditions described for the Locusta preparation. The lowering of the ionic concentration as a result of dialysis caused the formation of a heavy white precipitate, and this was removed by centrifugation at 1500 g for 30 min. The supernatant was concentrated on a 6 cm x 10 cm column of CM-50 Sephadex equilibrated in 50 mM-sodium phosphate buffer at pH 8.0. A white precipitate collected on the top of the resin bed and interfered with the elution step. The volume

of eluate was 650 ml and this was concentrated using a Diaflo ultra-filtration unit equipped with a UM-10 membrane operating at 70 p.s.i. and a flow rate of 30 ml/h at 4°C. The concentrated cytochrome c in 20 ml of solution was taken in 10% stages to 80% saturation with ammonium sulphate at pH 8.0 and 4°C, but no precipitation of cytochrome was observed, although large quantities of a copious white precipitate were removed by centrifugation at saturations greater than 40%, and these were discarded. The cytochrome c in 70 ml of 80% saturated ammonium sulphate was diluted twice and exhaustively dialysed against changes of 10 mM-sodium phosphate buffer at pH 7.2. 200 ml of dialysis residue was concentrated on a short column of CM-52 cellulose as described for the Locusta preparation and further purified by chromatography on Biogel P-30 (Figure 11). Partial separation into oxidised and reduced forms was observed and a yellow band was seen trailing the cytochrome bands, but this was not distinguished on spectrophotometric analysis of the eluted fractions. Those fractions having an $E^{C3+}_{410}/E^{C3+}_{280}$ spectral purity ratio of greater than 1.4 were pooled and the cytochrome c adsorbed to the top of a 1 cm x 10 cm column of CM-52 cellulose equilibrated in 10 mM-sodium phosphate buffer pH 7.2. The sample was eluted using a linear ionic gradient (Figure 12) and this step led to the collection of 15 mg of cytochrome c with an $E^{C3+}_{410}/E^{C3+}_{280}$ value of 4.1. Samples of cytochrome c were

concentrated on a 1 cm x 2 cm column of CM-52 cellulose equilibrated in 10 mM-sodium phosphate and eluted with 200 mM-sodium phosphate at pH 7.2. Samples were desalted on a 1 cm x 10 cm column of Sephadex G-10, freeze-dried and stored at -20°C. Details of the yields and $E^{C3+}_{410}/E^{C3+}_{280}$ purity ratios of Macrobrachium cytochrome c during the various stages of the purification of Batch I are shown in Table 3, and the final spectral ratios obtained for the pure protein are given in Table 4. The elution profiles of the chromatographic fractionation steps are shown in Figures 11 and 12.

Batch II

The second extraction was from 26 kg of frozen Macrobrachium abdomens, and this gave 120 l of homogenate which was treated in a similar manner to Batch I. At the Amberlite stage, the pressure eluate contained cytochrome c but none was seen in the column eluate, and this was discarded. Concentration after dialysis was on CM-50 Sephadex, the dilution of which was complicated by a thick white precipitate on the resin bed. The 300 ml of eluate was concentrated, after a dialysis step, on a 2 cm x 6 cm column of CM-52 cellulose equilibrated in 10 mM-sodium phosphate, pH 7.2, and eluted in a volume of 15 ml with 200 mM-sodium phosphate buffer. Large quantities of a thick white precipitate were removed by the ammonium sulphate fractionation, but precipitation of cytochrome c was not observed at saturations up to 80%. The dialysis and gel

TABLE 3.

The Purification of *Macrobrachium* Cytochrome *c* (Batch I)

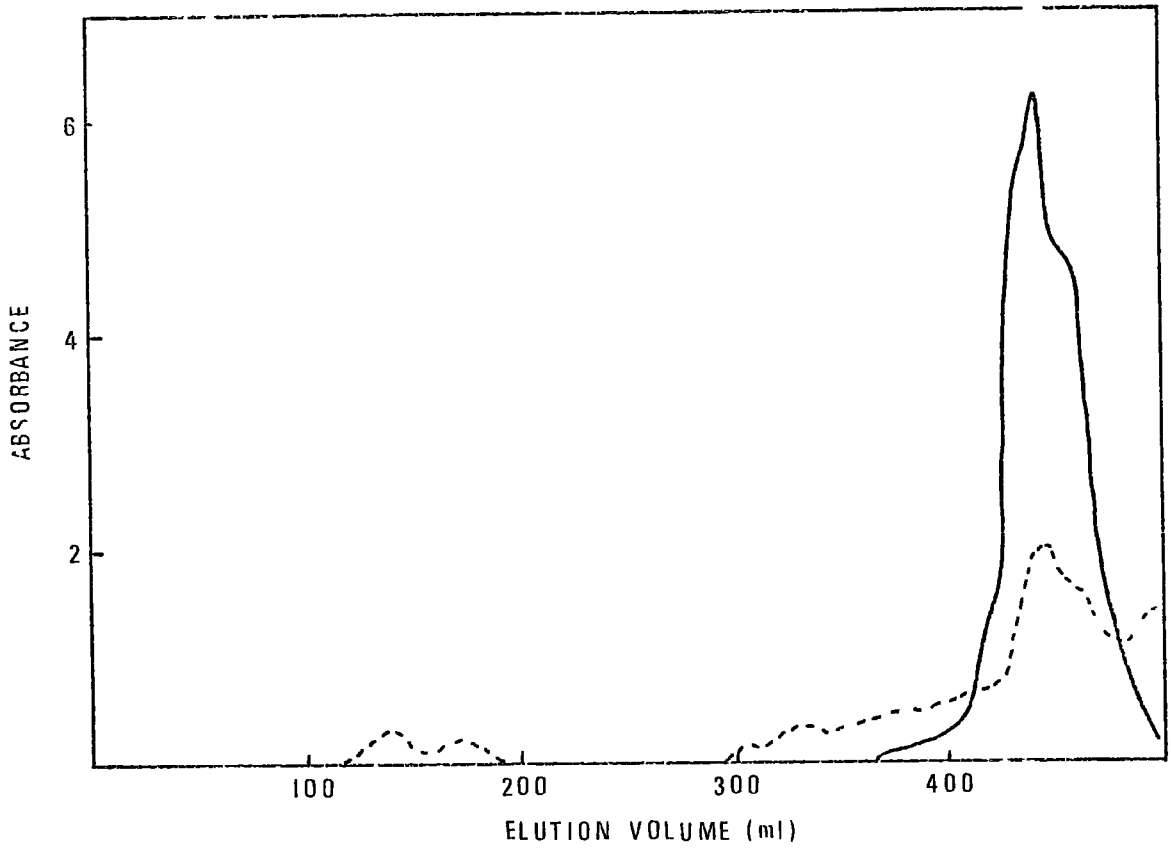
Step	Yield: mg cytochrome/ kg starting material	Purity Ratio: $E^{C3+}_{410}/E^{C3+}_{280}$
Amberlite CG-50 / CM-50 Sephadex	1.9-2.1	0.4
Ammonium Sulphate Fractionation	1.8	0.6
Biogel P-30	1.6	2.6
CM-52 Cellulose, Ionic Gradient	1.1	4.1
G-10 Sephadex	1.0	4.1

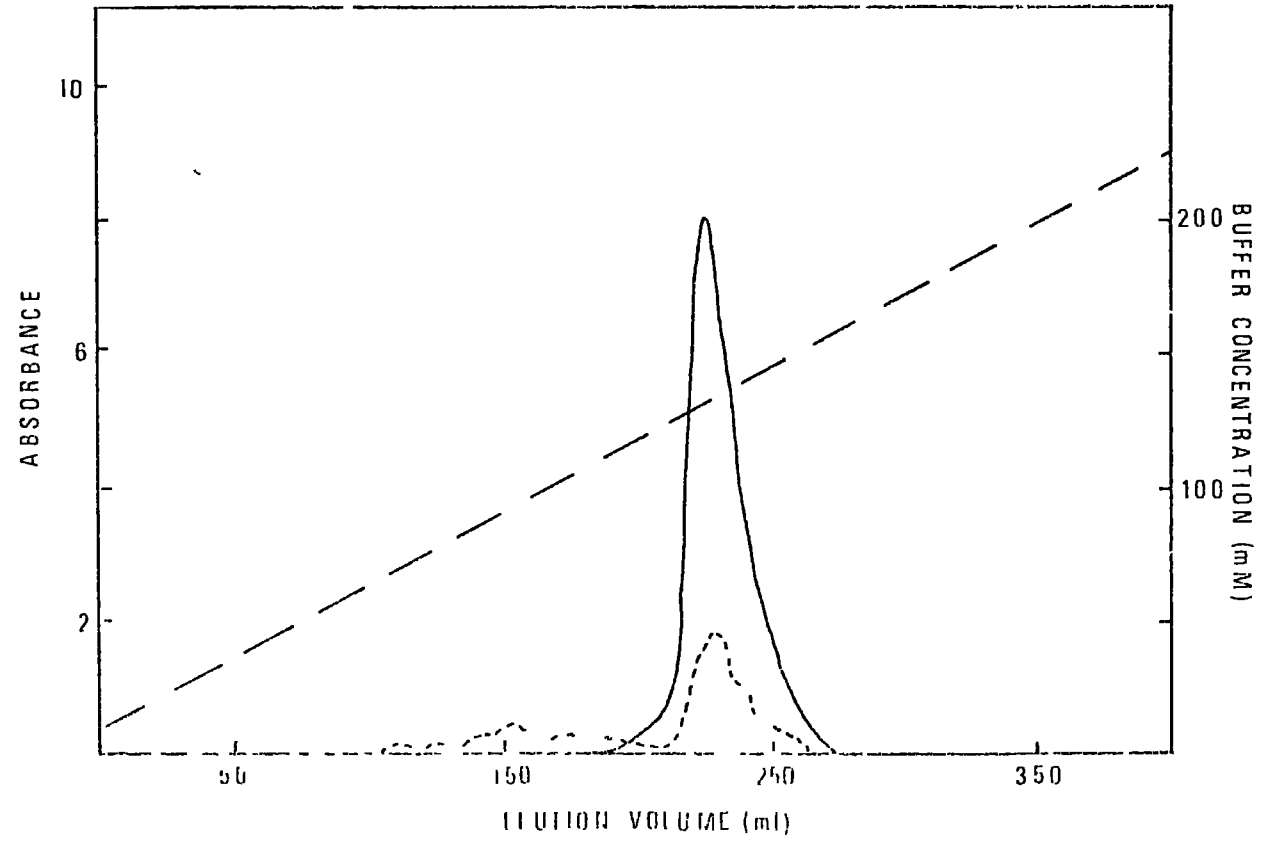
TABLE 4.

The Spectral Ratios of Purified *Macrobrachium* Cytochrome *c* (Batch I)

	Ratio
$E_{550}^{C2+}/E_{280}^{C3+}$	1.0
$E_{410}^{C3+}/E_{280}^{C3+}$	4.1
$E_{416}^{C2+}/E_{410}^{C3+}$	1.2
$E_{416}^{C2+}/E_{550}^{C2+}$	4.6

The mass was 13.5 mg salt-free.


ABSORBANCE
6
4
2
100
200
300
400
ELUTION VOLUME (ml)


ABSORBANCE
10
6
2
200
100
BUFFER CONCENTRATION (mM)
50
150
250
350
ELUTION VOLUME (ml)

filtration steps were carried out as described for the Locusta preparation to give pooled fractions with an $E_{410}^{C3+}/E_{280}^{C3+}$ ratio of 2.6 (Figure 13). Ion-exchange chromatography on CM-52 cellulose, using a linear ionic gradient from 10 mM- to 300 mM-phosphate at pH 7.2 on a 1 cm x 10 cm column only improved this purity ratio to 3.6 (Figure 14), and a gel filtration step using Sephadex G-75 was employed to further purify the protein. A 3 cm x 100 cm column of Sephadex G-75 equilibrated and eluted in 75 mM-Tris-KCl buffer at pH 7.5 was used. The flow rate was 30 ml/h in an upward direction and the pool of fractions selected from the elution profile gave an $E_{410}^{C3+}/E_{280}^{C3+}$ value of 4.3. Samples were desalted using a 1 cm x 10 cm of Sephadex G-10, freeze-dried and stored at $-20^{o}C$.

Details of the yields and $E_{410}^{C3+}/E_{280}^{C3+}$ purity ratios of Macrobrachium cytochrome c during the various stages of purification of Batch II are shown in Table 5 and the final spectral ratios obtained for the pure protein are given in Table 6. The elution profiles of the chromatographic fractionation steps are shown in Figures 13 and 14.

3. Purification of Eisenia cytochrome c

The cytochrome c from Eisenia was purified from three batches of starting material.

TABLE 5.

The Purification of *Macrobrachium* Cytochrome *c* (Batch II)

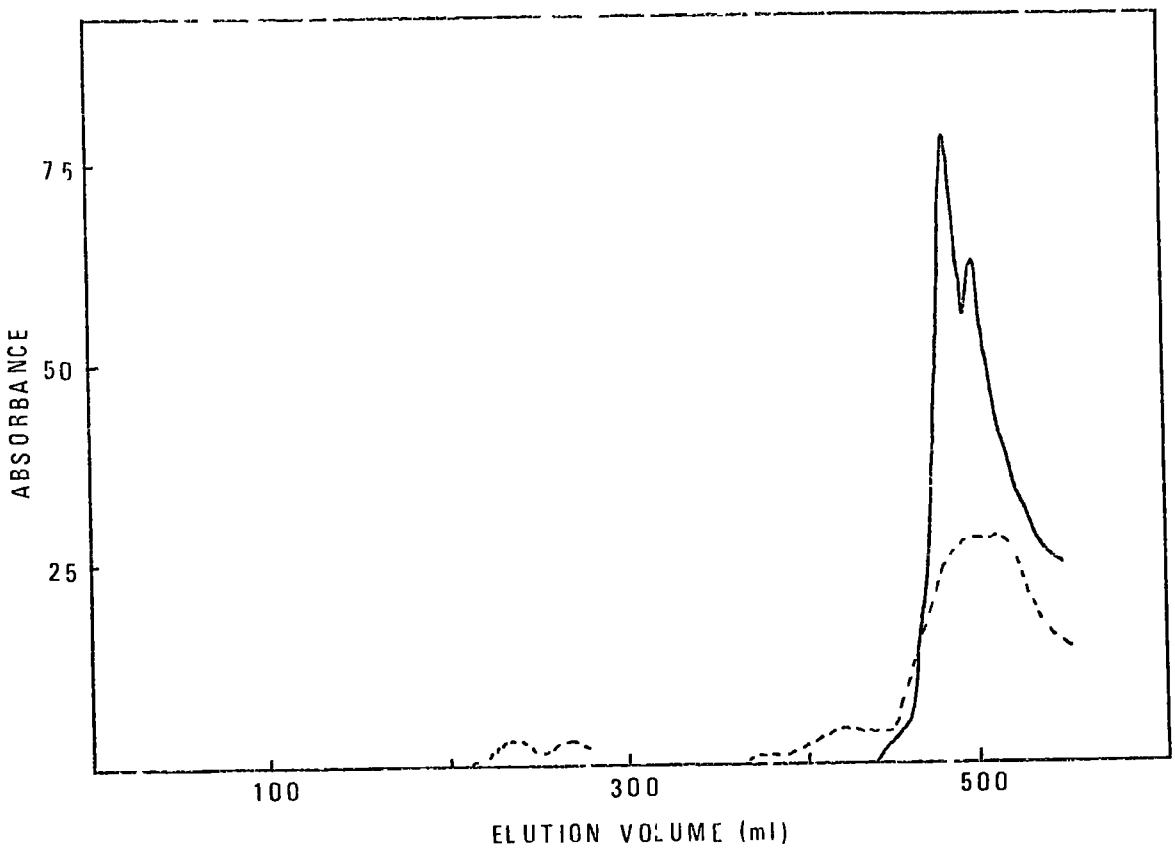
Step	Yield mg cytochrome/ kg starting material	Purity Ratio $E^{C3+}_{410}/E^{C3+}_{280}$
Amberlite CG-50	2.7	0.8
CM-50 Sephadex	2.2	1.1
Ammonium Sulphate Fractionation	2.0	1.6
Biogel P-30	1.2	2.6
CM-52 Cellulose, Ionic Gradient	0.8	3.6
G-75 Sephadex	0.6	4.3
G-10 Sephadex	0.5	4.1

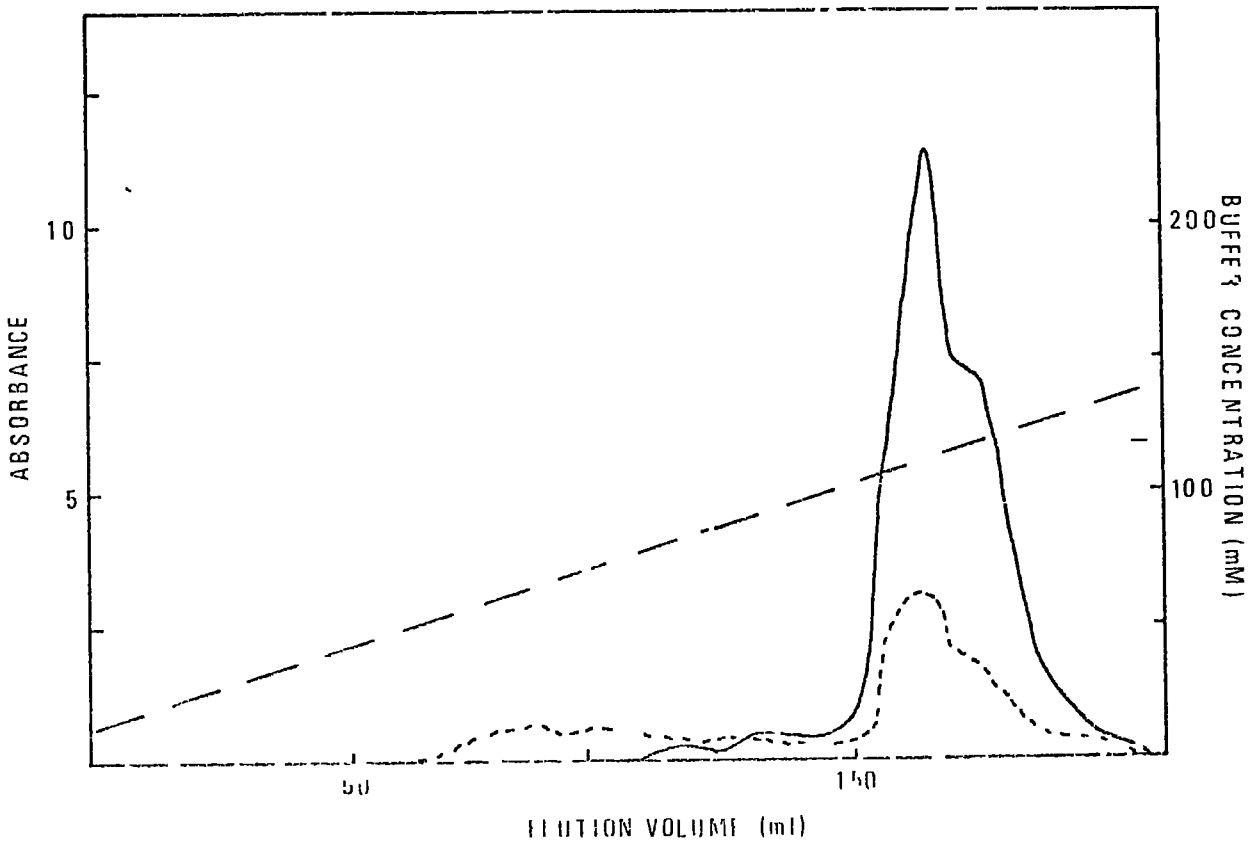
TABLE 6.

Spectral Ratios of Purified *Macrobrachium* Cytochrome *c* (Batch II)

	Ratio
$E^{C2+}_{550}/E^{C3+}_{280}$	1.0
$E^{C3+}_{410}/E^{C3+}_{280}$	4.1
$E^{C2+}_{416}/E^{C3+}_{410}$	1.1
$E^{C2+}_{416}/E^{C2+}_{550}$	4.9

The mass was 13.1 mg salt-free.


75
50
25
ABSORBANCE
100
300
500
ELUTION VOLUME (ml)


10
5
ABSORBANCE
200
100
BUFFER CONCENTRATION (mM)
150
ELUTION VOLUME (ml)

Batch I

27 kg of live worms were separated from Sphagnum moss by light avoidance and washed with tap water before homogenisation with ice in 10 mM-aluminium sulphate solution. The 90 l of homogenate was adjusted to pH 4.5 with 2 M-H_2SO_4 and stirred at 4°C for 2 h. Centrifugation using a 21" basket centrifuge was slow due to the slimy nature of the extract. The best results were obtained using a 30 min run followed by the replacement of the centrifuge bag, and this was repeated until the filtrate ran clean. The residues collected in this way were re-extracted as a precaution with 40 l. of 10 mM-aluminium sulphate. Further purification of the re-extracted filtrate yielded only a small quantity of cytochrome c and this was discarded.

The filtrate obtained from the first centrifugation was taken to pH 8.0 with 2 M-Tris and centrifuged to remove fresh precipitation. Eighty litres of filtrate was then passed through an 18 cm x 10 cm sintered glass funnel packed with Amberlite CG-50 resin regenerated in the NH_4^+ form. The flow rate was 15 l/h and the resin capacity judged to be 20 l. The filtrate from this step was then passed overnight through 10 cm x 56 cm columns of Amberlite CG-50, before being discarded. Both batches of resin were washed well with distilled water, and eluted by the method described for the Locusta preparation. The eluates from both Amberlite

adsorption procedures were found to contain cytochrome c and were combined to give a total volume of 10 l. Dialysis and concentration on CM-50 Sephadex were carried out as described for the Locusta preparation. A thick precipitate interfered with the Sephadex elution which gave a volume of 800 ml. This was concentrated using a Diaflo ultrafiltration unit equipped with a UM-2 membrane under a pressure of 75 p.s.i. maintaining a flow rate of 40 ml/h. The residue from this step was taken by 10% increases to 80% saturation with ammonium sulphate at pH 8.0. No cytochrome c precipitation was observed, although large quantities of a creamy-white precipitate were removed by centrifugation. The sample in 80% saturated ammonium sulphate was dialysed and concentrated on a short column of CM-52 cellulose as described for the Locusta preparation. The partially purified cytochrome c was passed through a 2 cm x 90 cm column of Biogel P-30 equilibrated in 10 mM-sodium phosphate at pH 7.2 (Figure 15) and the pooled fractions from this step had an $E_{410}^{C3+}/E_{280}^{C3+}$ purity ratio of less than 0.5. The cytochrome was then adsorbed onto a 1 cm x 15 cm column of CM-52 cellulose equilibrated in 10 mM-sodium phosphate buffer at pH 7.2, and eluted using a linear ionic gradient (Figure 16). The pooled fractions from this step gave an $E_{410}^{C3+}/E_{280}^{C3+}$ value of 2.1. After an appropriate dilution, the cytochrome c was adosrbed onto another 1 cm x 15 cm column of CM-52 cellulose

equilibrated in 10 mM-sodium phosphate buffer at pH 7.2 and oxidised and washed as before. The sample was eluted using a linear pH gradient (Figure 17) and fractions were collected into tubes containing 0.5 ml of 50 mM-sodium phosphate buffer at pH 7.2 in order to minimize the time of exposure to pH values greater than pH 10.0. The fractions selected from this purification step had a pooled $E^{C3+}_{410}/E^{C3+}_{280}$ purity ratio of 4.0 and these samples were desalted on a 1 cm x 15 cm column of Sephadex G-10, freeze-dried and stored at -20°C.

Details of the yields and $E^{C3+}_{410}/F^{C3+}_{280}$ purity ratios of Eisenia cytochrome c during the various stages of the purification of Batch I are shown in Table 7 and the final spectral ratios obtained from the pure protein are given in Table 8. The elution profiles of the chromatographic fractionation steps are shown in Figures 15, 16 and 17.

Batch II

22 kg of live worms in a rather poor condition were separated from Sphagnum moss by light avoidance and washed in tap water. The treatment was identical to Batch I except for a number of points:-

(1) The concentration step following the CM-50 Sephadex chromatography was by means of a 2 cm x 5 cm column of CM-50 Sephadex equilibrated in 50 mM-sodium phosphate at pH 8.0 and eluted with 500 mM-NaCl in 50 mM-sodium phosphate at pH 8.0.

(2) The pH gradient on CM-52 cellulose was applied before

TABLE 7.

The Purification of *Eisenia* Cytochrome *c*
(Batch I)

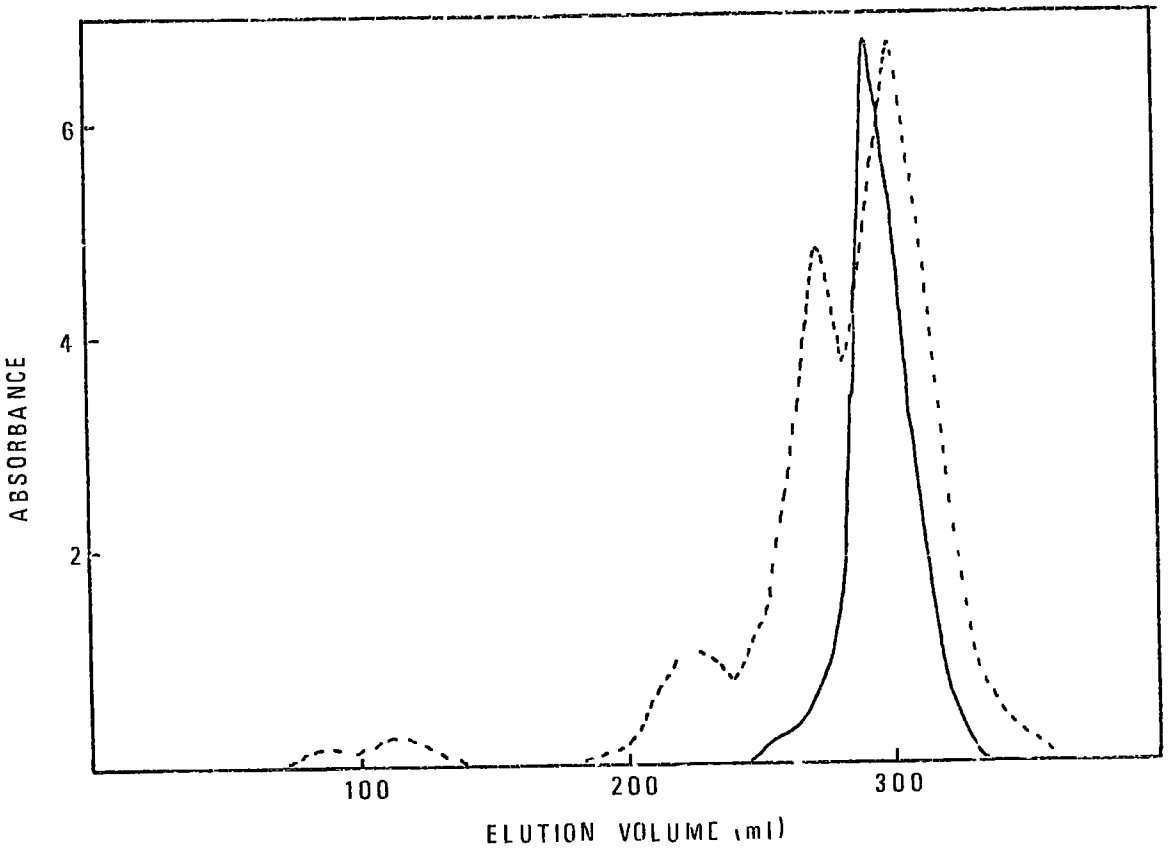
Step	Yield mg cytochrome/ kg starting material	Purity Ratio $E^{C3+}_{410}/E^{C3+}_{280}$
Amberlite CG-50 CM-50 Sephadex Ammonium Sulphate Fractionation	1.0	0.2
Biogel P-30	0.6	0.5
CM-52 Cellulose, Ionic Gradient	0.4	2.1
CM-52 Cellulose, pH Gradient	0.3	3.9
G-10 Sephadex	0.3	4.0

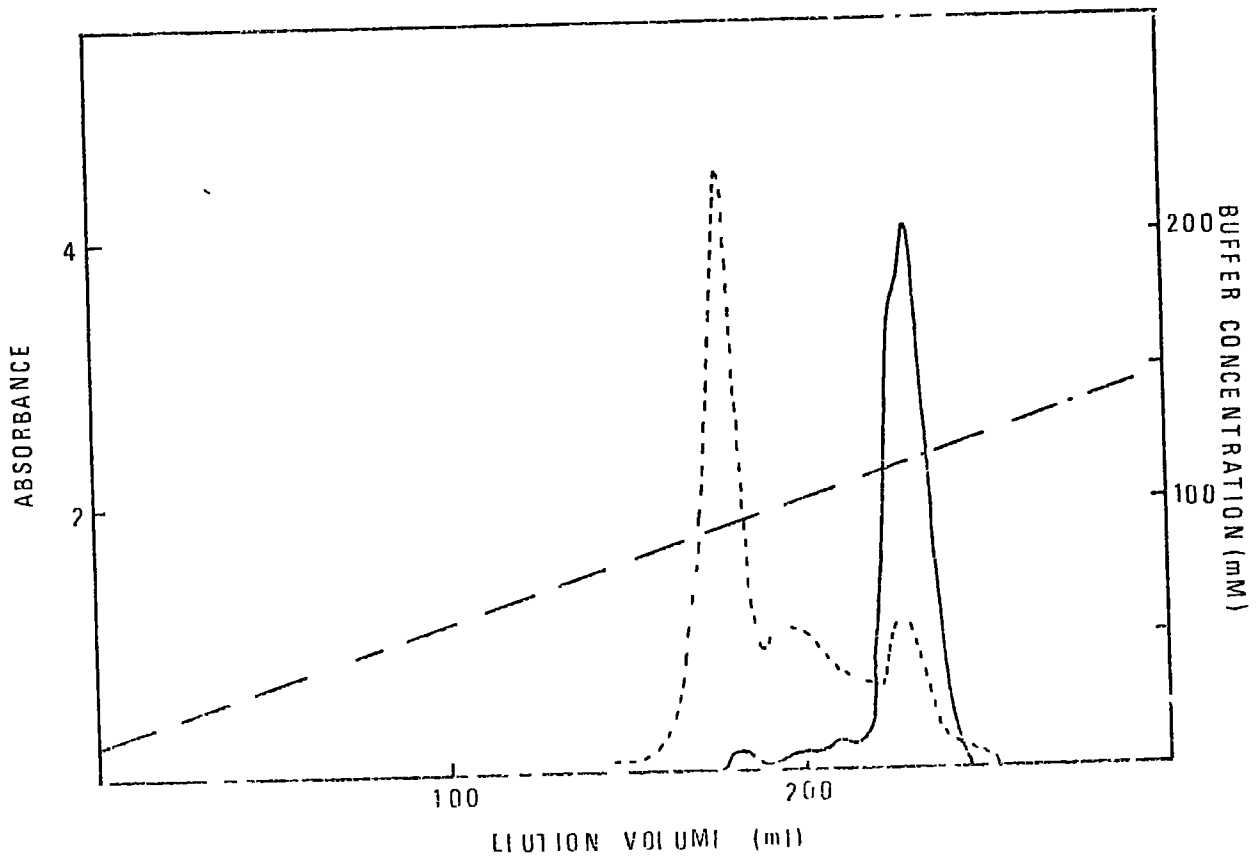
TABLE 8.

The Spectral Ratios of Purified *Eisenia* Cytochrome *c* (Batch I)

	Ratio
$E^{C2+}_{550}/E^{C3+}_{280}$	1.1
$E^{C3+}_{410}/E^{C3+}_{280}$	4.0
$E^{C2+}_{416}/E^{C3+}_{410}$	1.2
$E^{C2+}_{416}/E^{C2+}_{550}$	4.6

The mass was 6.3 mg salt-free.


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


4
2
ABSORBANCE
200
100
BUFFER CONCENTRATION (mM)
100
200
ELUTION VOLUME (ml)

the ionic gradient in order to avoid the adjusting of ionic strengths between steps.

Pure samples were desalted on a 1 cm x 10 cm column of Sephadex G-10, freeze-dried and stored at -20°C.

Details of the yields and $E_{410}^{C3+}/E_{280}^{C3+}$ purity ratios during the various stages of the purification of Batch II are shown in Table 9 and the final spectral ratios obtained from the pure protein are given in Table 10. The elution profiles of the chromatographic fractionation steps are shown in Figures 18, 19 and 20.

Batch III

22 kg of live worms in good condition were separated from Sphagnum moss by light avoidance and washed in tap water. Extraction was by the normal procedure. Filtration using the basket centrifuge was very slow and gave a poor filtrate but because of the time factor the extract was not filtered again at pH 8.0. 400 g of Amberlite CG-50 resin regenerated in the NH_4^+ form, was stirred with the extract (50 l) for 2 h at 4°C, allowed to settle out by standing, and reclaimed by decantation. The slimy supernatant was then passed through a sintered glass funnel (18 cm x 10 cm) packed with Amberlite CG-50. The flow rate was 5 l/h and the resin capacity was fixed at 10 l. After extensive washing with distilled water, the total Amberlite was eluted batchwise, using the method

TABLE 9.

The Purification of Eisenia Cytochrome c
(Batch II)

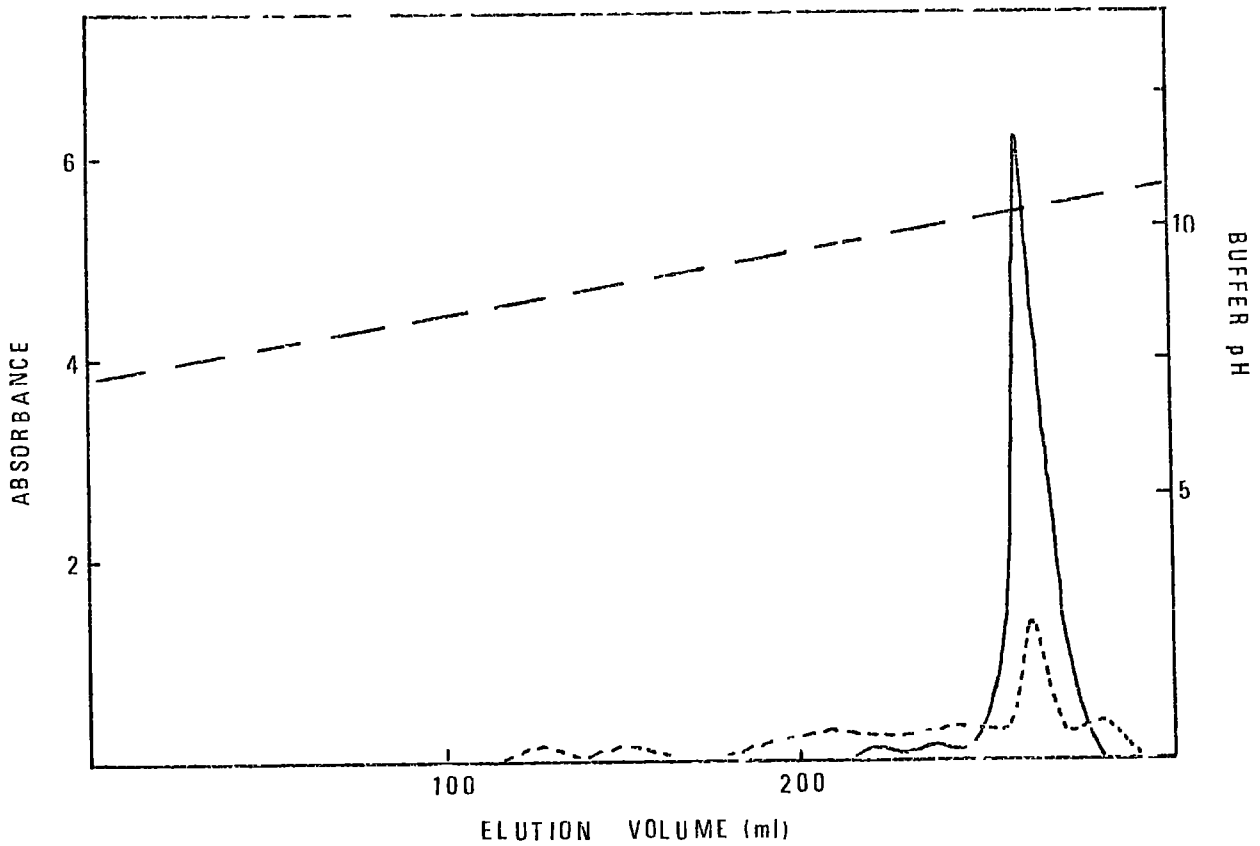
Step	Yield mg cytochrome/ kg starting material	Purity Ratio $E^{C3+}_{410}/E^{C3+}_{280}$
Amberlite CG-50 } CM-50 Sephadex }	2.0	0.8
Ammonium Sulphate Fractionation	1.5	1.0
Biogel P-30	1.0	1.4
CM-52 Cellulose, pH Gradient	0.8	2.6
CM-52 Cellulose, Ionic Gradient	0.6	4.0
G-10 Sephadex	0.6	4.0

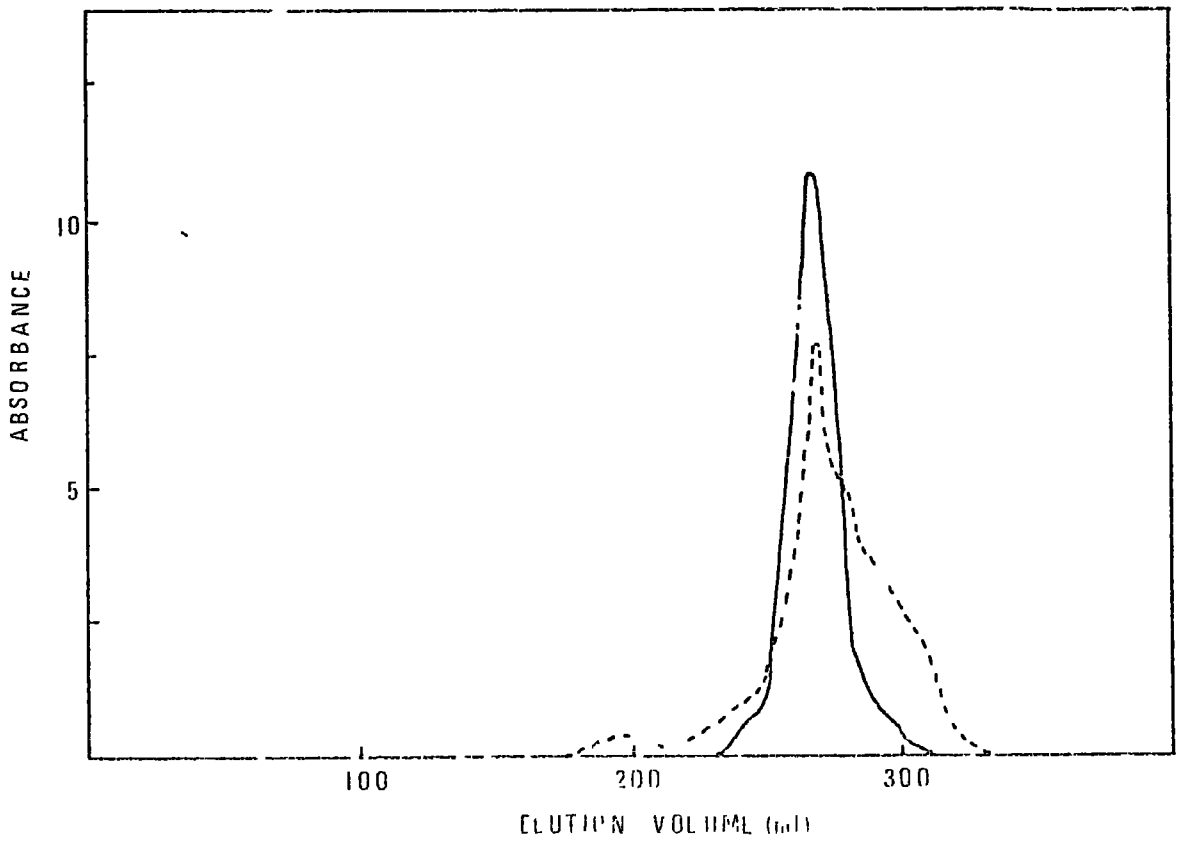
TABLE 10.

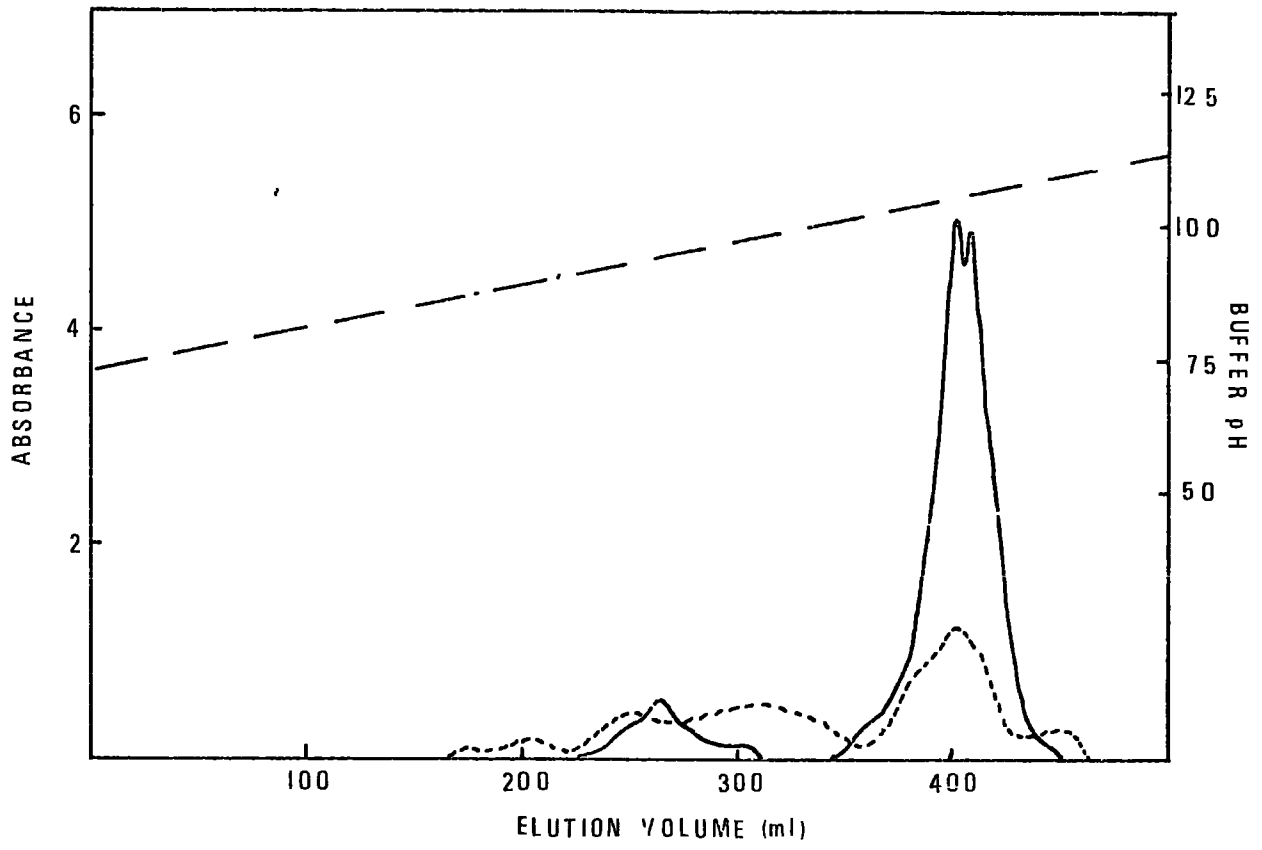
The Spectral Ratios of Purified Eisenia Cytochrome c (Batch II)

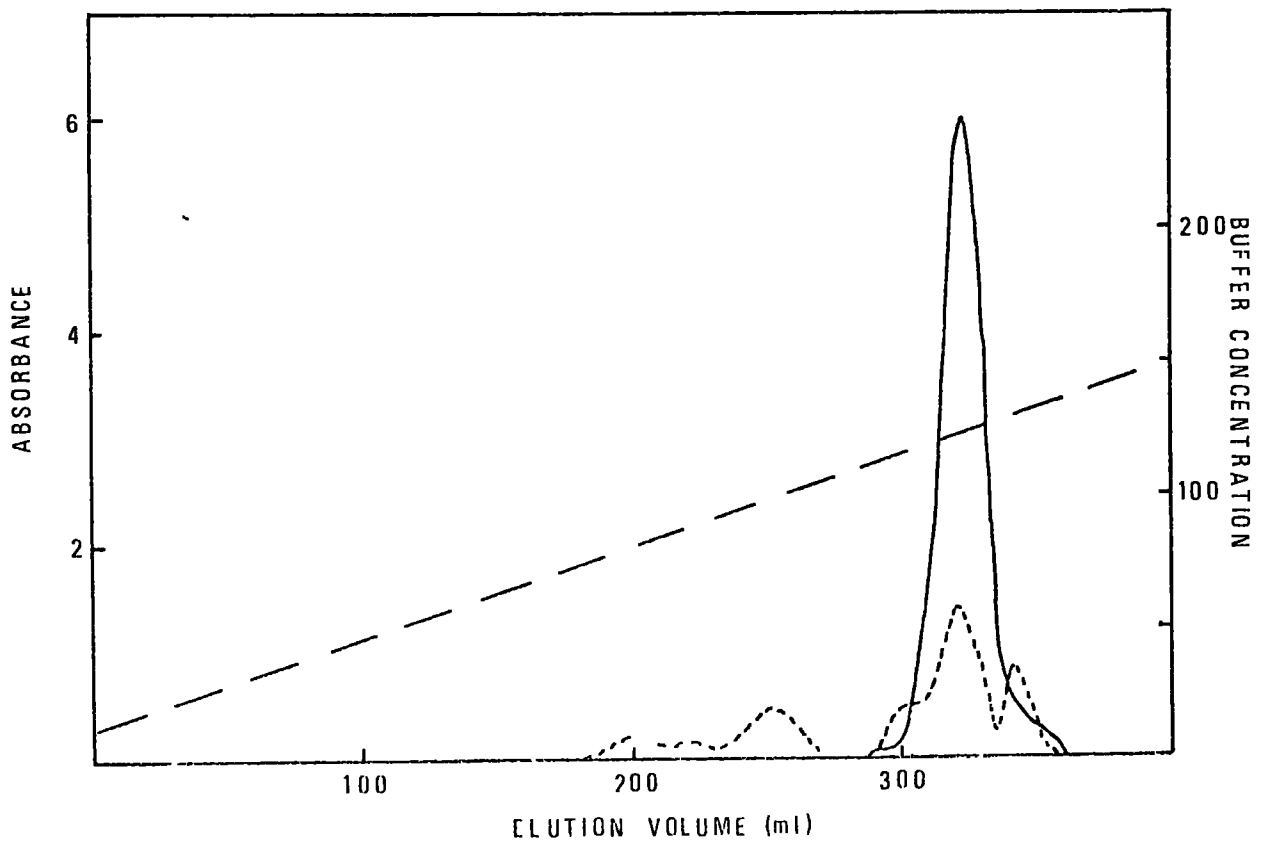
	Ratio
$E^{C2+}_{550}/E^{C3+}_{280}$	1.1
$E^{C3+}_{410}/E^{C3+}_{280}$	4.0
$E^{C2+}_{416}/E^{C3+}_{410}$	1.2
$E^{C2+}_{416}/E^{C2+}_{550}$	4.2

The mass was 13 mg salt-free.


ABSORBANCE
6
4
2
10
5
BUFFER pH
100
200
ELUTION VOLUME (ml)


ABSORBANCE
10
5
100
200
300
ELUTION VOLUME (ml)


ABSORBANCE
6
4
2
125
100
75
50
BUFFER pH
100
200
300
400
ELUTION VOLUME (ml)


ABSORBANCE
6
4
2
200
100
BUFFER CONCENTRATION
100
200
300
ELUTION VOLUME (ml)

described for the Locusta preparation. Six litres of eluate was dialysed and concentrated on CM-50 Sephadex as described for the Locusta preparation, to give a volume of 10 ml.

The partially purified cytochrome c was taken by 10% stages to 80% saturation with ammonium sulphate. Large amounts of a white precipitate were discarded after centrifugation but no precipitation of the cytochrome was observed. The solution was dialysed and concentrated on a short column of CM-52 cellulose as described for the Locusta preparation.

A gel filtration step using Biogel P-30 (Figure 21) was followed by concentration of the cytochrome c on a 1 cm x 15 cm column of CM-52 cellulose equilibrated in 10 mM-sodium phosphate at pH 7.2. After oxidation and washing, the column was eluted using a linear ionic gradient of 10 mM to 300 mM-sodium phosphate at pH 7.2 (Figure 22). The best $E_{410}^{C3+}/E_{280}^{C3+}$ purity ratios for this step were less than 2.5 and after an appropriate dilution the chromatography was repeated using a 1 cm x 30 cm column of CM-52 cellulose. This gave a pool of 8 mg of partially purified cytochrome c with a purity ratio of 3.6. In view of the expected losses from a further purification step, the cytochrome c was desalted on a 1 cm x 10 cm column of Amberlite MB-1 equilibrated in distilled water, freeze-dried and stored at -20°C.

Details of the yields and $E_{410}^{C3+}/E_{280}^{C3+}$ purity ratios during the various purification stages of Batch III are shown in

Table 11, and the final spectral ratios obtained for the protein are given in Table 12. The elution profiles of the chromatographic fractionation steps are shown in Figures 21, 22 and 23.

4. Purification of Asterias cytochrome c

The purification of Asterias cytochrome c was carried out in three batches.

Batch I

20 kg of frozen starfish were rapidly thawed at 4°C and homogenised as whole animals with ice in 10 mM-aluminium sulphate using a Waring blender. The homogenate was stirred for 2 h at pH 6.5 and 4°C before filtration through two layers of muslin to give 150 l of crude filtrate. The large amount of residue was discarded on the evidence of a test run, where re-extraction at this stage was shown to produce negligible quantities of cytochrome c. The filtrate was adjusted to pH 8.0 with 2 M-Tris and filtered using a 21" basket centrifuge fitted with a Terylene bag. Filtration was very slow, despite frequent changes of the centrifuge bag, and the processing of the total extract took 48 h. The filtrate was passed through 6 cm x 30 cm columns of Amberlite CG-50, regenerated in the NH_4^+ form. Constant repacking of the columns, together with the application of a partial vacuum, gave a flow rate of 2 l/h. The resin was well washed with

TABLE 11.

The Purification of Eisenia cytochrome c
(Batch III)

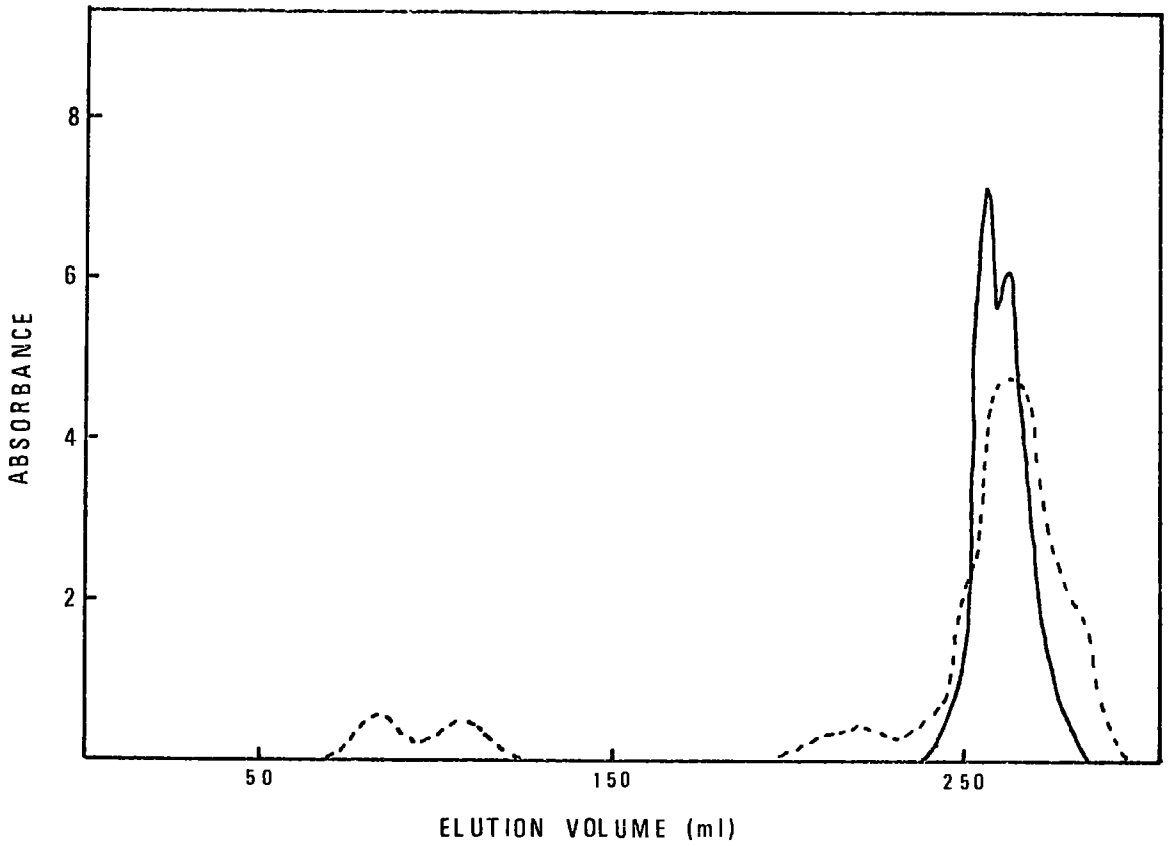
Step	Yield mg cytochrome / kg starting material	Purity Ratio $E^{C3+}_{410}/E^{C3+}_{280}$
Amberlite CG-50 CM-50 Sephadex Ammonium Sulphate Fractionation	1.0	0.5
Biogel P-30	0.8	1.0
CM-52 Cellulose, Short Column	0.5	2.0
CM-52 Cellulose, Long Column	0.4	3.6
Amberlite MB-1	0.3	3.5

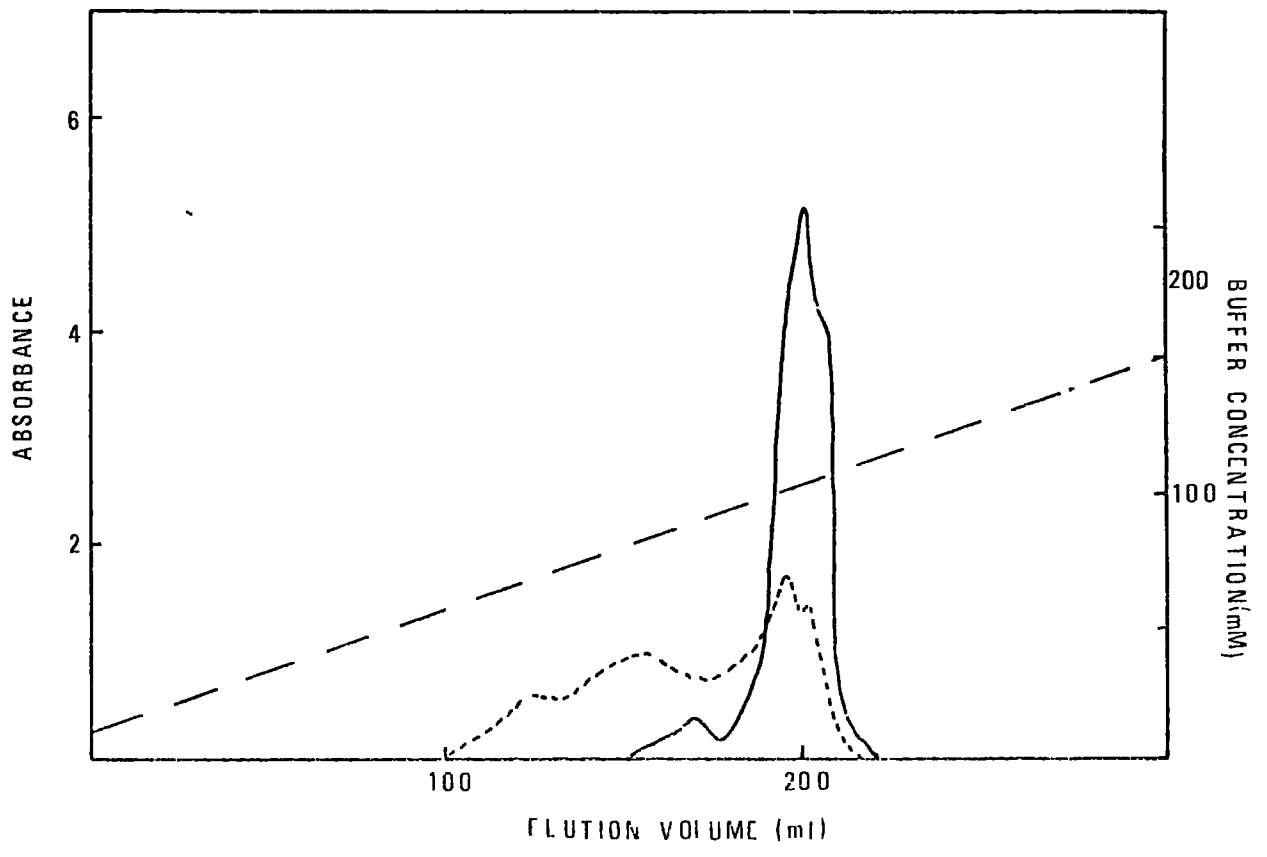
TABLE 12.

The Spectral Ratios of Purified Eisenia Cytochrome c (Batch III)

	Ratio
$E_{550}^{C2+}/E_{280}^{C3+}$	0.5
$E_{410}^{C3+}/E_{280}^{C3+}$	3.5
$E_{416}^{C2+}/E_{410}^{C3+}$	1.1
$E_{416}^{C2+}/E_{550}^{C2+}$	5.1

The mass was 6.4 mg salt-free.


8
6
4
2
ABSORBANCE
50
150
250
ELUTION VOLUME (ml)


6
4
2
ABSORBANCE
200
100
BUFFER CONCENTRATION(mM)
100
200
ELUTION VOLUME (ml)

distilled water and eluted batchwise using the method described for the Locusta preparation. Seven litres of eluate were dialysed against 120 l of 10 mM-sodium phosphate at pH 8.0 for 8 h and against 50 mM-sodium phosphate buffer at pH 8.0 for a further 8 h. After centrifugation at 2000 g for 30 min to remove a white precipitate, the cytochrome c was concentrated by passage through a 6 cm x 15 cm column of CM-50 Sephadex equilibrated in 50 mM-sodium phosphate at pH 8.0. Elution of the main fraction was with 0.5 M-NaCl in 50 mM-sodium phosphate buffer at pH 8.0, but examination of the column using a direct-vision, hand spectroscope, revealed that a substantial amount of cytochrome c remained bound to a white fat-like deposit that had collected on the top of the resin. The cytochrome was found to dissociate and eluate using 1 M-NaCl. Two litres of eluate were dialysed and concentrated on a short column of CM-50 Sephadex, as described for the Locusta preparation, to give a volume of 40 ml. This solution was taken in 10% stages to 80% saturation with ammonium sulphate at pH 8.0. Large amounts of a grey-white precipitate were removed by centrifugation but no precipitation of cytochrome c was observed. The solution was dialysed and concentrated on a short column of CM-52 cellulose, as described for the Locusta preparation, giving a volume of 5 ml. A gel filtration step with Biogel P-30 (Figure 24) was followed by the adsorption of the cytochrome c to a 1 cm x 10 cm

column of CM-52 cellulose equilibrated in 10 mM-sodium phosphate buffer at pH 7.2. The sample was eluted using a linear pH gradient between pH 7.2 and 11.5 (Figure 25). Fractions were collected in test-tubes containing 0.5 ml of 10 mM-sodium phosphate buffer at pH 7.2. The selected elution pool had a $E^{C3+}_{410}/E^{C3+}_{280}$ purity ratio of 2.1. After an appropriate pH adjustment, the sample was adsorbed onto a further 1 cm x 10 cm column of CM-52 cellulose and eluted using a linear ionic gradient (Figure 26). The selected elution pool contained 11 mg of cytochrome c with an $E^{C3+}_{410}/E^{C3+}_{280}$ purity ratio of 3.8. The sample was desalted using a 1 cm x 10 cm column of Sephadex G-10 equilibrated in distilled water, freeze-dried and stored at $-20^{o}C$.

Details of the yields and $E^{C3+}_{410}/E^{C3+}_{280}$ purity ratios at the various purification stages are shown in Table 13, and the final spectral ratios obtained for the protein are given in Table 14. The elution profiles of the chromatographic fractionation steps are shown in Figures 24, 25 and 26.

Batch II

27 kg of frozen starfish were rapidly thawed at $4^{o}C$ and homogenated in an identical manner to Batch I to give a volume of 150 l. Filtration was by two layers of muslin at pH 6.5 and the 21" basket centrifuge at pH 8.0. The filtrate was passed through a sintered glass funnel (18 cm x 10 cm) packed with Amberlite CG-50 regenerated in the NH_4^+ form. The flow

TABLE 13.

The Purification of *Asterias* Cytochrome *c* (Batch I)

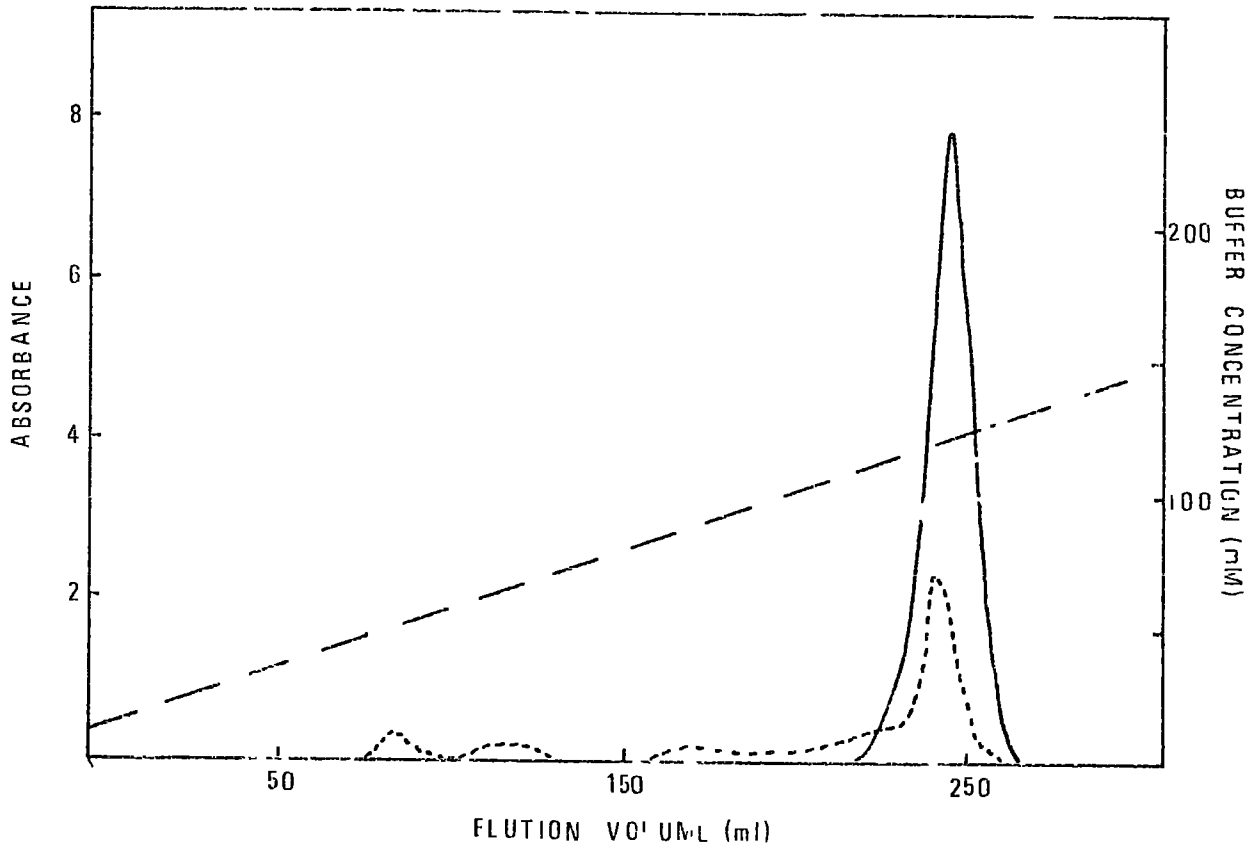
Step	Yield: mg cytochrome/ kg starting material	Purity Ratio: $E^{C3+}_{410}/E^{C3+}_{280}$
Amberlite CG-50	-	-
CM-50 Sephadex	2.0	0.5
Ammonium Sulphate Fractionation	1.8	0.7
Biogel P-30	1.1	0.9
CM-52 Cellulose, pH Gradient	0.8	2.1
CM-52 Cellulose, Ionic Gradient	0.6	3.8
G-10 Sephadex	0.5	3.8

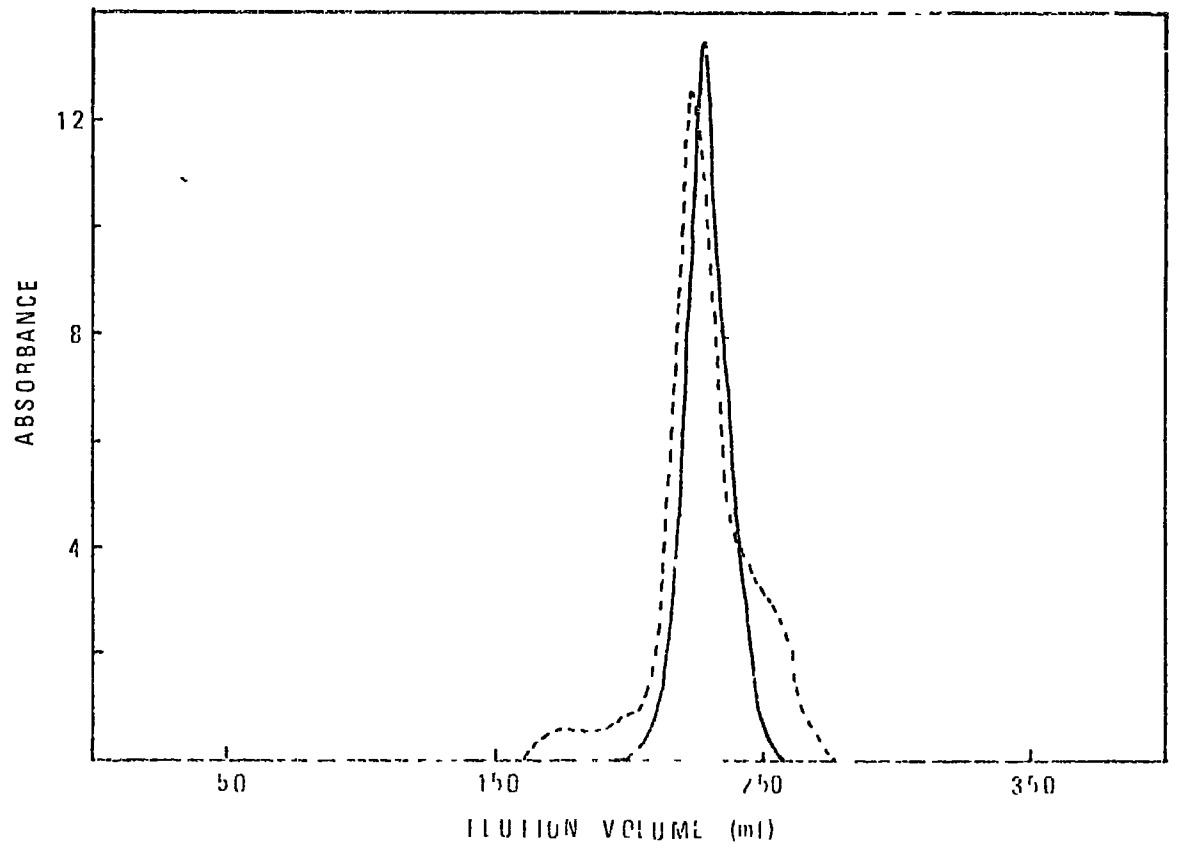
TABLE 14.

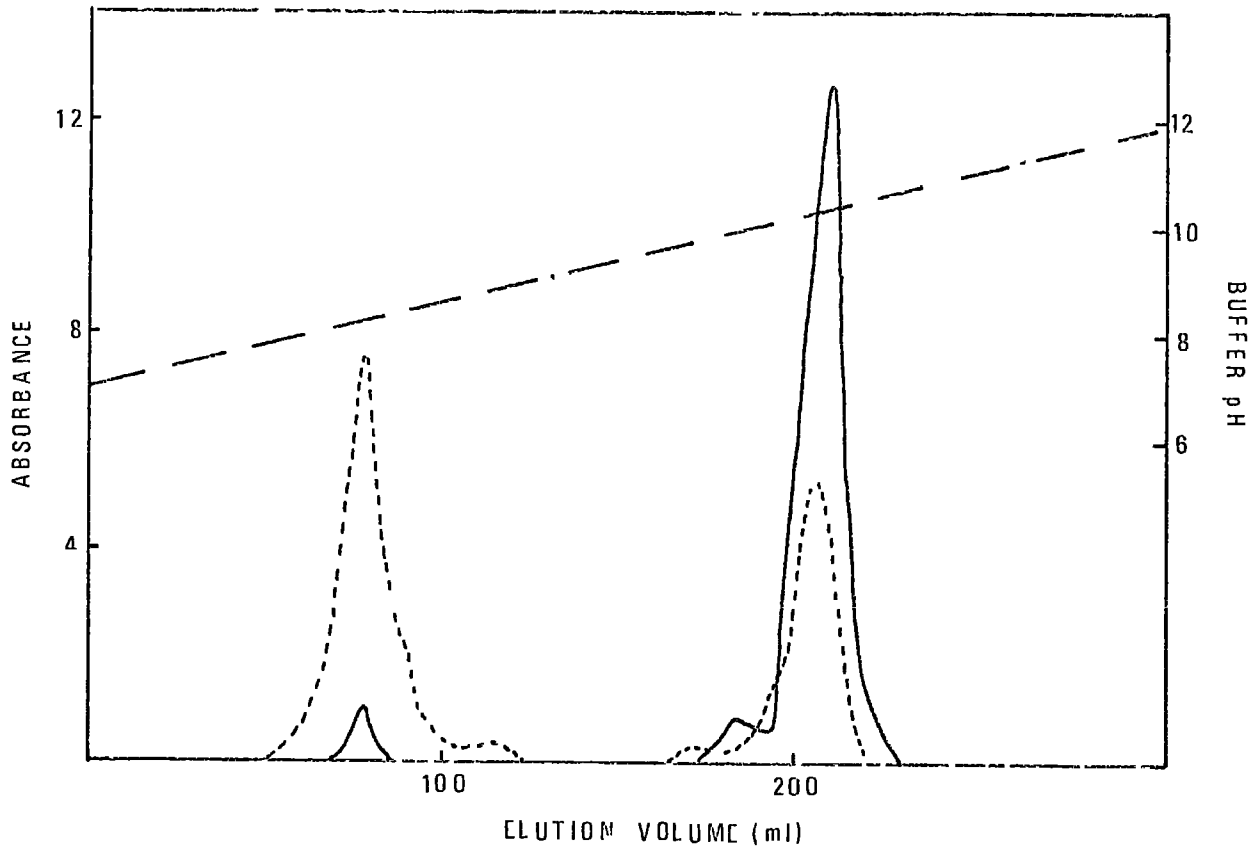
The Spectral Ratios of Purified Asterias Cytochrome c (Batch T)

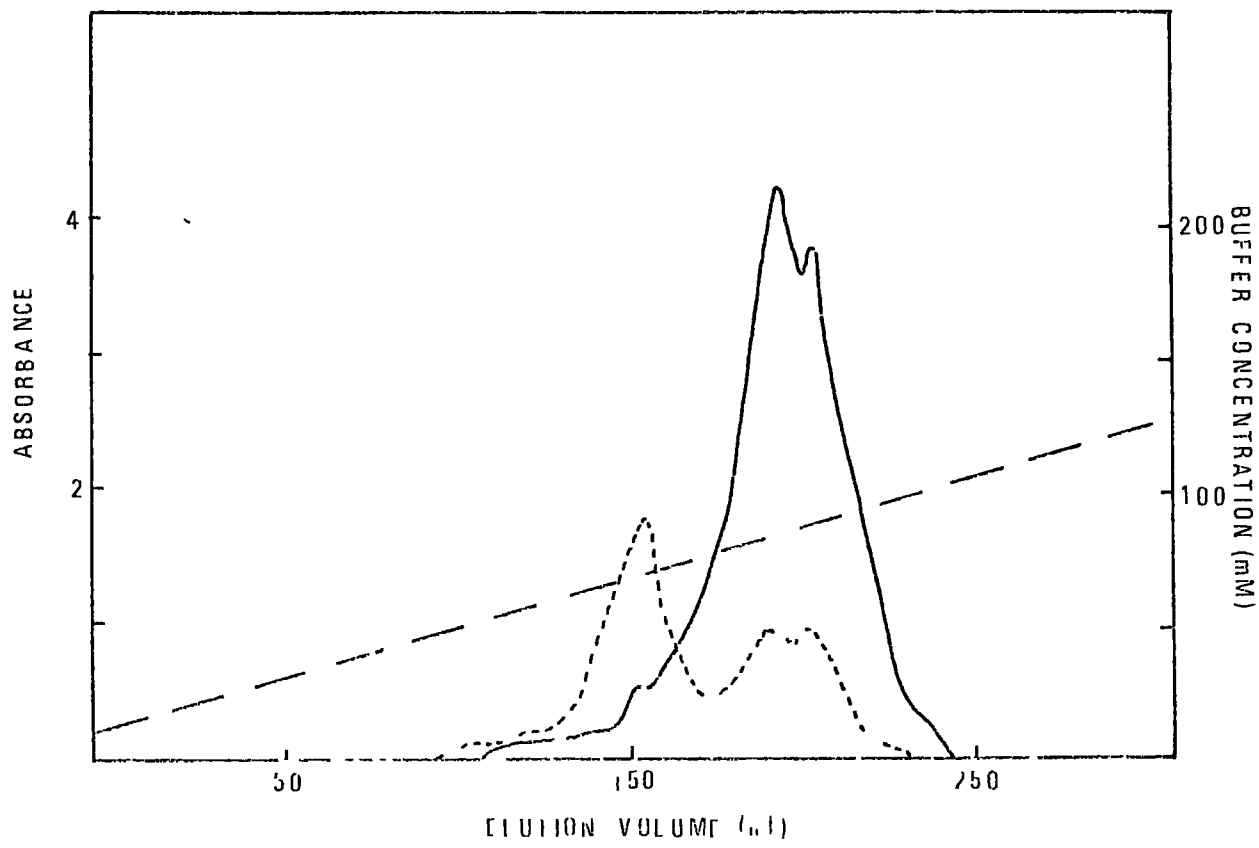
	Ratio
$E^{C2+}_{550}/E^{C3+}_{280}$	1.1
$E^{C3+}_{410}/E^{C3+}_{280}$	3.8
$E^{C2+}_{416}/E^{C3+}_{410}$	1.2
$E^{C2+}_{416}/E^{C2+}_{550}$	4.6

The mass was 8.4 mg salt-free.


ABSORBANCE
8
6
4
2
BUFFER CONCENTRATION (mM)
200
100
50
150
250
FLUTION VOLUME (ml)


ABSORBANCE
12
8
4
50
150
250
350
ELUTION VOLUME (ml)


12
8
4
ABSORBANCE
12
10
8
6
BUFFER pH
100
200
ELUTION VOLUME (ml)


4
2
ABSORBANCE
200
100
BUFFER CONCENTRATION (mM)
50
150
250
ELUTION VOLUME (ml)

rate was 15 l/h and the resin capacity was fixed at 20 l. The effluent from this step was then passed through three 6 cm x 30 cm columns of Amberlite CG-50 resin. Application of a partial vacuum gave a flow rate of 5 l/h. The columns were repacked every 2 h to maintain this rate. The Amberlite resin was washed and eluted and the eluate was dialysed and concentrated on CM-50 Sephadex as described for the Locusta preparation.

Extensive precipitation of impurities was observed during the ammonium sulphate fractionation, but no precipitation of cytochrome c occurred up to 80% saturation. The partially purified cytochrome c in 80% saturated ammonium sulphate was dialysed exhaustively, and concentrated on a short column of CM-52 cellulose as described for the Locusta preparation. Gel-filtration using a 2 cm x 100 cm column of Biogel P-30 was found to only marginally increase the $E_{410}^{C3+}/E_{280}^{C3+}$ purity ratio of the sample to 0.6, but an ionic gradient applied to CM-52 cellulose served to increase the ratio seven-fold. The sample was adsorbed to the top of a 1 cm x 30 cm column of CM-52 cellulose and was eluted using a linear ionic gradient (Figure 27). The selected elution pool contained 24 mg of cytochrome c with an $E_{410}^{C3+}/E_{280}^{C3+}$ purity ratio of 4.1. The sample was desalted using a 1 cm x 10 cm column of G-10 Sephadex, freeze-dried and stored at -20°C.

Details of the yields and $E_{410}^{C3+}/E_{280}^{C3+}$ purity ratios at

the various purification stages are shown in Table 15 and the final spectral ratios obtained for the protein are given in Table 16. The elution profile of the fractionation using CM-52 cellulose with an ionic gradient is shown in Figure 21.

Batch III

20 kg of frozen starfish were thawed rapidly at 4°C. The homogenisation and filtration procedures were identical to Batch I. One hundred litres of filtrate at pH 8.0 were passed through an 18 cm x 10 cm sintered glass funnel packed with Amberlite CG-50 resin in the NH_4^+ form. The flow rate was 15 l/h and the resin capacity was fixed at 20 l. The effluent from this step was dripped through 6 cm x 15 cm columns of Amberlite CG-50 at 2 l/h. The resin was eluted as before to give a volume of 6 l which was dialysed against running tap water for 8 h and against 60 l of 50 mM-sodium phosphate at pH 8.0 for a further 8 h. A thick white precipitate was removed from the dialysis residue by centrifugation at 2000 g for 30 min and the supernatant was concentrated on CM-50 Sephadex as described for the Locusta preparation. The ammonium sulphate fractionation removed large quantities of a white precipitate but no cytochrome c was precipitated at saturations up to 80%.

The final purification step involved the use of a 1 cm x 15 cm column of CM-52 cellulose. The adsorption and elution procedures were as described for the Locusta preparation

TABLE 15.

The Purification of *Asterias* Cytochrome *c* (Batch II)

Step	Yield mg cytochrome kg starting material	Purity Ratio $E^{C3+}_{410}/E^{C3+}_{280}$
Amberlite CG-50 CM-50 Sephadex Ammonium Sulphate Fractionation	2.0	0.4
Biogel P-30	1.0	0.6
CM-52 Cellulose, Ionic Gradient	0.8	4.1
G-10 Sephadex	0.6	3.8

TABLE 16.

The Spectral Ratios of Purified Asterias Cytochrome c (Batch II)

	Ratio
$E_{550}^{C2+}/E_{280}^{C3+}$	0.8
$E_{410}^{C3+}/E_{280}^{C3+}$	3.8
$E_{416}^{C2+}/E_{410}^{C3+}$	1.1
$E_{416}^{C2+}/E_{550}^{C2+}$	5.6

The mass was 24 mg salt-free.

and the pooled eluate had an $E_{410}^{C3+}/E_{280}^{C3+}$ purity ratio of 3.8. The sample was desalted using a 1 cm x 15 cm column of Amberlite MB-1 resin, freeze-dried and stored at -20°C.

Details of the yields and $E_{410}^{C3+}/E_{280}^{C3+}$ purity ratios at the various purification stages are shown in Table 17 and the final spectral ratios obtained for the protein are given in Table 18. The elution profiles of the chromatographic fractionation steps are shown in Figures 28 and 29.

5. Purification of Loligo Cytochrome c

11 kg of frozen squid were rapidly thawed at 4°C and washed in tap water. The animals were homogenised in a Waring blender with 10 mM-aluminium sulphate and ice to give a total volume of 35 l which was stirred at pH 4.5 and 4°C for 2 h. The filtration flow rate was only 2 l/h using the 21" basket centrifuge, despite hourly changes of the centrifuge bag. The filtrate at pH 8.0 was passed through a sintered glass funnel packed with Amberlite CG-50 regenerated in the NH_4^+ form. The flow rate was 15 l/h and the resin capacity was fixed at 20 l.

The effluent from this step was recycled and then discarded. The Amberlite resin was extensively washed with distilled water and eluted batchwise with 2 M-NaCl maintained at pH 8.0 with 2 M-NaOH.

The 4 l of eluate was dialysed against running tap water

TABLE 17.

The Purification of Asterias Cytochrome c
(Batch III)

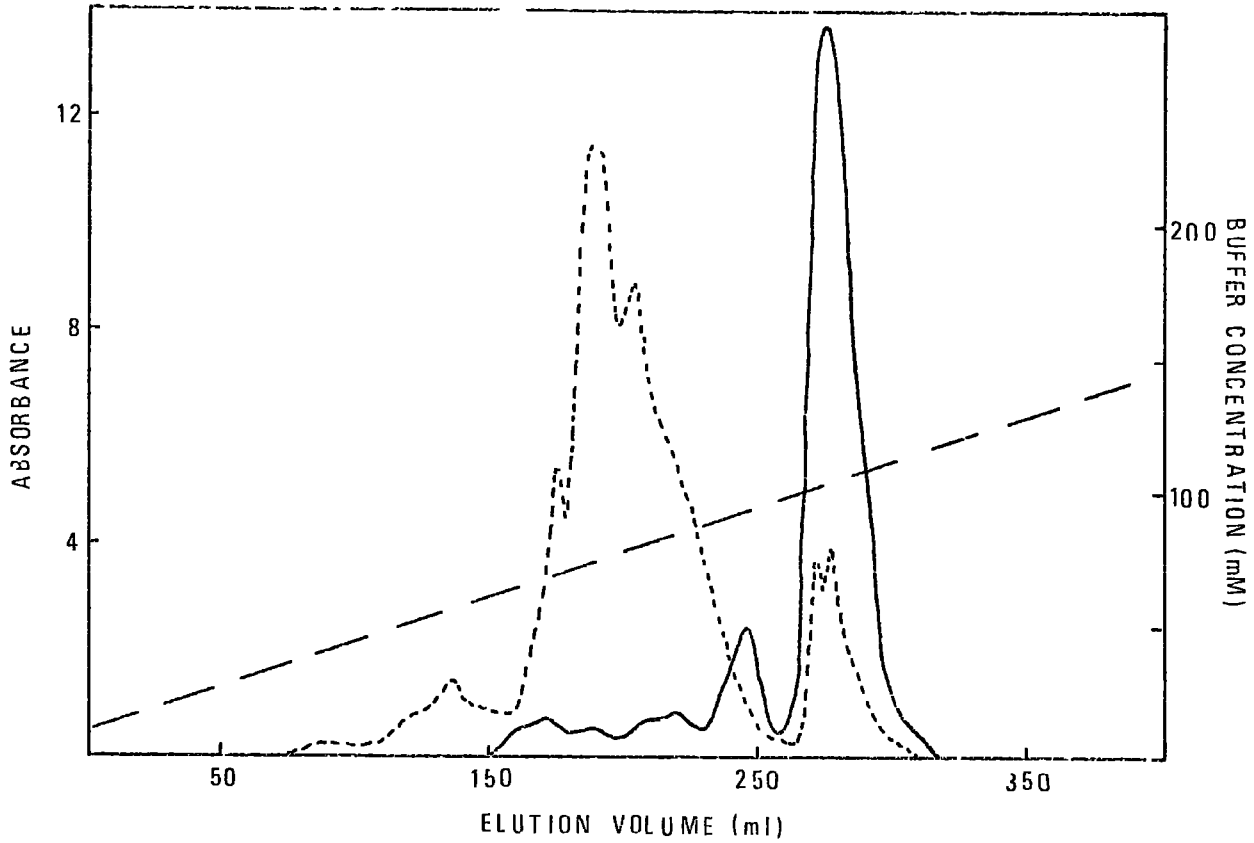
Step	Yield mg cytochrome/ kg starting material	Purity Ratio $E^{C3+}_{410}/E^{C3+}_{280}$
Amberlite CG-50 CM-50 Sephadex Ammonium Sulphate Fractionation	1.0	0.2
Biogel P-30	0.5	0.4
CM-52 Cellulose, Ionic Gradient	0.3	3.8
Amberlite MB-1	0.2	4.0

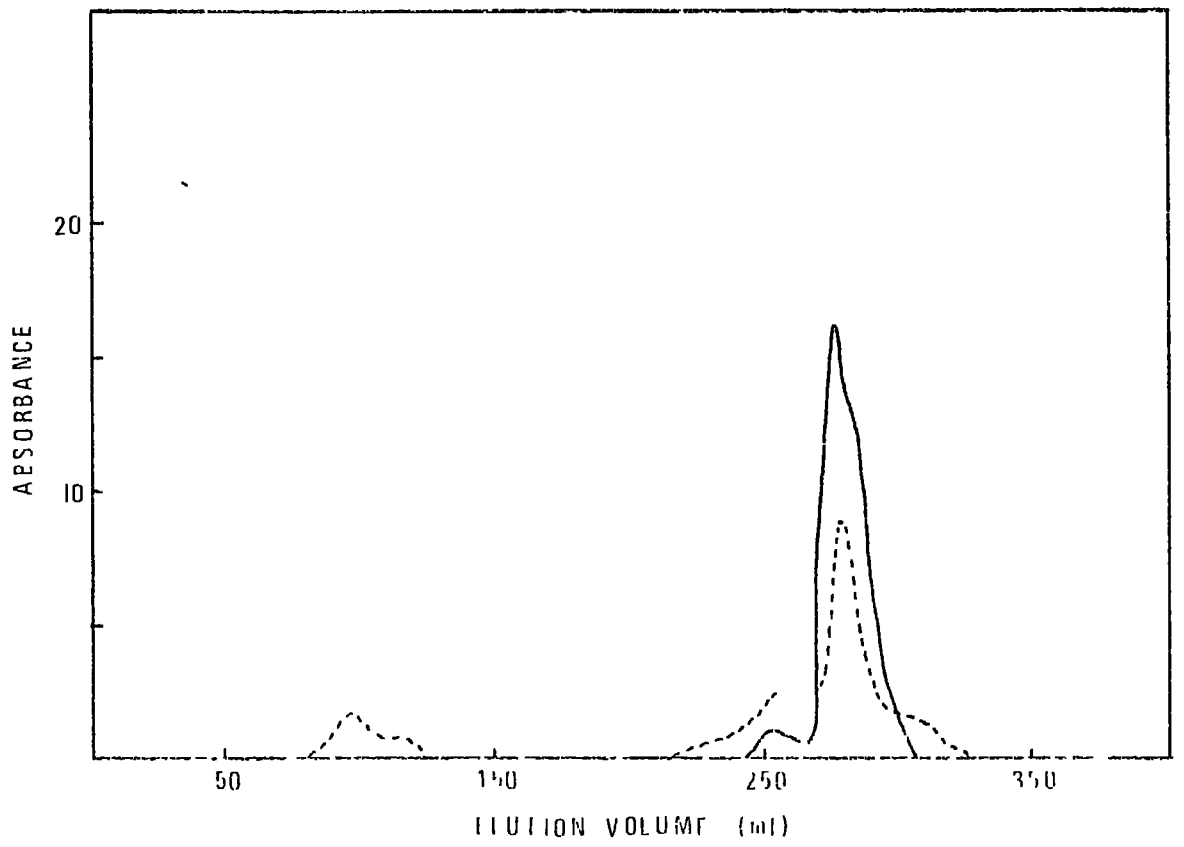
TABLE 18.

The Spectral Ratios of Purified Asterias Cytochrome c (Batch III)

	Ratio
$E_{550}^{C2+}/E_{280}^{C3+}$	0.8
$E_{410}^{C3+}/E_{280}^{C3+}$	4.0
$E_{416}^{C2+}/E_{410}^{C3+}$	1.2
$E_{416}^{C2+}/E_{550}^{C2+}$	6.0

The mass was 3.8 mg salt-free.


ABSORBANCE
12
8
4
200
100
BUFFER CONCENTRATION (mM)
50
150
250
350
ELUTION VOLUME (ml)


ABSORBANCE
20
10
50
150
250
350
ELUTION VOLUME (ml)

for 8 h and against 60 l of 50 mM-sodium phosphate buffer at pH 8.0 plus 4°C for a further 8 h. A white precipitate was partially removed from the dialysis residue by centrifugation at 2000 g for 30 min, and the cytochrome c was concentrated on CM-50 Sephadex as described for the Locusta preparation.

20 ml of partially purified cytochrome c was taken by 10% stages to 60% saturation with ammonium sulphate at pH 8.0. No precipitation of cytochrome c was observed but quantities of a white precipitate were removed by centrifugation. At saturations above 60% precipitation of cytochrome c occurred, but at 80% saturation a good deal of cytochrome c was still in solution despite overnight stirring at 4°C and prolonged centrifugation. The sample was diluted twice to redissolve the precipitated protein and dialysed exhaustively against changes of 10 mM-sodium phosphate buffer at pH 7.2. 300 ml of dialysis residue was then pumped through a 1 cm x 30 cm of CM-52 cellulose equilibrated in 10 mM-sodium phosphate buffer at pH 7.2. The cytochrome c adsorbed to the column and was eluted using a linear ionic gradient (Figure 30).

The best $E_{410}^{C3+}/E_{280}^{C3+}$ purity ratios from the eluted fractions had values of 3.7 to 4.2 and these were taken for desalting using a 1 cm x 10 cm column of Amberlite MB-1 resin equilibrated and eluated with distilled water. Salt-free samples were freeze-dried and stored at -20°C.

Details of the yields and $E_{410}^{C3+}/E_{280}^{C3+}$ purity ratios at

the various stages of purification are shown in Table 19 and the final spectral ratios obtained for the protein are given in Table 20. The elution profile of the CM-52 cellulose step is shown in Figure 30.

6. Extraction of Cytochrome c from Solaster

11 kg of Solaster were rapidly thawed at 4°C and homogenised in the normal manner with ice and 10 mM-aluminium sulphate. The homogenate was stirred at pH 6.5 for 2 h and filtered using a 21" basket centrifuge. The filtrate was adjusted to pH 8.0 with 2 M-Tris without any further precipitation. The filtrate was passed through an 18 cm x 10 cm sintered glass funnel packed with Amberlite and the resin was washed and eluted batchwise in the usual way. After the addition of a trace of ascorbic acid no cytochrome c could be seen in the eluate using the direct-vision hand spectroscope. The sample was dialysed against 50 mM-sodium phosphate at pH 8.0 and passed through a CM-50 Sephadex column. Only a faint trace of cytochrome c was found to be bound to the top of the column and the purification was terminated at this point due to the lack of material.

7. Extraction of Cytochrome c from Aphrodite

9 kg of Aphrodite were thawed at 4°C and washed thoroughly in tap water. The normal homogenisation procedure was

TABLE 19.

The Purification of Loligo Cytochrome c

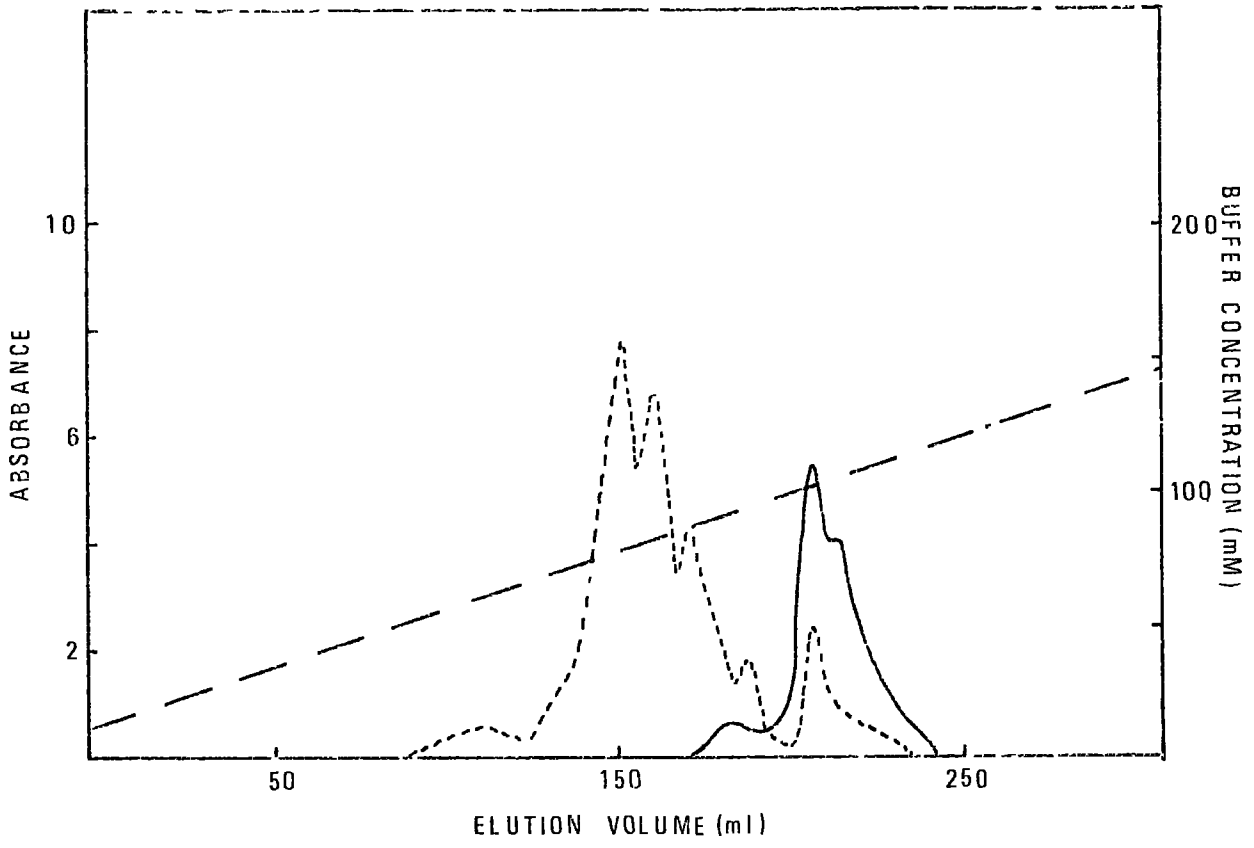
Step	Yield mg cytochrome/ kg starting material	Purity Ratio $E^{C3+}_{410}/E^{C3+}_{280}$
Amberlite CG-50 CM-50 Sephadex Ammonium Sulphate Fractionation	1.5	0.4
CM-52 Cellulose, Ionic Gradient	0.4	3.9
Amberlite MB-1	0.3	3.8

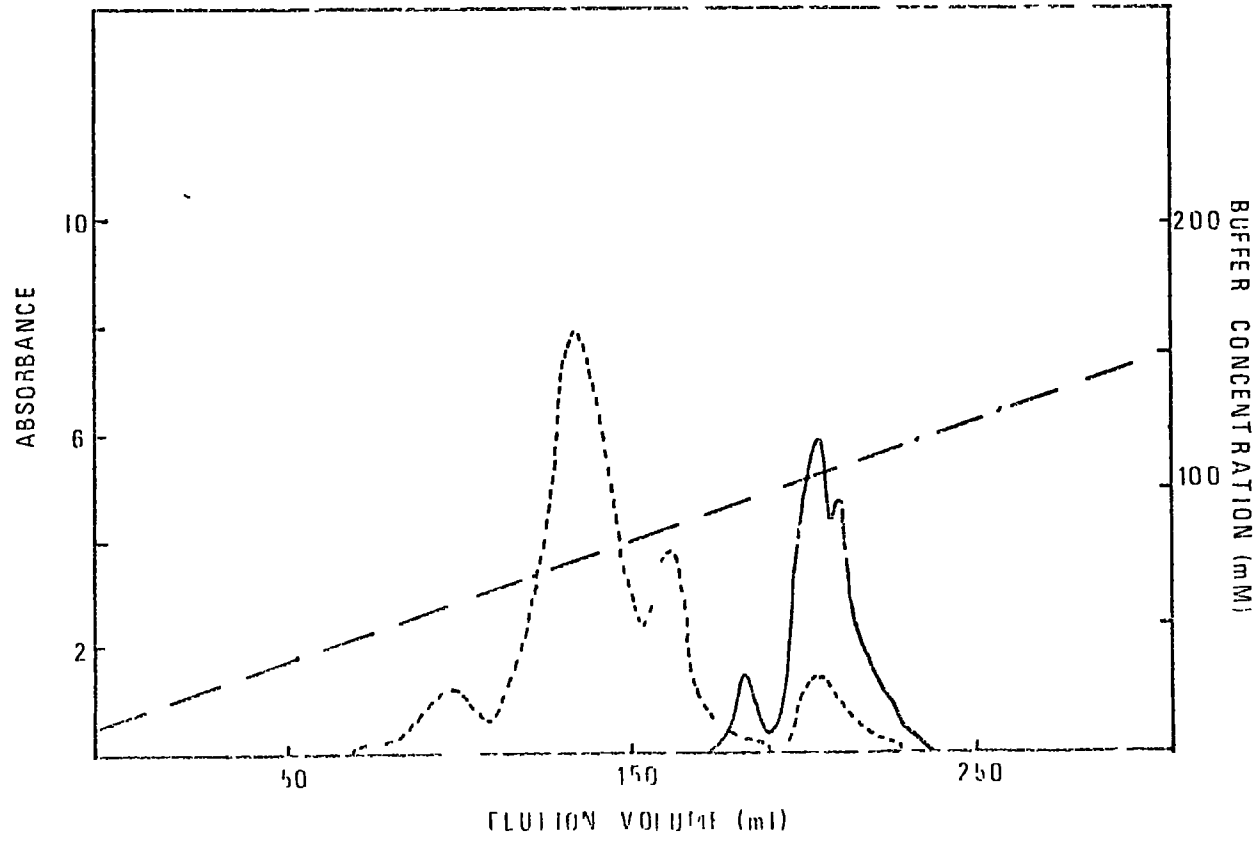
TABLE 20.

The Spectral Ratios of Purified *Loligo* Cytochrome *c*

	Ratio
$E^{C2+}_{550}/E^{C3+}_{280}$	0.9
$E^{C3+}_{410}/E^{C3+}_{280}$	3.8
$E^{C2+}_{416}/E^{C3+}_{410}$	1.1
$E^{C2+}_{416}/E^{C2+}_{550}$	4.6

The mass was 3.2 mg salt-free.


ABSORBANCE
10
6
2
50
150
250
ELUTION VOLUME (ml)
200
100
BUFFER CONCENTRATION (mM)


ABSORBANCE
10
6
2
50
150
250
ELUTION VOLUME (ml)
200
100
BUFFER CONCENTRATION (mM)

followed, although the animals proved resistant to blending. The extract was filtered after stirring in the normal way and treated to the usual Amberlite, dialysis and CM-50 Sephadex purification steps, but no cytochrome c was observed at any stage. The purification was terminated at this point due to the lack of material.

8. The Amino Acid Sequence of Locusta Cytochrome c

The amino acid sequence of Locusta cytochrome c was determined from the evidence of two chymotryptic digestions, one tryptic digestion and data obtained from the use of a Beckman Automatic Sequencer, using a total of 3 μmol (36 mg) of protein. The sequence is shown in Fig. 31 giving the points of enzyme cleavage together with the overlapping peptides from which the sequence was deduced. A list of chymotryptic peptides together with electrophoretic mobility and sequence data is given in Table 21 and a similar list of tryptic peptides is given in Table 22. The amino acid composition was obtained from three duplicate 50 μg samples hydrolysed for 24, 48 and 72 h respectively, and this is shown in Table 23.

Digestion

The oxidised protein readily denatured in 80% ethanol at room temperature and was adequately digested by both trypsin and chymotrypsin.

FIGURE 31.

The Amino Acid Sequence of Locusta Cytochrome c .

Residues which were identified by dansyl-Edman analysis are indicated by ——→; those identified by other means, e.g. by amino acid composition data, are indicated by --→; those identified by digestion with carboxypeptidase-A are indicated by ←——; arrows ⟹ indicate that the C-terminal residue was identified as the free amino acid. The arrows ↓↑ indicate points of complete enzymic cleavage, upwards for trypsin and downwards for chymotrypsin; ⇣⇡ indicates points of partial cleavage, and ⇓ indicates points of cleavage using chymotrypsin to digest the heme peptides. Residues -4 to 37, excluding 14 and 17, were determined using an automatic sequencer.

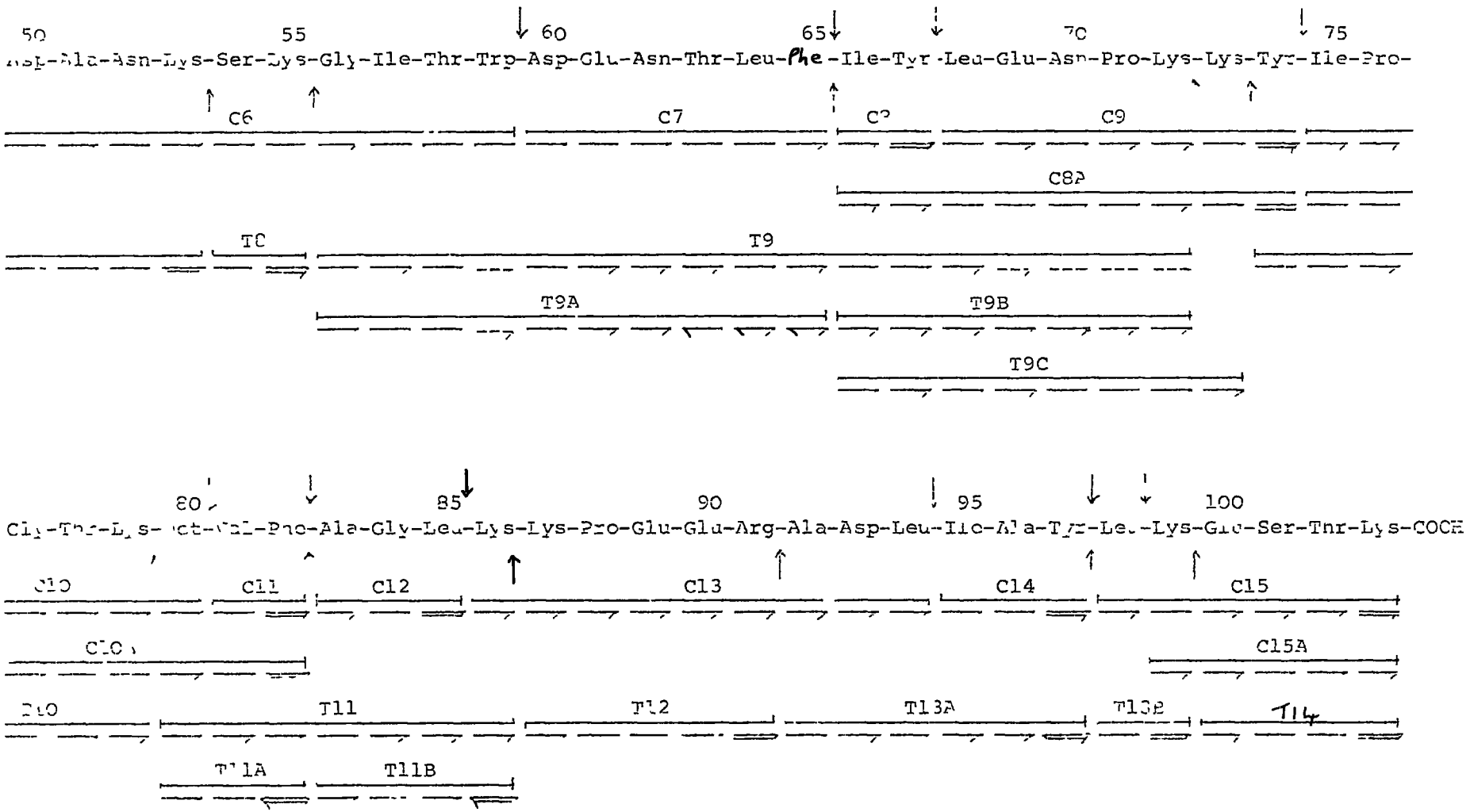

50
55
60
65
70
75
Asp-Ala-Asn-Lys-Ser-Lys-Gly-Ile-Thr-Trp-Asp-Glu-Asn-Thr-Leu-Phe-Ile-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr-Ile-Pro-
C6
C7
C8
C9
C8A
T8
T9
T9A
T9B
T9C
80
85
90
95
100
Gly-Thr-Lys-Met-Val-Pro-Ala-Gly-Leu-Lys-Lys-Pro-Glu-Glu-Arg-Ala-Asp-Leu-Ile-Ala-Tyr-Leu-Lys-Glu-Ser-Thr-Lys-COOH
C10
C11
C12
C13
C14
C15
C10A
C15A
T10
T11
T12
T13A
T13B
T14
T11A
T11B

TABLE 21.

Chymotryptic Peptides from Locusta Cytochrome c

Peptide/ Position	M (pH 6.5)	M (pH 1.9)	RDNS-ARG BAWP	Dansyl-Edman Results
C1 (-4-+10	0.50	0.90		Gly-Val-Pro-Glx-Gly-Asx-Val-Glx-Lys-Gly-Lys-Lys-Ile-Phe
C2 (11-26)	0	0.54	0.16	Val- (Heme peptide; see text)
C2CA (11-18)		0.32		Val-Glx-Arg-$CySO_3$-Ala-Glx-$CySO_3$-His
C2CB (19-26)	0.15	0.85	0.05	Thr-Val-Glx-Ala-Gly-Gly-Lys-His
C3 (27-36)	0.95	0.75		Lys-Thr-Gly-Pro-Asx-Leu-(His, Gly)-Leu-Phe
C3A (27-33)	1.48	1.01		Lys-Thr-Gly-Pro-Asx-Leu-His
C3B (34-36)	0	0.57	0.92	Gly-Leu-Phe
C4 (37-46)	1.21	0.85		Gly-Arg-Lys-Thr-Gly-Glx-Ala-Pro-Gly-Phe
C5 (47-48)	0	0.62	0.58	Ser-Tyr
C6 (49-59)	0.53	0.77		Thr-Asx-Ala-Asx-Lys-Ser-Lys-Gly-Ile-Thr-Trp
C7 (60-65)	-1.57	0.40		Asx-Glx-Asx-Thr-Leu-Phe
C8 (66-67)	0	0.57	0.57	Ile-Tyr
C8A (66-74)	0.81	0.76		Ile-Tyr-Leu-Glx-Asx-Pro-Lys-Lys-Tyr
C9 (68-74)	0.79	0.93		Leu-Glx-Asx-Pro-Lys-Lys-Tyr
C10 (75-80)	0.91	0.81		Ile-Pro-Gly-Thr-Lys-Met

TABLE 21 (Cont'd.)

C10A (75-82)	0.82	0.58		Ile-Pro-Gly-Thr-Lys-Met-Val-Phe
C11 (81-82)	0	0.62	0.98	Val-Phe
C12 (83-85)	0	0.57	0.77	Ala-Gly-Leu
C13 (86-94)	0	0.94		Lys-Lys-Pro-Glx-Glx-Arg-Ala-Asx-Leu
C14 (95-97)	0	0.57	0.94	Ile-Ala-Tyr
C15 (98-103)	0.85	1.08		Leu-Lys-Glx-Ser-Thr-Lys
C15A (99-103)	0.94	1.08		Lys-Glx-Ser-Thr-Lys

TABLE 22.

Tryptic Peptides from Locusta Cytochrome c

Peptide/ Position	M (pH 6.5)	M (pH 1.9)	RDNS-ARG BAWP	Dansyl-Edman Results
T1 (-4-+5)	-0.67	0.60		Gly-Val-Pro-Glx-Gly-Asx-Val-Glx-Lys
T1A (-4-+7)	0	0.69		Gly-Val-Pro-Glx-Gly-Asp-Val-Glx-Lys-Gly-Lys
T2 (6-7)	1.90	1.56		Gly-Lys
T3 (9-12)	0	0.74		Ile-Phe-Val-Glx-Arg
T4 (14-25)	-0.20	0.36		(Heme peptide; see text)
T4CA (14-18)		0.31		$CySO_3$-Ala-Glx-$CySO_3$-His
T4CB (19-25)		0.72		Thr-Val-Glx-Ala-Gly-Gly-Lys
T5 (26-27)	2.32	1.76		His-Lys
T6 (28-38)	0.84	0.62		Thr-Gly-Pro-Asx-Leu-His-Gly-Leu-Phe-Gly-Arg
T6A (28-39)	1.36	0.81		Thr-Gly-Pro-Asx-Leu-His-Gly-(Leu,Phe,Gly,Arg,Lys)
T7 (40-53)	0	0.42		Thr-Gly-Glx-Ala-Pro-Gly-Phe-Ser-Tyr-Thr-Asx-Ala-Asx-Lys
T8 (54-55)	1.86	1.37		Ser-Lys
T9 (56-73)	0	0.37		Gly-Ile-Thr-(Trp)-Asx-Glx-Asx-Thr-Leu-Phe-Ile-Tyr-Leu-(Glx, Asx,Pro,Lys,Lys)
T9A (56-65)	-0.90	0.27		Gly-Ile-Thr-(Trp)-Asx-Glx-Asx-Thr-Leu-Phe

TABLE 22 (cont'd.)

T9B (66-72)	0	0.67	Ile-Tyr-Leu-Glx-Asx-Pro-Lys
T9C (66-73)	0.68	0.84	Ile-Tyr-Leu-Glx-Asx-Pro-Lys-Lys
T10 (74-79)	0.87	0.76	Tyr-Ile-Pro-Gly-Thr-Lys
T11 (80-86)	0.79	0.54	Met-Val-Phe-Ala-Gly-Leu-Lys
T11A (80-82)	0	0.55	Met-Val-Phe
T11B (83-86)	1.32	1.02	Ala-Gly-Leu-Lys
T12 (87-91)	0	1.12	Lys-Pro-Glx-Glx-Arg
T13 (92-99)	Not isolated		
T13A (92-97)	-0.85	0.35	Ala-Asx-Leu-Ile-Ala-Tyr
T13B (98-99)	1.74	1.37	Leu-Lys
T14 (100-103)	0	0.92	Glx-Ser-Thr-Lys

TABLE 23.

The Amino Acid Composition of Locusta Cytochrome c

	Mean values 24 h hydrolysis	Mean values 48 h hydrolysis	Mean values 72 h hydrolysis	Mean Corrected Values	Amino Acid Analysis	Sequence Values
Asp	7.9	8.2	8.6	8.2	8	8
Thr	7.8	7.4	7.4	8.3*	8	8
Ser	4.0	3.8	3.7	4.2*	4	3
Glu	9.6	10.8	10.0	10.1	10	11
Pro	6.6	6.1	6.5	6.4	6	6
Gly	12.1	12.3	13.8	12.6	13	13
Ala	7.0	7.6	7.7	7.4	7	7
Cysteine	2.1	2.8	1.8	2.2	2	2
Val	4.6	5.2	4.9	4.8	5	5
Met	0.7	0.5	0.6	0.6	1	1
Ile	4.4	4.8	4.9	4.7	5	5
Leu	7.1	6.4	6.6	6.7	7	7
Tyr	3.3	4.3	4.1	3.9	4	4
Phe	4.1	4.3	4.1	4.2	4	5
His	2.8	2.8	2.8	2.8	3	3
Lys	17.2	17.0	15.1	16.4	16	15
Arg	3.3	3.2	3.4	3.3	3	3
Trp	-	-	-	-	+	1

The mean values were obtained from the analysis of 3 duplicate samples of 50 µg of cytochrome c at each hydrolysis time. The mean corrected values were obtained as an average of the six determinations with corrections made for the destruction of certain amino acids.

\+ Trp was not determined; the best spectral ratios indicate one residue is present.

* Calculated from 24 h and 72 h values assuming first order kinetics for destruction (Moore & Stein, 1963).

The chymotryptic digestions were performed on a 10 mg and a 12 mg sample of oxidised and denatured cytochrome c equilibrated at pH 8.0 and 37°C, under nitrogen. In both cases 2% (w/w) enzyme was added at zero time, a further 2% after 60 min and the digestion was terminated after 100 min. The tryptic digestion was performed on 10 mg of oxidised and denatured protein. 2% enzyme (w/w) was added at zero time, a further 2% after 80 min and the digestion was terminated after 100 min. The automatic sequencer data was obtained using 4 mg of oxidised cytochrome c denatured in 70% (v/v) formic acid.

Chymotryptic Peptides

Peptide C1 (-4-+10) (Gly-Val-Pro-Gln-Gly-Asp-Val-Glu-Gly-Lys-Lys-Ile-Phe)

Digestion with carboxypeptidase-A at 37°C and pH 8.0 for 3 h yielded phenylalanine as judged by dansyl analysis. Similar analysis after a 9 h digestion yielded the dansyl derivatives of isoleucine and phenylalanine. The pH 6.5 electrophoretic mobility indicated the presence of two acidic residues, and these were placed at positions 2 and 4 from evidence obtained from the automatic sequencer.

Peptide C2 (11-26) (Heme peptide)

Valine was shown to be the N-terminal amino acid of the heme peptide, which was dehemed in two ways. The first chymotryptic heme peptide was treated after the manner of

Ambler (1963) and the second chymotryptic heme peptide was mildly oxidised using performic acid as previously described. The product of both methods was digested with 5% chymotrypsin at pH 8.0 and 37°C for 30 min to yield peptides C2CA and C2CB which were separated on electrophoresis at pH 1.9.

Peptide C2CA (11-18) (Val-Gln-Arg-$CySO_3$-Ala-Gln-$CySO_3$-His)

The peptide gave a positive reaction with the phenanthraquinone reagent and arginine was placed at position-13 from the dansyl-Edman analysis. Digestion with carboxypeptidase-A for 3 h yielded bis-dansyl-histidine after dansyl analysis, as did dansylation without hydrolysis after seven Edman degradation steps. Residues 12 and 16 were both placed as glutamine from the data obtained using the automatic sequencer. Residue-12 could also be placed from the pH 6.5 electrophoretic mobility of peptide T3, although no such determinations could be made using the mobility data of C2CA because of the presence of histidine and cysteic acid.

Peptide C2CB (19-26) (Thr-Val-Glu-Ala-Gly-Gly-Lys-His)

Digestion with carboxypeptidase-A for 3 h yielded bis-dansyl-histidine after dansyl analysis as did dansylation without hydrolysis after seven Edman degradation steps. A sample of the peptide had a mobility of 0.15 on pH 6.5 electrophoresis which indicated that residue-21 was glutamic acid, assuming that histidine carried a charge of less than +1

at that pH. This result was confirmed by the data obtained from the automatic sequencer.

Peptide C3 (27-36 (Lys-Thr-Gly-Pro-Asn-Leu-His-Gly-Leu-Phe)

The dansyl-Edman analysis was inconclusive beyond five degradation steps and residues-33 and 34 were placed from the semi-quantitative amino acid composition data of C3 and the sequence data of C3A. Residues-35 and 36 were placed from dansyl analysis of carboxypeptidase-A digestions for 3 and 24 h, together with the sequence evidence for C3B. A semi-quantitative amino acid composition of C3 after six degradation steps contained bis-dansyl-histidine, but this was absent in a similar analysis after a further degradation step. Histidine and glycine were not identified in the N-terminal analysis following these two degradations. A consideration of the pH 6.5 electrophoretic mobility indicated that residue-31 was asparagine, assuming that histidine carried a charge of less than +1 at that pH. This conclusion was supported by the evidence from the automatic sequencer.

Peptide C3A (27-33) (Lys-Thr-Gly-Pro-Asn-Leu-His)

Digestion with carboxypeptidase-A for 3 h, followed by dansyl analysis, yielded bis-dansyl-histidine, as did dansylation without hydrolysis after six Edman degradation steps. The pH 6.5 electrophoretic mobility indicated that residue-31 was asparagine, assuming that histidine carried a charge of less than +1 at that pH. This conclusion was

supported by the evidence of the automatic sequencer data.

Peptide C3B (34-36) (Gly-Leu-Phe)

Dansylation without hydrolysis after two Edman degradation steps yielded dansyl-phenylalanine.

Peptide C4 (37-46 (Gly-Arg-Lys-Thr-Gly-Gln-Ala-Pro-Gly-Phe)

The peptide gave a positive result with the phenanthiaquinone reagent and arginine was placed at position-38 from the dansyl-Edman analysis. Dansylation without hydrolysis after nine Edman degradation steps yielded dansyl-phenylalanine. The pH 6.5 electrophoretic mobility was lower than expected, but indicated that residue 42 was an amide.

Peptide C5 (47-48) (Ser-Tyr)

Dansylation without hydrolysis after a single Edman degradation step yielded bis-dansyl-tyrosine.

Peptide C6 (49-59) (Thr-Asp-Ala-Asn-Lys-Ser-Lys-Gly-Ile-Thr-Trp)

The peptide gave a positive reaction with the Ehrlich reagent and showed a pink coloration during the trifluoroacetic acid stage of the first Edman degradation, indicating the presence of tryptophan (Uphaus et al., 1959). Digestion, with carboxypeptidase-A for 3 h followed by dansyl analysis, yielded dansyl-tryptophan, and a similar analysis after a 9 h digestion yielded the dansyl derivatives of tryptophan and threonine. The pH 6.5 electrophoretic mobility of C6 was lower than expected, possibly due to the presence of tryptophan,

but still indicated the presence of a single acidic residue. A sample of the peptide after two Edman degradation steps gave two ninhydrin positive zones on pH 6.5 electrophoresis at mobilities 0.62 and 1.15. This was interpreted to be due to the incomplete blocking of the ε-amino function of the two lysine residues by phenylthiocarbamyl groups as a result of their exposure to PITC. The mobility data was consistent with residue-50 being aspartic acid.

Peptide C7 (60-65) (Asp-Glu-Asn-Thr-Leu-Phe)

Digestion with carboxypeptidase-A for 3 and 9 h yielded phenylalanine and phenylalanine and leucine respectively, as judged by dansyl analysis.

The pH 6.5 electrophoretic mobility of the intact peptide indicated that two acidic residues were present and samples taken for pH 6.5 electrophoresis after one and two Edman degradation steps gave mobilities of -1.09 and zero respectively. This data was consistent with residues-60 and 61 being aspartic acid and glutamic acid respectively, and with residue-62 being asparagine.

Peptide C8 (66-67) (Ile-Tyr)

Dansylation without hydrolysis after a single Edman degradation step yielded bis-dansyl-tyrosine.

Peptide 8A (66-74) (Ile-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr)

Dansylation without hydrolysis after eight Edman degradations yielded bis-dansyl-tyrosine. The pH 6.5

electrophoretic mobility was greater than expected but indicated that a single acidic residue was present. This was placed from the evidence of C9 at position-69.

Peptide C9 (68-74) (Leu-Glu-Asn-Lys-Lys-Tyr)

Dansylation without hydrolysis following six Edman degradation steps yielded bis-dansyl-tyrosine. The pH 6.5 electrophoretic mobility of the intact peptide indicated the presence of one acidic residue and this was placed from the mobility evidence of the partially degraded peptide. After two Edman degradation steps, the peptide gave three ninhydrin positive zones on pH 6.5 electrophoresis of mobilities zero, 0.95 and 1.90 and this was interpreted as being due to the incomplete blocking of the ε-amino function of the two lysine residues by phenylthiocarbamyl groups due to their exposure to PITC. The evidence was consistent with residues-69 and 70 being glutamic acid and asparagine respectively.

Peptide C10 (75-80) (Ile-Pro-Gly-Thr-Lys-Met)

The peptide gave a pink coloration during the trifluoroacetic acid stage of the first Edman degradation step, but no other evidence for the presence of tryptophan could be found (see Uphaus et al., 1959).

Peptide C10A (75-82) (Ile-Pro-Gly-Thr-Lys-Met-Val-Phe)

Digestion with carboxypeptidase-A for 3 h followed by dansyl analysis yielded dansyl-phenylalanine together with some dansyl-valine. Dansylation without hydrolysis after

seven Edman degradation steps yielded dansyl-phenylalanine.

Peptide C11 (81-82) (Val-Phe)

Dansylation without hydrolysis after a single Edman degradation step yielded dansyl-phenylalanine.

Peptide C12 (83-85) (Ala-Gly-Leu)

Dansylation without hydrolysis after a single Edman degradation step yielded dansyl-leucine.

Peptide C13 (86-91) (Lys-Lys-Pro-Glu-Glu-Arg-Ala-Asp-Leu)

The peptide gave a positive reaction with the phenanthraquinone reagent and arginine was placed as residue-91 from the dansyl-Edman analysis. The pH 6.5 electrophoretic mobility indicated that residues-89, 90 and 93 were all acidic.

Peptide C14 (95-97) (Ile-Ala-Tyr)

Dansylation without hydrolysis after two Edman degradation steps yielded bis-dansyl-tyrosine.

Peptide C15 (98-103) (Leu-Lys-Glu-Ser-Thr-Lys)

Dansylation without hydrolysis after five Edman degradation steps yielded α-dansyl-ε-PTC lysine. The pH 6.5 electrophoretic mobility of C15 indicated that residue-100 was glutamic acid.

Peptide C15A (99-103) (Lys-Glu-Ser-Thr-Lys)

The pH 6.5 electrophoretic mobility was consistent with residue-100 being glutamic acid.

Tryptic Peptides

Peptide T1 (-4-+5) (Gly-Val-Pro-Gln-Gly-Asp-Val-Glu-Lys)

The pH 6.5 electrophoretic mobility indicated that two acidic residues were present and these were placed. from the data obtained from the automatic sequencer, at positions 2 and 4. Glutamine was positively identified at position -1 during the automatic sequencer analysis.

Peptide T1A (-4-+7) (Gly-Val-Pro-Gln-Gly-Asp-Val-Glu-Lys-Gly-Lys)

Dansylation without hydrolysis after ten Edman degradation steps yielded α-dansyl-ε-PTC-lysine. The pH 6.5 electrophoretic mobility of T1A was consistent with the automatic sequence data regarding the presence of two acidic residues.

Peptide T2 (6-7) (Gly-Lys)

Dansylation without hydrolysis after a single Edman degradation step yielded bis-dansyl-lysine.

Peptide T3 (9-12) (Ile-Phe-Val-Gln-Arg)

The peptide gave a positive result with the phenanthraquinone reagent and arginine was placed in position 12 from the results of the dansyl-Edman analysis, and dansylation without hydrolysis after four Edman degradation steps.

Peptide T4 (14-25) (Heme peptide)

The heme peptide underwent a mild performic oxidation and was then digested with 5% (w/w) chymotrypsin at pH 8.0

and 37°C for 30 min, yielding peptides T4CA and T4CB which were purified by pH 1.9 electrophoresis.

Peptide T4CA (14-18) ($CySO_3$-Ala-Gln-$CySO_3$-His)

Dansylation without hydrolysis after four Edman degradation steps yielded bis-dansyl-histidine. The presence of cysteic acid and histidine residues made the placing of amide residues from electrophoretic mobility data impossible, but residue-16 was fixed as glutamine from a positive identification during the automatic sequencer analysis.

Peptide T4CB (19-25) (Thr-Val-Glu-Ala-Gly-Gly-Lys)

The dansyl-Edman analysis was inconclusive beyond five degradation steps but dansylation without hydrolysis after the sixth degradation gave α-dansyl-ε-PTC-lysine despite the lack of a positive result from the N-terminal analysis at this point. Insufficient material made the pH 6.5 electrophoretic mobility determination of T4CB impossible so that residue-21 was placed as glutamic acid from the evidence of peptide C2CB and the automatic sequencer data.

Peptide T5 (26-27) (His-Lys)

Histidine was identified at residue-26 as its α-dansyl-derivative on N-terminal analysis of the intact peptide. The semi-quantitative amino acid composition of the intact peptide contained histidine as judged by the dansyl analysis of the hydrolysate, but this was absent in a similar analysis following a single Edman degradation step. Dansylation

without hydrolysis at this point yielded bis-dansyl-lysine.

Peptide T6 (28-38) (Thr-Gly-Pro-Asn-Leu-His-Gly-Leu-Phe-Gly-Arg)

The peptide gave a positive reaction with the phenanthraquinone reagent and arginine was placed in position 38 from the dansyl-Edman analysis, and dansylation without hydrolysis after ten Edman degradation steps. Histidine was identified as its α-dansyl derivative on N-terminal analysis after the fifth Edman degradation step. It was also fixed in position 33 by the presence of histidine in a semi-quantitative amino acid composition conducted after five degradation steps and the absence after six Edman degradations. The pH 6.5 electrophoretic mobility was consistent with the occurrence of asparagine in position 31, assuming that histidine carried a charge of less than +1 in that pH.

A peptide of identical sequence was isolated with a mobility of 0.20 on pH 6.5 electrophoresis. It was assumed that this was a deamidated form of T6 rather than a real alternative to the sequence.

Peptide T6A (28-39) (Thr-Gly-Pro-Asn-Leu-His-Gly-Leu-Phe-Gly-Arg-Lys)

The dansyl-Edman evidence was inconclusive beyond six Edman degradation steps and residues 35-39 were placed from amino acid composition data, the pH 1.9 electrophoretic mobility data which was consistent with a change of +3 for T6A at this pH and the sequence analysis of T6. The pH 6.5

electrophoretic mobility indicated that residue-31 was asparagine assuming that histidine carried a charge of less than +1 at that pH.

Peptide T7 (40-53) (Thr-Gly-Gln-Ala-Pro-Gly-Phe-Ser-Tyr-Thr-Asp-Ala-Asn-Lys)

Dansylation without hydrolysis after thirteen Edman degradation steps yielded α-dansyl-ε-PTC-lysine. The pH 6.5 electrophoretic mobility indicated the presence of one acidic residue and this was placed as aspartic acid-50 from the mobility evidence of the intact and partially degraded forms of peptide C6. Residue-42 was also placed as glutamine from the evidence of peptide C4.

Peptide T8 (54-55) (Ser-Lys)

Dansylation without hydrolysis after a single Edman degradation step yielded bis-dansyl-lysine.

Peptide T9 (56-73) (Gly-Ile-Thr-Trp-Asp-Glu-Asn-Thr-Leu-Phe-Ile-Tyr-Leu-Glu-Asn-Pro-Lys-Lys)

The peptide gave a positive reaction with the Ehrlich reagent and showed a rose-pink coloration during the trifluoroacetic acid stage of the first Edman degradation step, indicative of tryptophan (Uphaus et al., 1959). However, tryptophan was not identified by N-terminal analysis after three Edman degradation steps due to the acid destruction of the dansyl derivatives, and it was therefore placed from peptide C6. The dansyl-Edman analysis was inconclusive

beyond thirteen Edman degradation steps and residues 69-73 were placed from the semi-quantitative amino acid composition data of T9, and the sequence analysis of T9B. Both electrophoretic mobilities were poor, possibly due to a combination of the presence of tryptophan and the molecular weight in excess of 2,000, and this made the fixing of acidic residues difficult. Positions 60, 61 and 69 were placed as acidic residues and 62 and 70 as amides from the mobility evidence of peptides T9A, T9B, C7 and C9.

Peptide T9A (56-65) (Gly-Ile-Thr-Trp-Asp-Glu-Asn-Thr-Leu-Phe)

The peptide gave a positive reaction with the Ehrlich reagent and gave a pink coloration during the trifluoroacetic acid stage of the first Edman degradation step, indicative of tryptophan (Uphaus et al., 1959), but tryptophan was not identified on N-terminal analysis after three degradation steps because of the acid destruction of dansyl-tryptophan. Digestion with carboxypeptidase-A for three hours yielded dansyl-phenylalanine, whilst a similar analysis after 9 h of digestion yielded the dansyl derivatives of phenylalanine, leucine and threonine. The pH 6.5 mobility was less acidic than expected, possibly due to the presence of tryptophan, but it still indicated a charge of -2 consistent with the mobility evidence of the intact peptide C7. Positions 60 and 61 were fixed as aspartic acid and glutamic acid, and position 62 as asparagine from the mobility evidence of

partially degraded samples of C7.

Peptide T9B (66-72) (Ile-Tyr-Leu-Glu-Asn-Pro-Lys)

Residues 66 and 70 were placed as glutamic acid and asparagine respectively, from the evidence of C9.

Peptide T9C (66-73) (Ile-Tyr-Leu-Glu-Asn-Pro-Lys-Lys)

The pH 1.9 electrophoretic mobility indicated the presence of the two lysine residues and the pH 6.5 mobility was consistent with the placing of residues 69 and 70 as glutamic acid and asparagine as outlined in the evidence for C9.

Peptide T10 (74-79) (Tyr-Ile-Pro-Gly-Thr-Lys)

Peptide T11 (80-86) (Met-Val-Phe-Ala-Gly-Leu-Lys)

Peptide T11A (80-82) (Met-Val-Phe)

Digestion with carboxypeptidase-A for 3 h yielded the dansyl derivative of phenylalanine together with some valine after dansyl analysis, whilst dansylation without hydrolysis after two Edman degradation steps yielded dansyl-phenylalanine.

Peptide T11B (83-86) (Ala-Gly-Leu-Lys)

Dansylation without hydrolysis after three Edman degradation steps yielded α-dansyl-ε-PTC-lysine.

Peptide T12 (87-91) (Lys-Pro-Glu-Glu-Arg)

The peptide gave a positive reaction with the phenanthraquinone reagent and arginine was fixed in position 91 from the dansyl-Edman analysis and dansylation without

hydrolysis after four Edman degradation steps. The pH 6.5 electrophoretic mobility indicated that residues 89 and 90 were both glutamic acid and samples of the peptide after two and three Edman degradations had pH 6.5 mobilities of -1.25 and zero. This data was consistent with the mobility evidence from C13.

Peptide T13 (92-99)

The expected tryptic peptide T13 was not isolated.

Peptide T13A (92-97) (Ala-Asp-Leu-Ile-Ala-Tyr)

Digestion with carboxypeptidase A for 2 h yielded bis-dansyl-tyrosine after dansyl analysis as did dansylation without hydrolysis after five Edman degradation steps. The pH 6.5 electrophoretic mobility indicated that residue-93 was aspartic acid.

Peptide T13B (98-99) (Leu-Lys)

Dansylation without hydrolysis after a single Edman degradation step yielded bis-dansyl-lysine.

Peptide T14 (100-103) (Glu-Ser-Thr-Lys)

Dansylation without hydrolysis after three Edman degradation steps yielded α-dansyl-ε-PTC-lysine. The pH 6.5 electrophoretic mobility indicated that residue -100 was glutamic acid.

Automatic Sequence Analysis

```
 -4                1                                                  10
Gly-Val-Pro-Gln-Gly-Asp-Val-Glu-Lys-Gly-Lys-Lys-Ile-Phe-

                                                  20
Val-Gln-Arg- * -Ala-Gln- * -His-Thr-Val-Glu-Ala-Gly-Gly-

                          30
Lys-His-Lys-Thr-Gly-Pro-Asn-Leu-His-Gly-Leu-Phe-Gly
```

Residues -4-+37 were determined using a Beckman 890c sequencer operating on the quadrol fast protein double cleavage programme.

The glutamic acid residues at positions 4 and 21, the glutamines at -1, 12 and 16, the aspartic acid at 2 and the asparagine at position 31, were identified as their PTH-derivatives by thin-layer chromatography (TLC) and by gas chromatography (GLC) after "on column" trimethylsilylation using a 200-270°C, 6° per min temperature programme. The isoleucine residue at position 9 and the leucine residues at positions 32 and 34 were identified as their PTH-derivatives on TLC, and from GLC of the trimethylsilyl derivatives formed "on column" using an isothermal development at 200°C. The glycine residues at positions 29 and 37 and the proline at 30 were identified as their PTH-derivatives on TLC and direct GLC, together with the dansyl analysis of the regenerated parent amino acid. The lysine residues in positions 5, 7, 8, 25 and 27 were identified as their PTH-derivatives on TLC and by dansyl-analysis of the regenerated parent amino acid. The PTH-derivatives of arginine and histidine were not soluble

in ethylacetate but remained in the aqueous layer following the acid conversion of the 2-anilino-5-thiazolinone derivatives to the PTH-derivatives and their subsequent extraction with ethyl acetate. Arginine was fixed in position 13 from the regeneration of such an aqueous phase followed by dansyl analysis, together with a spot test using the phenonthraquinone reagent. The aqueous phases for positions 18, 26 and 33, where histidine was suspected, were split into two parts. The parent amino acid was regenerated normally from the first part, whilst the second part was taken to pH 12.5 with ammonia before extraction with ethyl acetate. The unprotonated histidine was more soluble under these conditions, and the ethyl acetate extract could be regenerated in the normal manner. Both methods yielded bis-dansyl-histidine after dansyl analysis. The cysteine residues at 14 and 17 were not identified as such, and were placed from the manual sequence analysis. The cytochrome c was not dehemed before analysis, so that the cysteine remained bound covalently to the heme after the eighteenth degradation, and the two cysteine residues were released bound to the heme after the twenty first degradation step. The analysis of the eighteenth and twenty first fractions was inconclusive, but the latter did exhibit a pink coloration due to the presence of the heme-cysteine complex.

The remaining residues fixed by the automatic sequencer were identified as their PTH derivatives on TLC, and direct

GLC using a temperature programme of 160-270°C at 8° per min.

Amino Acid Sequence

The overlapping chymotryptic + tryptic peptides gave the sequence of Locusta cytochrome c as shown in Figure 31. The sequence data agreed with the amino acid composition data shown in Table 23 with a number of exceptions. The values obtained for serine and lysine were routinely high in all determinations, but no evidence for an impurity responsible for such discrepancies could be seen in the sequence analysis. Tryptophan was not determined as part of the amino acid composition, but the best spectral ratios of the purified Locusta cytochrome c indicated that one residue was present.

From Figure 31 it can be seen that all the residues were positively identified in both chymotryptic and tryptic digestions except for those in positions 8, 39 and 59. These could be placed from positive identifications by automatic sequence analysis for positions 8 and 39, or by the sequence analysis in only one of the digestions, from amino acid composition data or a combination of these last two. The lysine residues at positions 8 and 39 were not identified in the tryptic digest because of the nature of the enzyme cleavage at X-Lys-Lys-X and X-Arg-Lys-X sequences, and tryptophan was not identified in the appropriate tryptic peptide because of the destruction of the dansyl derivative during the acid hydrolysis stage of dansyl analysis.

Residues -4 to +37 were positively identified, with the exception of 14 and 17, using the automated sequence data. No residues were placed without positive identification from at least one digestion.

All positive overlaps between chymotryptic and tryptic peptides were observed except in the region C14, C15/T13A, T13B, where an anomalous tryptic break duplicated the chymotryptic digestion. However, the ordering of peptides is clear from the overlaps obtained at C13, C14/T13A and C15, C15A/T13B, T14.

All the acidic and amide residues indicated in the sequence were placed, with the exception of positions -1, +2, 4, 16, 21 and 31, from the pH 6.5 electrophoretic mobilities of intact or partially degraded peptides taken from either, or both, of the chymotryptic and tryptic digestions. The residues in positions -1, +2, 4, 21 and 31 were placed from a combination of the above, together with determinations associated with the automatic sequence analysis. Glutamine-16 was placed from the automatic sequence data alone.

The observed electrophoretic alternative to peptide T6 was assumed to be due to deamidation during preparation rather than to represent a real alternative to the sequence.

The observed enzyme specificities were consistent with those expected (Smyth, 1967), except that only partial chymotryptic cleavage was observed at tyrosine-67. In addition,

full chymotryptic cleavage was observed at histidine-26, leucine-85 and leucine-94, and partial chymotryptic cleavage at histidine-33, methionine-80 and leucine-98. Partial tryptic cleavage was observed at lysine-5, arginine-38 and lysine-72, together with phenylalanine-65 and phenylalanine-82. Full tryptic cleavage was observed at tyrosine-97.

9. The Amino Acid Sequence of Macrobrachium Cytochrome c

The amino acid sequence of Macrobrachium cytochrome c was determined from the evidence of two chymotryptic digestions and two tryptic digestions using a total of 2 μmol (24 mgs) of protein. The sequence is given in Figure 32 showing the points of enzyme cleavage together with the overlapping peptides from which it was deduced. A list of chymotryptic peptides, together with mobility and sequence data, is given in Table 24, and a similar list of tryptic peptides is given in Table 25. The amino acid composition of Macrobrachium cytochrome c was obtained from three duplicate 50 μg samples hydrolysed for 24, 48 and 72 hours respectively, and this is given in Table 26. The amino acid composition of certain peptides is given in Table 27.

Digestion

The oxidised protein denatured easily in 80% ethanol and was adequately digested by both trypsin and chymotrypsin.

FIGURE 32.

The Amino Acid Sequence of Macrobrachium Cytochrome c.

Residues which were identified by dansyl-Edman analysis are indicated by ——→; those identified by other means, e.g. by amino acid composition data, are indicated by --→; those identified by digestion with carboxypeptidase-A are indicated by ←——; arrows ⟹ indicate that the C-terminal residue was identified as the free amino acid. The arrows ↓↑ indicate points of complete enzymic cleavage, upwards for trypsin and downwards for chymotrypsin; ⇣⇡ indicate points of partial cleavage, and ⇓ indicates points of cleavage using chymotrypsin to digest the heme peptides. The amino acid composition of peptides marked * is given in Table 27.

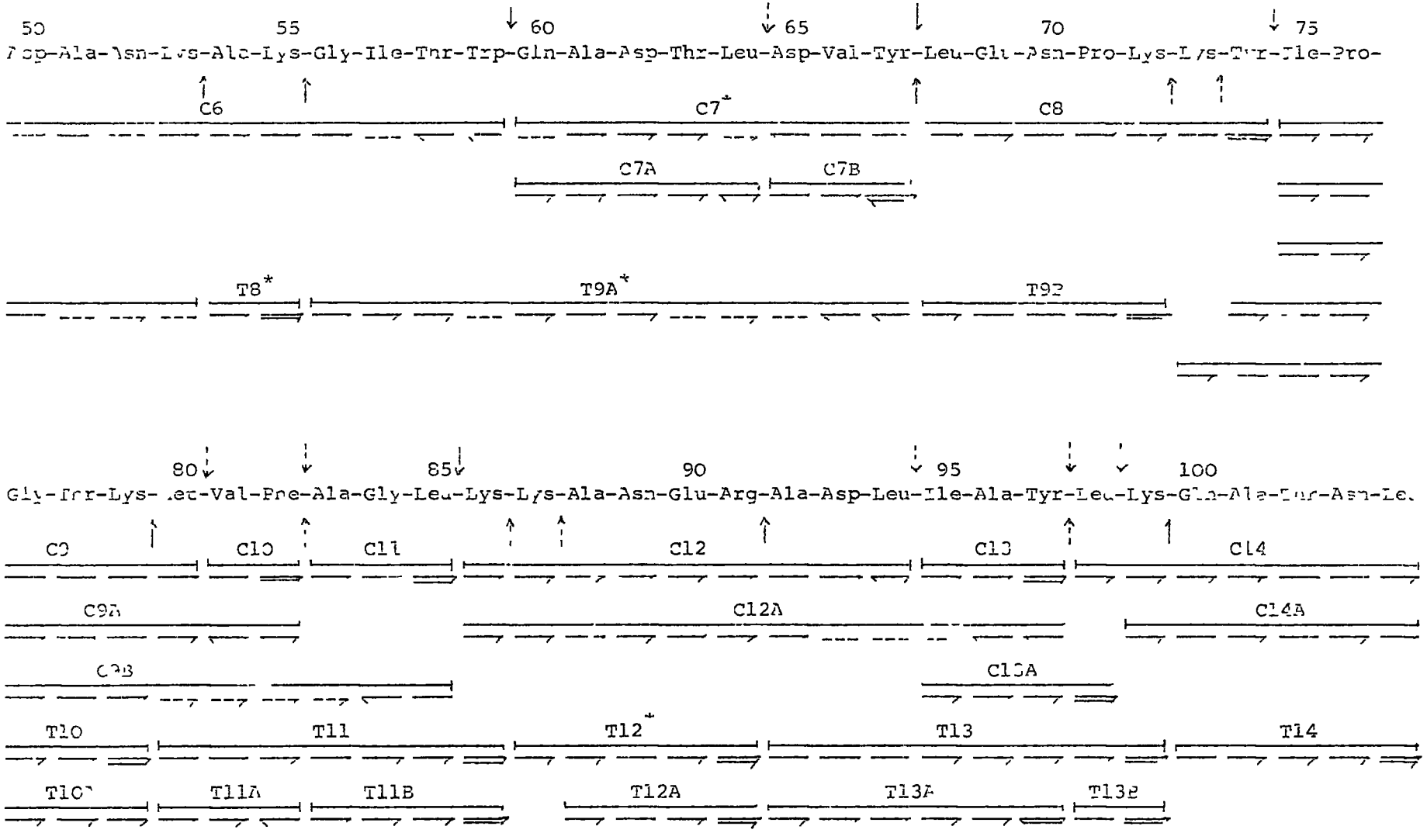

50
55
60
65
70
75
Asp-Ala-Asn-Lys-Ala-Lys-Gly-Ile-Thr-Trp-Gln-Ala-Asp-Thr-Leu-Asp-Val-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr-Ile-Pro-
C6
C7*
C8
C7A
C7B
T8*
T9A*
T9B
80
85
90
95
100
Gly-Thr-Lys-Met-Val-Phe-Ala-Gly-Leu-Lys-Lys-Ala-Asn-Glu-Arg-Ala-Asp-Leu-Ile-Ala-Tyr-Leu-Lys-Gln-Ala-Thr-Asn-Leu
C9
C10
C11
C12
C13
C14
C9A
C12A
C14A
C9B
C13A
T10
T11
T12*
T13
T14
T10A
T11A
T11B
T12A
T13A
T13B

TABLE 24

Chymotryptic peptides from Macrobrachium cytochrome c

Peptide/ position	M (pH 6.5)	M (pH 1.9)	RDNS-ARG BAWP	Dansyl-Edman Results
C1 (1-10)	0	0.80		(X-Gly,Asx,Val,Glx,Lys,Gly, Lys)-Lys-Ile-Phe
C2 (11-26)	0.20	0.39	0.14	Val- (Heme peptide; see text)
C2CA (11-18)		0.28		Val-Glx-Arg-$CySO_3$-(Ala,Glx, $CySO_3$)-His
C2CB (19-26)	1.10	1.15		Ser-Ala-Glx-Ala-Asx (Leu,Lys) His
C2CC (19-24)	0	0.46		Ser-Ala-Glx-Ala-Asx-Leu
C2CD (25-26)		1.76		Lys-His
C3 (27-36)	0.67	0.53	0.39	Lys-Thr-Gly-Pro-Asx-Leu-Asx-Gly-Leu-Phe
C3A (27-32)	0.95	0.77		Lys-Thr-Gly-Pro-Asx-Leu
C3B (33-36)	0	0.54	0.11	Asx-Gly-Leu-Phe
C3C (27-33)	0.88	0.72		Lys-Thr-Gly-Pro-Asx-Leu-Asx
C3D (34-36)	0	0.55	0.85	Gly-Leu-Phe
C4 (37-46)	0.65	0.53	0.03	Gly-Arg-Glx-Thr-Gly-Glx-Ala-Ser-Gly-Tyr
C4A (37-48)	0	0.47		Gly-Arg-Glx-Thr-Gly-Glx-Ala-Ser-Gly-Tyr-Val-Tyr
C5 (47-48)	0	0.59	0.89	Val-Tyr
C6 (49-59)	0.65	0.79		Thr-Asx-Ala-Asx-Lys-Ala-Lys-Gly-(Ile)-Thr-Trp

TABLE 24 (cont'd.)

C7 (60-67)	1.53	0.39		(Glx)-Ala-Asx-Thr-(Leu)-Asx-Val-Tyr
C7A (60-64)	-1.92	0.44		Glx-Ala-Asx-Thr-Leu-Asx-Val-Tyr
C7B (65-67)	-1.19	0.45		Asx-Val-Tyr
C8 (68-74)	0.20	0.95		Leu-Glx-Asx-Pro-Lys-Lys-Tyr
C9 (75-80)	0.99	0.82	0.21	Ile-Pro-Gly-Thr-Lys-Met
C9A (75-82)	0.75	0.61		Ile-Pro-Gly-Thr-Lys-Met-Val-Phe
C9B (75-85)	0.69	0.57		Ile-Pro-Gly-Thr-Lys-(Met, Val,Phe,Ala)-Gly-Leu
C10 (81-82)	0	0.67	0.91	Val-Phe
C11 (83-85)	0	0.67	0.70	Ala-Gly-Leu
C12 (86-94)	0.65	1.09		Lys-Lys-Ala-Asx-Glx-Arg-Ala-Asx-Leu
C12A (86-97)	0.65	0.94	0.04	Lys-Lys-Ala-Asx-Glx-Arg-Ala-(Asx,Leu,Ile)-Ala-Tyr
C13 (95-97)	0	0.54	0.91	Ile-Ala-Tyr
C13A (95-98)	0	0.46		Ile-Ala-Tyr-Leu
C14 (98-104)	0.75	0.65		Leu-Lys-Glx-Ala-Thr-Asx-Leu
C14A (99-104)	0.94	0.69	0.09	Lys-Glx-Ala-Thr-Asx-Leu

TABLE 25.

Tryptic Peptides from Macrobrachium Cytochrome c

Peptide/ Position	M (pH 6.5)	M (pH 1.9)	RDNS-ARG BAWP	Dansyl-Edman Results
T1 (1-5)	-2.00	0.40		(X-Glx,Asx,Val,Glx)-Lys
T2 (6-7)	1.93	1.45	0.17	Gly-Lys
T2A (6-8)	2.73	1.61	0.07	Gly-Lys-Lys
T3 (9-13)	0.86	0.73	0.55	Ile-Phe-Val-Glx-Arg
T4 (14-25)	0.29	0.41		(Heme peptide; see text)
T4CA (14-18)		0.38		$CySO_3$-Ala-Glx-($CySO_3$)-His
T4CB (19-25)	0.82	0.67		Ser-Ala-Glx-Ala-Asx-(Leu,Lys)
T5 (26-27)	2.38	1.60	0.08	His-Lys
T6 (28-38)	0.69	0.53		Thr-Gly-Pro-Asx-Leu-Asx-Gly-Leu-Phe-Gly-Arg
T7 (39-53)	0	0.43		Glx-Thr-Gly-Glx-Ala-Ser-Gly-Tyr-Val-Tyr-Thr-Asx (Ala,Asx, Lys)
T8 (54-55)	2.08	1.45	0.24	Ala-Lys
T9 (56-73)	Not isolated.			
T9A (56-67)	-0.90	0.15		Gly-Ile-Thr-(Trp)-Glx-Ala-Asx-(Thr,Leu,Asx)-Val-Tyr
T9B (68-72)	0	0.90		Leu-Glx-Asx-Pro-Lys

TABLE 25 (cont'd.)

T10 (74-79)	0.91	0.72	0.45	Tyr-Ile-Pro-Gly-Thr-Lys
T10A (73-79)	1.52	1.02	0.26	Lys-Tyr-Ile-Pro-Gly-Thr-Lys
T11 (80-86)	0.74	0.71	0.75	Met-Val-Phe-Ala-Gly-Leu-Lys
T11A (80-82)	0	0.53		Met-Val-Phe
T11B (83-86)	1.30	1.04	0.34	Ala-Gly-Leu-Lys
T12 (87-91)	0.99	1.10	0.13	Lys-Ala-Asx-Glx-Arg
T12A (88-91)	0	0.82		Ala-Asx-Glx-Arg
T13 (92-99)	0	0.70		Ala-Asx-Leu-Ile-Ala-Tyr-Leu-Lys
T13A (92-97)	-0.86	0.44		Ala-Asx-Leu-Ile-Ala-Tyr
T13B (98-99)	1.70	1.40		Leu-Lys
T14 (100-104)	0	0.49		Glx-Ala-Thr-Asx-Leu

TABLE 26.

The Amino Acid Composition of *Macrobrachium* cytochrome *c*

	Mean values 24 h hydrolysis	Mean values 48 h hydrolysis	Mean values 72 h hydrolysis	Mean Corrected Values	Amino Acid Analysis	Sequence Values
Asp	10.7	11.5	11.0	11.0	11	12
Thr	6.0	5.7	5.8	6.2*	6	7
Ser	2.6	2.2	2.3	2.9*	3	2
Glu	9.9	10.1	10.2	10.2	10	10
Pro	4.1	4.1	4.0	4.0	4	3
Gly	10.5	10.6	10.7	10.7	11	10
Ala	10.9	10.5	10.8	10.8	11	12
Cysteine	1.8	2.2	2.0	2.0	2	2
Val	4.8	5.0	5.0	5.0	5	5
Met	0.9	1.4	1.0	1.0	1	1
Ile	3.4	3.5	3.6	3.6	4	4
Leu	8.6	8.5	8.7	8.7	9	9
Tyr	3.6	3.7	3.5	3.5	4	4
Phe	2.7	3.1	2.8	2.8	3	4
His	1.6	1.8	1.8	1.8	2	2
Lys	16.0	16.7	16.3	16.3	16	13
Arg	3.7	3.8	3.6	3.6	4	3
Trp	-	-	-	-	+	1

The mean values were obtained from the analyses of 2 samples of 50 µg of cytochrome *c* at each hydrolisation. The mean corrected values were obtained as an average of the six determinations with corrections made for the destruction of certain amino acids.

* Calculated from 24 h and 72 h values assuming first order kinetics for destruction (Moore & Stein, 1963).

\+ Trp was not determined; the best spectral ratios of the cytochrome indicated the presence of one residue.

TABLE 27.

The Amino Acid Composition of Chymotryptic and Tryptic Peptides from *Macrobrachium*.

	Peptide C1	Peptide C7	Peptide T1	Peptide T6	Peptide T7	Peptide T8	Peptide T9A	Peptide T12
Asp	1.3(1)	2.4(2)	1.1(1)	2.1(2)	2.1(2)		2.4(2)	1.4(1)
Thr		0.7(1)		0.6(1)	2.1(2)		1.7(2)	
Ser					1.2(1)			
Glu	0.8(1)	0.8(1)	1.0(1)		2.0(2)		1.1(1)	0.9(1)
Pro				1.0(1)				
Gly	2.3(2)		1.1(1)	2.8(3)	2.2(2)		0.7(1)	
Ala		1.2(1)			2.0(2)	1.0(1)	1.2(1)	0.6(1)
Cysteine								
Val	0.7(1)	0.9(1)	0.8(1)		1.0(1)		0.8(1)	
Met								
Ile	0.8(1)						0.9(1)	
Leu		1.3(1)		1.7(2)			1.3(1)	
Tyr						1.6(2)		
Phe	0.8(1)	0.9(1)		0.8(1)			0.6(1)	
His								
Lys	3.3(3)		0.8(1)		1.0(1)	0.7(1)		1.4(1)
Arg				0.9(1)				0.6(1)
Trp							+(1)	

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

+Tryptophan was not determined.

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

Six mg of cytochrome c was oxidised and denatured, and equilibrated at 37°C and pH 8.0 for both of the chymotryptic digestions. 2% (w/w) enzyme was added at zero time, a further 2% after 50 min and the digestion was terminated after 90 min. Six mg of protein was used for both of the tryptic digestions. After denaturation and equilibration at 37°C and pH 8.0, 2% (w/w) trypsin was added at zero time, a further 2% at 80 min and the digestion was terminated after 120 min.

Chymotryptic Peptides

Peptide C1 (1-10) (Gly-Asp-Val-Glu-Lys-Gly-Lys-Lys-Ile-Phe)

The peptide failed to give an N-terminal result with the dansyl method and would not undergo any degradation with the Edman method, indicating that the peptide had a blocked N-terminus. Digestion with carboxypeptidase-A at 37°C and pH 8.5 for 3 h yielded dansyl-phenylalanine after dansyl analysis, whilst 24 h of digestion yielded quantities of isoleucine and phenylalanine together with a significant amount of lysine. Duplicate samples of C1 were taken and hydrolysed for 24 h to obtain quantitative amino acid composition data for the peptide. The following mean values were obtained: Asp (1.3), Glu (0.8), Gly (2.3), Val (0.7), Ile (0.8), Phe (0.8), and Lys (3.3). These were in accord with the proposed sequence of peptide C1. Residues 5-10 can be more directly placed from the sequence evidence of peptides T2

and T3 and the C-terminal determinations of peptide T1. Residues 1-4 were placed from the quantitative amino acid composition data of C1 and T1 together with the evidence of the corresponding residues of cytochromes c published to date (Dayhoff, 1972).

Acetyl determinations on samples of peptide C1 proved inconclusive.

Peptide C2 (11-26) (Heme peptide)

The heme peptide was purified by electrophoresis and paper chromatography. Valine was shown to be the N-terminal residue using the dansyl method. After a mild performic oxidation, the peptide was digested with 5% chymotrypsin (w/w) at pH 8.0 and 37°C to yield peptides C2CA, C2CB, C2CC and C2CD which were separated by electrophoresis at pH 1.9.

Peptide C2CA (11-18) (Val-Gln-Arg-$CySO_3$-Ala-Gln-$CySO_3$-His)

The peptide gave a positive result with both Pauly and phenanthraquinone reagents, and digestion with carboxypeptidase-A followed by dansyl analysis indicated that histidine was the C-terminal amino acid. The dansyl-Edman analysis was inconclusive beyond four degradation steps so that residues 15-17 were placed from T4CA and amino acid composition data. Residue-12 was fixed as glutamine from the pH 6.5 electrophoretic mobility of peptide T3, and a comparison with the mobility of the equivalent chymotryptic peptide in Locusta cytochrome c indicated that position 16 was occupied by

glutamine. The method of Offord (1966) for the placement of amide residues from electrophoretic mobilities was not applicable, due to the presence of histidine and cysteic acid residues.

Peptide C2CB (19-26) (Ser-Ala-Gln-Ala-Asn-Leu-Lys-His)

The peptide gave a positive reaction with the Pauly reagent and digestions with carboxypeptidase-A for 3 h and 9 h followed by dansyl analysis confirmed histidine as the C-terminal residue. Dansyl-Edman analysis was inconclusive beyond five degradations and residues-24 and 25 were placed from the semi-quantitative amino acid composition data of C2CB and from the sequence analysis of C2CC. The mobility on pH 6.5 electrophoresis indicated that residues-21 and 23 were amides. A peptide with an identical sequence to C2B was isolated in low yield having a zero mobility on pH 6.5 electrophoresis. It was assumed to have been deamidated during preparation, rather than to be a real alternative form of the sequence.

Peptide C2CC (19-26) (Ser-Ala-Gln-Ala-Asn-Leu)

Dansylation without hydrolysis after five Edman degradation steps confirmed leucine as the C-terminal residue. A consideration of the pH 6.5 electrophoretic mobility placed residues-21 and 23 as amides.

Peptide C2CD (27-28) (Lys-His)

Dansylation without hydrolysis after one Edman

degradation step yielded bis-dansyl-histidine.

Peptide C3 (27-36) (Lys-Thr-Gly-Pro-Asn-Leu-Asn-Gly-Leu-Phe)

Digestion with carboxypeptidase-A for 3 h followed by dansylation yielded an excess of dansyl-phenyalanine, whilst a similar 9 h digestion and analysis yielded excesses of phenylalanine and leucine together with a significant quantity of glycine. The pH 6.5 electrophoretic mobility indicated that residues-31 and 33 were amides.

Peptide C3A (27-32) (Lys-Thr-Gly-Pro-Asn-Leu)

Dansylation without hydrolysis after five Edman degradations confirmed leucine as the C-terminal residue. Residue-31 was fixed as asparagine from the pH 6.5 electrophoretic mobility.

Peptide C3B (33-36) (Asn-Gly-Leu-Phe)

Digestion with carboxypeptidase-A for 3 h followed by dansyl analysis yielded dansyl-phenylalanine together with some dansyl-leucine, and dansylation without hydrolysis after three Edman degradations confirmed phenylalanine as the C-terminus of C3B. The pH 6.5 electrophoretic mobility indicated that residue-33 was asparagine.

Peptide C3C (27-33) (Lys-Thr-Gly-Pro-Asn-Leu-Asn)

Dansylation without hydrolysis after six Edman degradations yielded dansyl derivatives which co-chromatographed with dansyl-asparagine and dansyl-aspartic acid on polyamide

layers. The pH 6.5 electrophoretic mobility indicated that residues-31 and 33 were amides.

Peptide C3D (34-36) (Gly-Leu-Phe)

Dansylation without hydrolysis after two Edman degradation steps confirmed phenylalanine as the C-terminal residue.

Peptide C4 (37-46) (Gly-Arg-Gln-Thr-Gly-Gln-Ala-Ser-Gly-Tyr)

The peptide gave a positive reaction with the phenanthraquinone reagent, indicating the presence of arginine, and this was confirmed at position 38 by dansyl-Edman analysis. Digestion with carboxypeptidase-A for 3 h followed by dansyl analysis yielded an excess of bis-dansyl-tyrosine together with some dansyl-glycine. Dansyl analysis after 9 h digestion yielded similar quantities of the dansyl derivatives of tyrosine, glycine and serine as judged by an assessment of the fluorescence of these dansyl derivatives on polyamide sheets. The electrophoretic mobility at pH 6.5 of C4 indicated that residues-39 and 42 were amides.

Peptide C4A (37-48) (Gly-Arg-Gln-Thr-Gly-Gln-Ala-Ser-Gly-Tyr-Val-Tyr)

Dansylation without hydrolysis after eleven Edman degradation steps confirmed tyrosine as the C-terminal residue. From a comparison of the pH 6.5 electrophoretic mobilities of C4 and C4A it was assumed that C4CA was a deamidated form of C4 and as such, a preparation artefact, rather than a real

alternative from the sequence.

Peptide C5 (47-48) (Val-Tyr)

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

Peptide C6 (49-59) (Thr-Asp-Ala-Asn-Lys-Ala-Lys-Gly-Ile-Thr-Trp)

The peptide gave a positive result with the Ehrlich reagent and digestion with carboxypeptidase for 3 h followed by dansyl analysis yielded dansyl-tryptophan, whilst a similar analysis after 9 h of digestion yielded quantities of dansyl-threonine and dansyl-tryptophan. The dansyl-Edman analysis was inconclusive beyond eight degradations, but a semi-quantitative amino acid composition taken at this point showed significant amounts of isoleucine and threonine. The pH 6.5 electrophoretic mobility indicated the presence of one acidic residue and this was placed as aspartic acid-50 from a consideration of the pH 6.5 mobility after two Edman degradations. After this particular electrophoresis, three ninhydrin positive zones were detected with mobilities of 0, 0.66 and 1.24 and these were assumed to represent degrees of blocking of the lysine ε-amino groups in the peptide by the phenylthiocarbamyl group as a result of their exposure to PITC during two Edman degradations. The three mobility values were consistent with residue-50 being aspartic acid.

Peptide C7 (60-67) (Gln-Ala-Asp-Thr-Leu-Asp-Val-Tyr)

The peptide gave a positive result with the ninhydrin reagent but failed to give an N-terminal result after the normal dansyl analysis and would not undergo any degradation with the Edman method, indicating that the peptide had no free N-terminal group. Digestion with carboxypeptidase-A for 3 h followed by dansyl analysis yielded an excess of bis-dansyl-tyrosine, whilst digestion for 6 h gave dansyl-valine and dansyl-tyrosine. A 24 h digest yielded equivalent amounts of dansyl-valine and dansyl-tyrosine after dansyl analysis together with a quantity of dansyl-aspartic acid, as judged by their fluorescence on polyamide sheets under u.v. light. Analysis of duplicate samples of C7 on a Locarte amino acid analyser gave the following values: Asp (2.4), Thr (0.7), Glu (0.8), Ala (1.2), Val (0.9), Leu (1.3), Tyr (0.9). Digestion of the peptide for 30 min with pyrrolidonyl peptidase prepared from Pseudomonas fluorescens, followed by protein precipitation with absolute ethanol, gave a peptide in the supernatant with alanine as the N-terminus and containing tyrosine as shown by dansyl analysis. Dansyl-Edman analysis of this new peptide was inconclusive beyond two degradation steps which placed residues-62 and 63. It was concluded from the evidence that the loss of a free N-terminal group had occurred after the peptide had been eluted from the pH 1.9 electrophoresis paper, and that this loss was due to the

cyclisation of glutamine to pyrrolidone-carboxylic acid. The pH 1.9 electrophoretic mobility of 0.30 indicated a charge of +1 which would be due to the N-terminal amino group, and the electrophoretic mobility of pH 6.5 indicated the presence of two acidic residues. Aspartic acid-65 could be independently placed from C7B and residue-60 was placed as glutamine from the observed susceptibility to cyclisation and the subsequent affinity for pyrrolidonyl-peptidase. Residue-62 was therefore placed as aspartic acid and this was compatible with the evidence of C7A.

Peptide C7A (60-64) (Gln-Ala-Asp-Thr-Leu)

Digestion of the peptide with carboxypeptidase-A followed by dansyl analysis gave dansyl-leucine after 3 h and quantities of the dansyl-derivatives of leucine, threonine and aspartic acid after 24 h. The pH 6.5 electrophoretic mobility indicated the presence of two acid residues but residue-60 was fixed as an amide from the evidence outlined in C7. It was assumed that this peptide represented an example of deamidation at residue-60 rather than an alternative form of the sequence.

Peptide C7B (65-67) (Asp-Val-Tyr)

Digestion with carboxypeptidase-A for 3 h yielded bis-dansyl-tyrosine after dansyl analysis, and tyrosine was confirmed as the C-terminal residue by dansylation without hydrolysis after two Edman degradation steps. The pH 6.5

electrophoretic mobility placed residue-C5 as aspartic acid.

Peptide C8 (68-74) (Leu-Glu-Asn-Pro-Lys-Lys-Tyr)

Digestion with carboxypeptidase-A for 3 h yielded bis-dansyl-tyrosine after dansyl analysis and tyrosine was confirmed as the C-terminal residue by dansylation after six Edman degradation steps. A consideration of the pH 6.5 electrophoretic mobility indicated the presence of one amide residue and this was placed at position 70 from the pH 6.5 mobility after two Edman degradation steps.

Peptide C9 (75-80) (Ile-Pro-Gly-Thr-Lys-Met)

Peptide C9A (75-82) (Ile-Pro-Gly-Thr-Lys-Met-Val-Phe)

Digestion with carboxypeptidase-A for 3 and 24 h followed by dansyl analysis yielded dansyl-phenylalanine and dansyl-phenylalanine and dansyl-valine respectively.

Peptide C9B (75-85) (Ile-Pro-Gly-Thr-Lys-Met-Val-Phe-Ala-Gly-Leu)

Digestion with carboxypeptidase-A for 3 and 9 h yielded dansyl-leucine and the dansyl derivatives of leucine and glycine respectively after dansyl analysis. The dansyl-Edman analysis was inconclusive beyond five degradation steps and residues 80-83 were placed from composition data and the sequences of C9, C10 and C11.

Peptide C10 (81-82) (Val-Phe)

Dansylation without hydrolysis after a single Edman degradation step confirmed phenylalanine as the C-terminal residue.

Peptide C11 (83-85) (Ala-Gly-Leu)

Dansylation without hydrolysis after two Edman degradations yielded dansyl-leucine.

Peptide C12 (86-94) (Lys-Lys-Ala-Asn-Glu-Arg-Ala-Asp-Leu)

The peptide gave a positive result with the phenanthraquinone reagent and digestion for 3 h with carboxypeptidase-A followed by dansyl-analysis indicated that leucine was the C-terminal residue. The pH 6.5 electrophoretic mobility indicated the presence of two acidic residues and these were placed as glutamic acid-90 and aspartic acid-93 from the tryptic peptides T12 and T13 respectively.

Peptide C12A (86-97) (Lys-Lys-Ala-Asn-Glu-Arg-Ala-Asp-Leu-Ile-Ala-Tyr)

The peptide gave a positive reaction with the phenanthraquinone reagent and digestions with carboxypeptidase-A for 3 and 9 h gave bis-dansyl-tyrosine, and bis-dansyl-tyrosine together with dansyl-alanine respectively, after dansyl analysis. The dansyl-Edman analysis was inconclusive beyond seven degradation steps and residues 93-95 were placed from the semi-quantitative amino acid analysis of C12A together with the sequence evidence for peptides C12 and C13. The pH 6.5 electrophoretic mobility was consistent with peptide C12 with regard to the placement of the acidic residues at positions 90 and 93.

Peptide C13 (95-97) (Ile-Ala-Tyr)

Dansylation without hydrolysis after two dansyl-Edman degradation steps confirmed tyrosine as the C-terminal residue.

Peptide C13A (95-98) (Ile-Ala-Tyr-Leu)

Dansylation without hydrolysis after three dansyl-Edman degradations yielded dansyl-leucine.

Peptide C14 (98-104) (Leu-Lys-Gln-Ala-Thr-Asn-Leu)

Digestion with carboxypeptidase-A for 3 h followed by dansyl analysis yielded dansyl-leucine, whilst similar analysis after 9 h of digestion gave quantities of dansyl-leucine and dansyl-asparagine. The pH 6.5 electrophoretic mobility of C14 indicated that residues-100 and 103 were both amides. A peptide of identical sequence to C14 was isolated from the neutral zone of a pH 6.5 electrophoresis, but this was placed as a deamidated form of C14 rather than a real sequence alternative.

Peptide C14A (99-104) (Lys-Gln-Ala-Thr-Asn-Leu)

The pH 6.5 electrophoretic mobility indicated that residues-100 and 103 were both amides. A peptide of identical sequence but with a neutral pH 6.5 mobility was isolated and attributed to deamidation in a similar fashion to C14.

Tryptic Peptides

Peptide T1 (1-5) (Gly-Asp-Val-Glu-Lys)

The peptide gave a positive result with the ninnydrin

reagent but failed to react with PITC or dansyl-chloride in the Edman and dansyl procedures, indicating a blocked N-terminus. Digestion for 24 h with carboxypeptidase-A followed by dansyl analysis yielded a small amount of bis-dansyl-lysine.

A quantitative amino acid analysis conducted on T1 gave the following values. Asp (1.1), Glu (1.0), Gly (1.1), Val (0.8), Lys (0.8). Residues 1-4 were placed from the quantitative amino acid compositions of T1 and C1 together with a consideration of the corresponding residues in the cytochromes c published to date (see Dayhoff, 1972). The pH 6.5 electrophoretic mobility of T1 indicated that residues 2 and 4 were both acidic amino acids.

Acetyl determinations on samples of peptide T1 proved inconclusive.

Peptide T2 (6-7) (Gly-Lys)

Dansylation without hydrolysis after one Edman degradation yielded bis-dansyl-lysine.

Peptide T2A (6-8) (Gly-Lys-Lys)

Dansylation without hydrolysis after a single Edman degradation yielded a fluorescent zone which did not co-chromatography with bis-dansyl-lysine but which had similar Rf in all three solvent systems. This was assumed to be bis-dansyl-lysyl-ε-dansyl-lysine. Dansylation without hydrolysis after two Edman degradation steps yielded bis-dansyl-lysine.

Peptide T3 (9-13) (Ile-Phe-Val-Gln-Arg)

The peptide gave a positive reaction with the phenanthraquinone reagent and dansylation without hydrolysis after four Edman degradation steps yielded dansyl-arginine. The pH 6.5 electrophoretic mobility of T3 placed residue-12 as glutamine.

Peptide T4 (14-25) (Heme peptide)

The heme peptide was purified by electrophoresis, dehemed by means of a mild performic oxidation and re-digested with 5% chymotrypsin at pH 8.0 and 37°C for 30 min. The products of this digestion were separated by electrophoresis at pH 1.9.

Peptide T4CA (14-18) ($CySO_3$-Ala-Gln-$CySO_3$-His)

Digestion with carboxypeptidase for 3 h yielded bis-dansyl-histidine after dansyl analysis. The dansyl-Edman analysis was unclear beyond three degradation steps, but a semi-quantitative amino acid composition of the peptide residue at this point contained cysteic acid and histidine. Residue-17 was placed from composition data and glutamine-16 was fixed from the evidence of C2CA. The pH 6.5 electrophoretic mobility could not be used to place amide residues due to the presence of cysteic acid and histidine.

Peptide T4CB (19-25) (Ser-Ala-Gln-Ala-Asn-Leu-Lys)

The dansyl-Edman analysis was unclear beyond four degradation steps and residues 23-25 were placed from composition

data and the evidence of peptides C2CC and C2CD. A sample of the peptide had a mobility of 0.82 on pH 6.5 electrophoresis indicating that residues-21 and 23 were amides.

Peptide T5 (26-27) (His-Lys)

The peptide gave a positive reaction with the Pauly reagent and an amino acid composition containing histidine and lysine. The N-terminal residue was observed as α-dansyl-histidine after dansylation and hydrolysis of the peptide, and dansylation without hydrolysis after a single Edman degradation yielded bis-dansyl-lysine. The amino acid composition at this point contained no histidine.

Peptide T6 (28-38) (Thr-Gly-Pro-Asn-Leu-Asn-Gly-Leu-Phe-Gly-Arg)

The peptide gave a positive reaction with the phenanthraquinone reagent and dansylation without hydrolysis after ten Edman degradation steps gave dansyl-arginine. A quantitative amino acid analysis on a sample of T6 gave the following values: Asp (2.1), Thr (0.6), Pro (1.0), Gly (2.8), Leu (1.7), Phe (0.8), Arg (0.9). The pH 6.5 electrophoretic mobility of T6 indicated that residues-31 and 33 were amides, although a peptide of identical sequence was isolated from the neutral zone of a pH 6.5 electrophoresis. This example was assumed to be a deamidated form of T6 rather than a real alternative to the sequence.

Peptide T7 (39-53) (Gln-Thr-Gly-Gln-Ala-Ser-Gly-Tyr-Val-Tyr-Thr-Asp-Ala-Asn-Lys)

A quantitative amino acid composition of the peptide gave the following values: Asp (2.1), Thr (2.1), Ser (1.2), Glu (2.0), Gly (2.2), Ala (2.0), Val (1.0), Tyr (1.6), Lys (1.0). The dansyl-Edman analysis was unclear beyond twelve degradation steps and residues 51-53 were placed from the composition data and from the sequence of peptide T6. The pH 6.5 electrophoretic mobility indicated the presence of a single acidic residue. Residues-39 and 42 were placed as amides from peptide C4 and residues-50 and 52 as aspartic acid and asparagine respectively from the evidence of peptide C6.

Peptide T8 (54-55) (Ala-Lys)

A quantitative amino acid composition gave the following values: Ala (1.0) and Lys (0.7). Dansylation without hydrolysis after a single Edman degradation step gave bis-dansyl-lysine.

Peptide T9 (56-72)

The expected peptide was not isolated.

Peptide T9A (56-67) (Gly-Ile-Thr-Trp-Gln-Ala-Asp-Thr-Leu-Asp-Val-Tyr)

The peptide gave a positive reaction with the Ehrlich reagent and showed a rose-pink coloration during the trifluoro-acetic acid stage of the first Edman degradation step, indicative of tryptophan (Uphaus et al., 1959). A quantitative amino acid composition gave the following values:Asp (2.4),

Thr (1.7), Glu (1.1), Gly (0.7), Ala (1.2), Val (0.8), Ile (0.9), Leu (1.3), Tyr (0.6). Tryptophan was not determined. Digestion with carboxypeptidase-A for 3 and 9 h gave bis-dansyl-tyrosine and bis-dansyl-tyrosine together with dansyl-valine respectively, following dansyl analysis. The dansyl-Edman analysis was inconclusive beyond seven degradation steps and residues 63-67 were placed from the amino acid composition data and the sequence of peptides C8, C8A and C8B.

Peptide T9B (68-72) (Leu-Glu-Asn-Pro-Lys)

Dansylation without hydrolysis after four Edman degradation steps yielded bis-dansyl-lysine. The pH 6.5 electrophoretic mobilities of the purified peptide and the peptide after two degradation steps indicated that residue-69 was glutamic acid and residue-70 was asparagine. The degraded peptide gave two ninhydrin positive spots with mobilities equal to zero and 1.45 as a result of only a proportion of the ε-amino group of lysine having reacted with PITC.

Peptide T10 (74-79) (Tyr-Ile-Pro-Gly-Thr-Lys)

Dansylation without hydrolysis after five Edman degradation steps yielded α-dansyl-ε-PTC-lysine.

Peptide T10A (73-79) (Lys-Tyr-Ile-Pro-Gly-Thr-Lys)

Dansylation without hydrolysis after six Edman degradations yielded α-dansyl-ε-PTC-lysine.

Peptide T11 (80-86) (Met-Val-Phe-Ala-Gly-Leu-Lys)

Dansylation without hydrolysis after six Edman degradations yielded α-dansyl-ε-PTC-lysine.

Peptide T11A (80-82) (Met-Val-Phe)

Digestion with carboxypeptidase-A for 3 h followed by dansyl analysis yielded dansyl-phenylalanine whilst both phenylalanine and valine were present as their dansyl derivatives after similar analysis of a 9 h digestion. Dansylation without hydrolysis after two Edman degradation steps yielded dansyl-phenylalanine.

Peptide T11B (83-86) (Ala-Gly-Leu-Lys)

Dansylation without hydrolysis after three Edman degradation steps yielded α-dansyl-ε-PTC-lysine.

Peptide T12 (87-91) (Lys-Ala-Asn-Glu-Arg)

The peptide gave a positive result with the phenanthraquinone reagent, and dansylation without hydrolysis after four Edman degradation steps fixed arginine as residue-91 and the C-terminus of T12. A quantitative amino acid composition gave the following values: Asp (1.4), Glu (0.9), Ala (0.6), Lys (1.4), Arg (0.6). The peptide had a 6.5 electrophoretic mobility of 0.99 indicating the presence of one acid and one amide residue. The pH 6.5 mobility of the peptide after three Edman degradations was zero indicating that residue-90 was glutamic acid, and residue-89 was accordingly fixed as asparagine.

Peptide T12A (88-91) (Ala-Asn-Glu-Arg)

The peptide gave a positive reaction with the phenanthraquinone reagent and dansylation without hydrolysis after three Edman degradation steps yielded dansyl-arginine. The pH 6.5 mobility indicated that the peptide contained one acidic residue and this was placed at residue-90 from T12.

Peptide T13 (92-99) (Ala-Asp-Leu-Ile-Ala-Tyr-Leu-Lys)

Dansylation without hydrolysis after seven Edman degradations yielded α-dansyl-ε-PTC-lysine and consideration of the zero pH 6.5 electrophoretic mobility fixed residue-93 as aspartic acid.

Peptide T13A (92-97) (Ala-Asp-Leu-Ile-Ala-Tyr)

Digestion with carboxypeptidase-A for 3 h followed by dansyl analysis yielded bis-dansyl-tyrosine and dansyl-alanine, whilst dansylation without hydrolysis after five Edman degradations gave bis-dansyl-tyrosine. The pH 6.5 electrophoretic mobility of - 0.86 fixed residue-93 as aspartic acid.

Peptide T13B (98-99) (Leu-Lys)

Dansylation without hydrolysis after a single Edman degradation step gave bis-dansyl-lysine.

Peptide T14 (100-104) (Gln-Ala-Thr-Asn-Leu)

Digestion with carboxypeptidase-A for 3 h followed by dansyl analysis and dansylation without hydrolysis after four Edman degradations both yielded dansyl-leucine. The

pH 6.5 electrophoretic mobility indicated that residues-100 and 103 were both amides, although a peptide of identical sequence was isolated with a pH 6.5 mobility equal to -1.90. This was attributed to deamidation during preparation rather than a real alternative to the sequence.

Amino Acid Sequence

The overlapping chymotryptic and tryptic peptides gave the sequence of Macrobrachium cytochrome c as shown in Figure 32. The sequence agreed with the amino acid composition data shown in Table 26 with a number of exceptions. In all determinations the values obtained for serine and lysine were routinely high, but no evidence for an impurity responsible for these discrepancies could be found from the sequence analysis. Tryptophan was not determined as part of the amino acid composition, but the best spectral ratios of purified Macrobrachium cytochrome c indicated that one residue was present.

From Figure 32 it can be seen that all the residues were positively identified by sequence analysis in both chymotryptic and tryptic digestions with the exception of positions 1-7, 15-17, 24, 25, 51-53, 57, 59 and 63-65. These were placed from positive identification by sequence analysis from only one of the digestions, from amino acid composition data or from a combination of the two. Tryptophan-59 could not be positively identified in the appropriate tryptic peptide

because of the destruction of dansyl-tryptophan during the acid hydrolysis stage of dansyl analysis. Residues 1-4 and cysteine-17 were not seen in either digestion and could only be placed from composition data. The nature of the N-terminal blocking group was not determined.

All possible overlaps between chymotryptic and tryptic peptides were observed except in the region C1/T1-T2, T2A where the peptide order is clear from the position of the blocked N-terminus, and in the region C7, C7B-C8/T9A-T9B where the point of cleavage duplicates the chymotryptic evidence due to an anomalous tryptic break at tyrosine-67. In this case the peptide order is made clear from a consideration of the appropriate region in other cytochromes c.

All the acid and amide residues indicated in the sequence were placed, with the exception of glutamine-16, from the pH 6.5 electrophoretic mobilities of intact or partially degraded peptides taken from either, or both, of the chymotryptic and tryptic digestions. Glutamine-16 was placed from a comparison of mobilities with the equivalent peptides in Locusta cytochrome c where the amide content of the heme region was unequivocally determined. The electrophoretic alternatives of peptides C2CB, C4A, C7A, C14, C14A, T6 and T14 that were isolated were attributed to the deamidation of amide groups during preparation of the peptides.

The observed enzyme specificities were consistent with those expected (Smyth, 1967) except that only partial chymotryptic

cleavage was observed at tyrosine-46, phenylalanine-82 and tyrosine-97. In addition, full chymotryptic cleavage occurred at histidine-26 and leucine-85, whilst partial cleavage occurred at leucine-24, leucine-32, asparagine-33, leucine-64, methionine-80, leucine-94 and leucine-97. Partial tryptic cleavage was observed at lysine residues-7, 72 and 87 together with phenylalanine-82 and tyrosine-97. Full tryptic cleavage was observed at tyrosine-67.

10. The Amino Acid Sequence of Eisenia Cytochrome c

The amino acid sequence of Eisenia cytochrome c was determined from the evidence of two chymotryptic digestions and two tryptic digestions using a total of 2 μmol (24 mg) of protein. The sequence is shown in Figure 33 giving the points of enzyme cleavage together with the overlapping peptides from which the sequence was deduced. A list of chymotryptic peptides, together with electrophoretic mobility and sequence data, is given in Table 28, and a similar list of tryptic peptides is given in Table 29. The amino acid composition of Eisenia cytochrome c was obtained from two duplicate 50 μg samples hydrolysed for 24, 48 and 72 h respectively, and this is given in Table 30.

Digestion

The oxidised protein readily denatured in 80% ethanol and was adequately digested by both trypsin and chymotrypsin.

The first chymotryptic digest was performed on 8 mg of

FIGURE 33.

The Amino Acid Sequence of *Eisenia* Cytochrome *c*.

Residues which were identified by dansyl-Edman analysis are indicated by ⟶; those identified by other means, e.g. by amino acid composition data, are indicated by --⟶; those identified by digestion with carboxypeptidase-A are indicated by ⟵; arrows ⟹ indicate that the C-terminal residue was identified as the free amino acid. The arrows ↓↑ indicate points of complete enzymic cleavage, upwards for trypsin and downwards for chymotrypsin; ⇣⇡ indicate points of partial cleavage, and ⇓ indicates points of cleavage using chymotrypsin to digest the heme peptides.

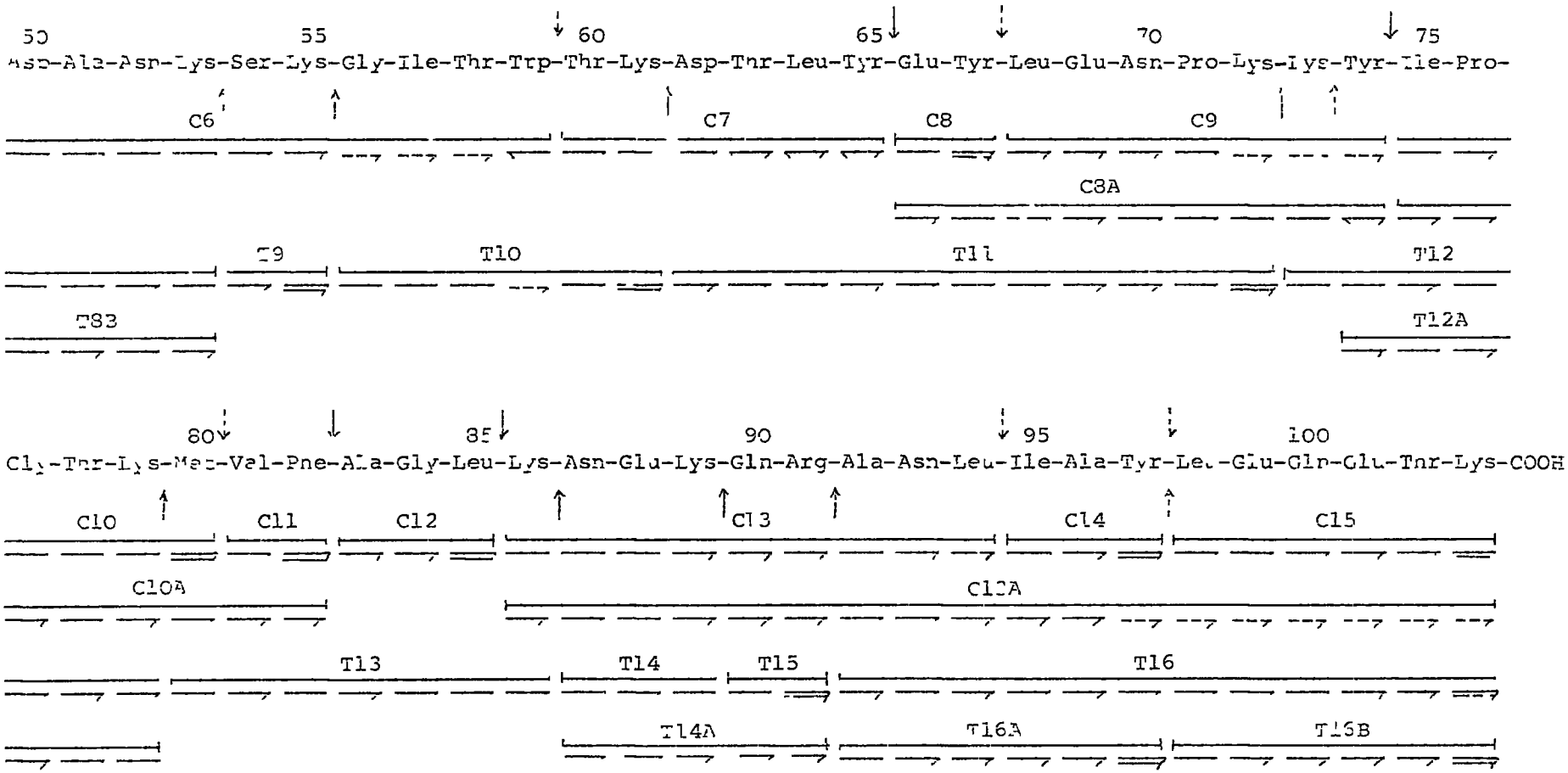

50 55 60 65 70 75
Asp-Ala-Asn-Lys-Ser-Lys-Gly-Ile-Thr-Trp-Thr-Lys-Asp-Thr-Leu-Tyr-Glu-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr-Ile-Pro-
C6
C7
C8
C9
C8A
T9
T10
T11
T12
T8B
T12A
80 85 90 95 100
Gly-Thr-Lys-Met-Val-Phe-Ala-Gly-Leu-Lys-Asn-Glu-Lys-Gln-Arg-Ala-Asn-Leu-Ile-Ala-Tyr-Leu-Glu-Gln-Glu-Thr-Lys-COOH
C10
C11
C12
C13
C14
C15
C10A
C12A
T13
T14
T15
T16
T14A
T16A
T16B

TABLE 28.

Chymotryptic Peptides from Eisenia Cytochrome c

Peptide/ Position	M (pH 6.5)	M (pH 1.9)	RDNS-ARG BAWP	Dansyl-Edman Results
C1 (-5-10)	0	0.70		Gly-Gly-Ile-Pro-Ala-Glx-Asx-Val-Glx-Lys-Gly-Lys-Thr-Ile-Phe
C2 (11-26)	0.44	0.49		Lys- (Heme peptide; see text)
C2CA (11-18)		0.58		Lys-Glx-Arg-$CySO_3$-Ala-(Glx, $CySO_3$)-His
C2CB (19-26)	0.44	0.94		Thr-Val-Asx-Lys-Gly-Gly-Pro-His
C3 (27-36)	1.09	0.93		Lys-Thr-Gly-Pro-Asx-Leu-His-Gly-Ile-Phe
C3A (27-33)	Not isolated.			
C3B (34-36)	0	0.56		Gly-Ile-Phe
C4 (37-46)	0.79	0.66		Gly-Arg-Ala-Thr-Gly-Glx-Ala-Ala-Gly-Phe
C5 (47-48)	0	0.59	0.61	Ala-Tyr
C6 (49-59)	0.56	0.68		Thr-Asx-Ala-Asx-Lys-Ser-Lys-(Gly,Ile,Thr)-Trp
C7 (60-65)	0	0.63	0.55	Thr-Lys-Asx-Thr-Leu-Tyr
C8 (66-67)	-1.64	0.54		Glx-Tyr
C8A (66-74)	0	0.76	0.44	Glx-Tyr-Leu-Glx-Asx-Pro-Lys-Lys-Tyr
C9 (68-74)	0.79	0.93		Leu-Glx-Asx-Pro-(Lys,Lys,Tyr)

TABLE 28 (Cont'd.)

C10 (75-80)	0.92	0.75		Ile-Pro-Gly-Thr-Lys-Met
C10A (75-82)	0.75	0.66		Ile-Pro-Gly-Thr-Lys-Met-Val-Phe
C11 (81-82)	0	0.56	0.96	Val-Phe
C12 (83-85)	0	0.69	0.92	Ala-Gly-Leu
C13 (86-94)	1.31	1.10		Lys-Asx-Glx-Lys-Glx-Arg-Ala-Asn-Leu
C13A (86-103)	0.44	0.83		Lys-Asx-Glx-Lys-Glx-Arg-Ala-Asn-Leu-Ile-Ala-(Tyr,Leu,Glx, Glx,Glx,Thr,Lys)
C14 (95-97)	0	0.54	0.84	Ile-Ala-Tyr
C15 (98-103)	-0.83	0.70		Leu-Glx-Glx-Glx-Thr-Lys

TABLE 29.

Tryptic Peptides from Eisenia Cytochrome c

Peptide/ Position	M (pH 6.5)	M (pH 1.9)	RDNS-ARG BAWP	Dansyl-Edman Results
T1 (-5-5)	-0.67	0.61		Gly-Gly-Ile-Pro-Ala-Gly-Asx-Val-Glx-Lys
T1A (-5-7)	0	0.65		Gly-Gly-Ile-Pro-Ala-Gly-Asx-Val-Glx-Lys-Gly-Lys
T2 (6-7)	2.14	1.47		Gly-Lys
T3 (8-11)	1.03	0.89		Thr-Ile-Phe-Lys
T4 (12-13)	1.54	1.16	0.21	Glx-Arg
T5 (14-22)	0.20	0.35		(Heme peptide; see text)
T5A (14-22)		0.42		$CySO_3$-Ala-Glx-$CySO_3$-(His,Thr, Val,Asx,Lys)
T6 (23-27)	Not isolated.			
T7 (28-38)	1.03	0.69		Thr-Gly-Pro-Asx-Leu-His-Gly-Ile-Phe-Gly-Arg
T8 (39-53)	0	0.47		Ala-Thr-Gly-Glx-Ala-Ala-Gly-Phe-Ala-Tyr-Thr-Asx-Ala-Asx-Lys
T8A (39-48)	0	0.24		Ala-Thr-Gly-Glx-Ala-Ala-Gly-Phe-Ala-Tyr
T8B (49-53)	0	0.97		Thr-Asx-Ala-Asx-Lys
T9 (54-55)	1.98	1.47		Ser-Lys
T10 (56-61)	0.83	0.71	0.33	Gly-Ile-Thr-(Trp)-Thr-Lys

TABLE 29 (Cont'd.)

T11 (62-72)	-1.02	0.44	0.49	Asx-Thr-Leu-Tyr-Glx-Tyr-Leu-Glx-Asx-Pro-Lys
T12 (73-79)	1.47	1.00		Lys-Tyr-Ile-Pro-Gly-Thr-Lys
T12A (74-79)	0.83	0.77		Tyr-Ile-Pro-Gly-Thr-Lys
T13 (80-86)	0.83	0.72		Met-Val-Phe-Ala-Gly-Leu-Lys
T14 (87-89)	0	1.05		Asx-Glx-Lys
T14A (87-91)	0.92	1.02		Asx-Glx-Lys-Glx-Arg
T15 (90-91)	1.50	1.19	0.21	Glx-Arg
T16 (92-103)	-0.62	0.46	0.41	Ala-Asx-Leu-Ile-Ala-Tyr-Leu-Glx-Glx-Glx-Thr-Lys
T16A (92-97)	0	0.35	0.43	Ala-Asx-Leu-Ile-Ala-Tyr
T16B (98-103)	0.88	0.76		Leu-Glx-Glx-Glx-Thr-Lys

TABLE 30.

The Amino Acid Composition of Eisenia Cytochrome c

	Mean Values 28 h hydrolysis	Mean Values 48 h hydrolysis	Mean Values 72 h hydrolysis	Mean Corrected Values	Amino Acid Analysis	Sequence Values
Asp	9.8	9.1	9.0	9.3	9	9
Thr	9.3	8.6	8.2	9.5*	10	10
Ser	2.8	2.0	1.8	3.2*	3	1
Glu	11.0	10.5	11.4	11.0	11	11
Pro	5.1	4.8	5.0	5.0	5	5
Gly	14.0	13.7	13.6	13.8	14	14
Ala	10.3	10.2	9.9	10.1	10	10
Cysteine	1.9	1.4	1.6	1.6	2	2
Val	2.9	2.6	2.6	2.7	3	3
Met	1.0	0.9	0.9	0.9	1	1
Ile	5.9	6.3	6.2	6.1	6	6
Leu	5.7	5.5	6.3	5.8	6	6
Tyr	4.9	4.5	4.2	4.5	5	5
Phe	3.2	2.8	2.7	2.9	3	4
His	2.9	2.9	2.9	2.9	3	3
Lys	16.9	15.8	15.7	16.1	16	14
Arg	4.1	4.1	4.0	4.1	4	3
Trp	-	-	-	-	+	1

The mean values were obtained from the analysis of 2 samples of 50 µg of cytochrome c at each hydrolysis time. The mean corrected values were obtained as an average of the six determinations with corrections for the destruction of certain amino acids.

*Calculated from 24 h and 72 h values assuming first order kinetics for destruction (Moore & Stein, 1963).

+Trp was not determined; the best spectral ratios of the cytochrome indicated the presence of one residue.

oxidised and denatured cytochrome c equilibrated at pH 8.0 and 37°C under nitrogen. 2% (w/w) enzyme was added at zero time, a further 2% at 60 min, and the digestion was terminated after 150 min. The digest was separated into two fractions after passage through a 1 cm x 110 cm column of G-15 Sephadex equilibrated and eluted in 10% acetic acid. The peptides in these two fractions were then separately purified in the normal manner. The second chymotryptic digest was performed on 4 mg of oxidised and denatured cytochrome c equilibrated as before. 2% (w/w) enzyme was added at zero time and after 45 min, and the digestion was terminated after 60 min. The peptides were then purified in the normal way. The two tryptic digestions were performed on 4 mg and 8 mg of cytochrome c respectively under identical conditions to the chymotryptic work. In the first instance 2% (w/w) enzyme was added at zero time and 90 min, and the digestion was terminated after 120 min, and in the second, 2% (w/w) enzyme was added at zero time and after 20 min, and the digestion was terminated after 50 min of incubation.

Chymotryptic Peptides

Peptide C1 (-5-+10) (Gly-Gly-Ile-Pro-Ala-Gly-Asp-Val-Glu-Lys-Gly-Lys-Thr-Ile-Phe)

Digestion with carboxypeptidase-A for 3 h at 37°C and pH 8.5, followed by dansyl analysis, yielded dansyl-phenylalanine, whilst a similar 9 h digestion yielded the

dansyl derivatives of phenylalanine and isoleucine after analysis. A consideration of the pH 6.5 electrophoretic mobility fixed residues-2 and 4 as aspartic acid and glutamic acid respectively.

Peptide C2 (11-20) (Heme peptide)

Lysine was shown to be the N-terminal amino acid of the heme peptide. After a mild performic oxidation, the peptide was digested with 5% (w/w) chymotrypsin at 37°C and pH 8.0 for 30 min and the products were separated by pH 1.9 electrophoresis as pepcides C2CA and C2CB.

Peptide C2CA (11-18) (Lys-Gln-Arg-$CySO_3$-Ala-Gln-$CySO_3$-His)

The peptide gave a positive reaction with the phenanthraquinone reagent and arginine was fixed at position 13 by sequence analysis. Digestion with carboxypeptidase-A for 3 h yielded bis-dansyl-histidine after dansyl analysis. The dansyl-Edman analysis was inconclusive beyond five degradation steps and residues-16 and 17 were placed from the semi-quantitative amino acid compositions of the intact peptide and the peptide after five degradations. Residue-12 was placed as glutamine from the pH 6.5 electrophoretic mobility of peptide T4, and residue-16 as glutamine from the pH 6.5 mobility of T5.

Peptide C2CB (19-26) (Thr-Val-Asp-Lys-Gly-Gly-Pro-His)

Digestion with carboxypeptidase-A for 9 h yielded bis-dansyl-histidine after dansyl analysis and dansylation

without hydrolysis after seven Edman degradations fixed histidine as the C-terminus of C2CB. A consideration of the pH 6.5 electrophoretic mobility indicated that residue-21 was aspartic acid, assuming that histidine would carry a change of less than +1 at pH 6.5.

Peptide C3 (27-36) (Lys-Thr-Gly-Pro-Asn-Leu-His-Gly-Ile-Phe)

Digestion with carboxypeptidase-A followed by dansyl-analysis yielded dansyl-phenylalanine after 3 h and dansyl-phenylalanine, together with dansyl-isoleucine after 9 h. Histidine was fixed as residue-33 from the identification of α-dansyl-histidine on N-terminal analysis after six Edman degradations and from semi-quantitative amino acid composition determinations after six and seven Edman degradation steps. The pH 6.5 electrophoretic mobility indicated that residue-31 was an amide assuming that histidine carried a charge of less than +1 at pH 6.5 (see C2CB).

Peptide C3A (27-33)

The peptide was not isolated.

Peptide C3B (34-36) (Gly-Ile-Phe)

Dansylation without hydrolysis gave dansyl-phenyl-alanine after two Edman degradation steps.

Peptide C4 (37-46) (Gly-Arg-Ala-Thr-Gly-Gln-Ala-Ala-Gly-Phe)

The peptide gave a positive reaction with the phenanthraquinone reagent and arginine was fixed as residue-38

from the dansyl-Edman analysis. Dansylation without hydrolysis after nine Edman degradation steps gave dansyl-phenylalanine The pH 6.5 electrophoretic mobility of C4 indicated that residue-42 was glutamine.

Peptide C5 (47-48) (Ala-Tyr)

Dansylation without hydrolysis after a single Edman degradation yielded bis-dansyl-tyrosine.

Peptide C6 (49-59) (Thr-Asp-Ala-Asn-Lys-Ser-Lys-Gly-Ile-Thr-Trp)

The peptide gave a positive reaction with the Ehrlich reagent and showed a pink coloration during the trifluoroacetic acid stage of the first Edman degradation, indicating the presence of tryptophan (Uphaus et al., 1959). Digestion with carboxypeptidase-A for 3 h followed by dansyl analysis yielded dansyl-tryptophan. A consideration of the pH 6.5 electrophoretic mobility indicated the presence of one acidic residue and this was placed as aspartic acid-50 from the pH 6.5 mobility of a peptide sample after three Edman degradations. This sample yielded three ninhydrin positive zones of mobilities zero, 0.72 and 1.35, due to the incomplete blocking of the ε-amino function of the lysine residues by phenylthiocarbamyl groups, as a result of their exposure to PITC. These values are consistent with residues-50 and 52 being aspartic acid and asparagine respectively.

Peptide C7 (60-68) (Thr-Lys-Asp-Thr-Leu-Tyr)

Digestion with carboxypeptidase-A for 3, 9 and 24 h, followed by dansyl analysis, fixed the sequence of residues-66-68 and this was confirmed by the dansyl-Edman analysis. The pH 6.5 electrophoretic mobility indicated that residue-62 was aspartic acid.

Peptide C8 (66-67) (Glu-Tyr)

Dansylation without hydrolysis after a single Edman degradation gave bis-dansyl-tyrosine. The pH 6.5 electrophoretic mobility of C8 indicated that residue-66 was glutamic acid.

Peptide C9 (68-74) (Leu-Glu-Asn-Pro-Lys-Lys-Tyr)

The dansyl-Edman analysis was unclear beyond four degradation steps and residues 72-74 were placed from the composition data of C9 and the sequence evidence for C8A. A sample of the peptide residue after two Edman degradations had a maximum pH 6.5 electrophoretic mobility of 1.92 indicating that residues-69 and 70 were glutamic acid and asparagine respectively.

Peptide C10 (75-80) (Ile-Pro-Gly-Thr-Lys-Met)

Dansylation without hydrolysis after five Edman degradation steps confirmed methionine as the C-terminal residue of C10.

Peptide C10A (75-82) (Ile-Pro-Gly-Thr-Lys-Met-Val-Phe)

Digestion with carboxypeptidase-A followed by dansyl

analysis yielded dansyl-phenylalanine after 1 h and dansyl-phenylalanine together with a significant amount of dansyl-valine after 3 h.

Peptide C11 (81-82) (Val-Phe)

Dansylation without hydrolysis after a single Edman degradation yielded dansyl-phenylalanine.

Peptide C12 (83-85) (Ala-Gly-Leu)

Dansylation without hydrolysis after two Edman degradation steps confirmed leucine as the C-terminus of C12.

Peptide C13 (86-94) (Lys-Asn-Glu-Lys-Gln-Arg-Ala-Asn-Leu)

The peptide gave a positive reaction with the phenanthraquinone reagent and dansylation without hydrolysis after eight Edman degradation steps yielded dansyl-leucine. The pH 6.5 electrophoretic mobility indicated the presence of one acidic residue and this was placed as glutamic acid-88 from the evidence of T14. A peptide of identical sequence was isolated with a pH 6.5 electrophoretic mobility of 0.77 and this was assumed to be a deamidated form of C13 rather than a real alternative to the sequence.

Peptide C13A (86-103) (Lys-Asn-Glu-Lys-Arg-Ala-Asn-Leu-Ile-Ala-Tyr-Leu-Glu-Gln-Glu-Thr-Lys)

The dansyl-Edman analysis was unclear beyond eleven degradation steps and residues 97-103 were placed from the semi-quantitative amino acid composition of C13A, the electrophoretic mobilities of C13A and the sequence evidence for peptides C14

and C15.

Peptide C14 (95-97) (Ile-Ala-Tyr)

Dansylation without hydrolysis after two Edman degradation steps yielded bis-dansyl-tyrosine.

Peptide C15 (98-103) (Leu-Glu-Gln-Glu-Thr-Lys)

Dansylation without hydrolysis after five Edman degradations yielded bis-dansyl-lysine. A consideration of the pH 6.5 electrophoretic mobility indicated the presence of two acid residues and these were placed at positions 99 and 101 from the mobility evidence of partially degraded samples. A sample of C15 after two degradation steps gave two ninhydrin positive zones of mobilities zero and -1.20, fixing residue-101 as glutamic acid. The placing of glutamic acid-99 and glutamine-100 is consistent with the evidence, as an amide at position 99 would have led to an additional more acidic spot in the mobility determination of the peptide after two degradation steps.

Tryptic Peptides

Peptide T1 (-5-+5) (Gly-Gly-Ile-Pro-Ala-Gly-Asp-Val-Glu-Lys)

The pH 6.5 electrophoretic mobility indicated that residues-2 and 4 were aspartic acid and glutamic acid respectively.

Peptide T2A (-5-17) (Gly-Gly-Ile-Pro-Ala-Gly-Asp-Val-Glu-Lys-Gly-Lys)

Dansylation without hydrolysis after eleven Edman degradation steps yielded α-dansyl-ε-PTC-lysine. The pH 6.5 electrophoretic mobility indicated that residues-2 and 4 were aspartic acid and glutamic acid respectively.

Peptide T2 (6-7) (Gly-Lys)

Peptide T3 (8-11) (Thr-Ile-Phe-Lys)

Peptide T4 (12-13) (Gln-Arg)

The peptide gave a positive result with the phenanthraquinone reagent and dansylation without hydrolysis after a single Edman degradation step placed arginine as the C-terminal residue of T4. The pH 6.5 electrophoretic mobility indicated that residue-12 was glutamine.

Peptide T5 (14-22) (Heme peptide)

The peptide was dehemed by a mild performic oxidation and purified on pH 1.9 electrophoresis as peptide T5A.

Peptide T5A (14-22) ($CySO_3$-Ala-Gln-$CySO_3$-His-Thr-Val-Asp-Lys)

The dansyl-Edman analysis was unclear beyond four Edman degradation steps and residues 18-22 were placed from composition data for T5A and the sequence evidence of C2CA and C2CB. The electrophoretic mobility could not be used for the fixing of amide residues due to the presence of cysteic acid and histidine residues, but the aspartic acid

at position 21 could be placed from the evidence of C2CB. A comparison of the mobilities of the tryptic heme peptides of Asterias and Eisenia, which were identical in the sequence except for Asterias having glutamic acid at position 21 instead of the aspartic acid of Eisenia, indicated that position 16 in Eisenia was glutamine.

Peptide T6 (23-27)

The expected tryptic peptide T6 was not isolated.

Peptide T7 (28-38) (Thr-Gly-Pro-Asn-Leu-His-Gly-Ile-Phe-Gly-Arg)

The peptide gave positive reactions with both the phenanthraquinone and Pauly reagents. Histidine at position 33 was identified as the α-dansyl-derivative on N-terminal analysis of the peptide after five degradation steps, and as the bis-dansyl-derivative in a semi-quantitative amino acid composition conducted at this point. Bis-dansyl histidine was absent from a semi-quantitative composition conducted after six Edman degradation steps. Dansylation without hydrolysis after ten Edman degradations yielded dansyl-arginine. Despite the presence of histidine, the pH 6.5 electrophoretic mobility indicated that residue-31 was an amide. A peptide of identical sequence was isolated with a pH 6.5 electrophoretic mobility of 0.30. It was assumed to be a deamidated form of T7 resulting from the preparation, rather than a real alternative to the sequence.

Peptide T8 (39-53) (Ala-Thr-Gly-Gln-Ala-Ala-Gly-Phe-Ala-Tyr-Thr-Asp-Ala-Asn-Lys)

The pH 6.5 electrophoretic mobility indicated that one acidic residue was present and this was placed as aspartic acid-50 from the mobility evidence of C6. Residues-42 and 52 were therefore placed as amides agreeing with the evidence of peptides C4 and C6.

Peptide T8A (39-48) (Ala-Thr-Gly-Gln-Ala-Ala-Gly-Phe-Ala-Tyr)

Dansylation without hydrolysis after nine Edman degradation steps yielded bis-dansyl-tyrosine. The zero pH 6.5 electrophoretic mobility fixed residue-42 as glutamine.

Peptide T8B (49-53) (Thr-Asp-Ala-Asn-Lys)

A sample of the peptide gave two ninhydrin positive zones of mobilities zero and 1.30 on pH 6.5 electrophoresis after having undergone two Edman degradation steps. This evidence, together with pH 6.5 mobility of the intact peptide, fixed residue-50 as aspartic acid.

Peptide T9 (54-55) (Ser-Lys)

Dansylation without hydrolysis after a single Edman degradation yielded bis-dansyl-lysine.

Peptide T10 (56-61) (Gly-Ile-Thr-Trp-Thr-Lys)

The peptide gave a positive result with the Ehrlich reagent but tryptophan was not positively identified as the N-terminus following three Edman degradation steps due to the acid destruction of the dansyl-derivative in the N-terminal

method. The residue-59 was therefore placed from the chymotryptic evidence. Dansylation without hydrolysis after five Edman degradations yielded α-dansyl-ε-PTC-lysine.

Peptide T11 (62-72) (Asp-Thr-Leu-Tyr-Glu-Tyr-Leu-Glu-Asn-Pro-Lys)

Dansylation without hydrolysis after ten Edman degradation steps yielded α-dansyl-ε-lysine. A consideration of the pH 6.5 electrophoretic mobility indicated that three acidic residues were present and these were placed at positions 62, 66, and 69 from the evidence of peptides C7, C8 and C9 respectively.

Peptide T12 (73-79) (Lys-Tyr-Ile-Pro-Gly-Thr-Lys)

Peptide T12A (74-79) (Tyr-Ile-Pro-Gly-Thr-Lys)

Peptide T13 (80-86) (Met-Val-Phe-Ala-Gly-Leu-Lys)

Peptide T14 (87-89) (Asn-Glu-Lys)

The pH 6.5 electrophoretic mobility of the intact peptide indicated that a single acid residue is present. A sample of the peptide taken after a single Edman degradation gave two ninhydrin positive zones on pH 6.5 electrophoresis of mobilities zero and -1.62 due to the incomplete blocking of the ε-amino function of lysine by a phenylthiocarbamyl group after exposure to PITC. This evidence was consistent with residue-88 being glutamic acid.

Peptide T14A (88-91) (Asn-Glu-Lys-Ala-Arg)

The peptide gave a positive reaction with the

phenanthraquinone reagent and dansylation without hydrolysis confirmed arginine as the C-terminus of T14A. The pH 6.5 electrophoretic mobility of T14A indicated the presence of a single acidic residue and this was placed as glutamic acid-89 from the evidence of T14. A peptide of identical sequence to T14A was isolated having a neutral pH 6.5 electrophoretic mobility and this was assumed to be a deamidated form of T14A rather than a real alternative to the sequence.

Peptide T15 (90-91) (Gln-Arg)

The peptide gave a positive reaction with the phenanthraquinone reagent, and dansylation without hydrolysis after a single Edman degradation confirmed arginine as the C-terminus of T15. The pH 6.5 electrophoretic mobility indicated that residue-90 was glutamine although, as with T14A, a small proportion of the peptide appeared to be deamidated as judged by the isolation of a peptide in low yield having a pH 6.5 mobility of zero.

Peptide T16 (92-103) (Ala-Asn-Leu-Ile-Ala-Tyr-Leu-Glu-Gln-Glu-Thr-Lys)

Dansylation without hydrolysis after eleven Edman degradation steps yielded α-dansyl-ε-PTC-lysine. The pH 6.5 electrophoretic mobility of T16 indicated the presence of two acid residues and these were placed from the evidence of C15 at positions 99 and 101. Residue-93 was placed as asparagine from peptide T16A.

Peptide T16A (92-97) (Ala-Asn-Leu-Ile-Ala-Tyr)

Dansylation without hydrolysis after five Edman degradation steps yielded bis-dansyl-tyrosine. The pH 6.5 electrophoretic mobility indicated that residue-93 was asparagine.

Peptide T16B (98-103) (Leu-Glu-Gln-Glu-Thr-Lys)

Dansylation without hydrolysis after five Edman degradation steps gave α-dansyl-ε-PTC-lysine. The pH 6.5 electrophoretic mobility indicated the presence of two acidic residues and these were placed from the evidence of peptide C15.

Amino acid sequence

The overlapping chymotryptic and tryptic peptides gave the sequence of Eisenia cytochrome c as shown in Figure 33. The sequence agreed with the amino acid composition data shown in Table 30 with a number of exceptions. In all the determinations the values obtained for serine, lysine and arginine were routinely high, although no evidence for an impurity responsible for such discrepancies could be found from the sequence analysis. Tryptophan was not determined as part of the amino acid composition, but the best spectral ratios of the purified Eisenia cytochrome c indicated that one residue was present.

From Figure 33 it can be seen that all the residues were positively identified by sequence analysis in both chymotryptic

and tryptic digestions, with the exception of positions 16-27 and 56-59. These were placed from the positive identification by sequence analysis in only one of the digestions, from amino acid composition data or from a combination of the two. Tryptophan was not identified in the appropriate tryptic peptide because of the destruction of dansyl tryptophan during the acid hydrolysis stage of dansyl analysis. No residues were placed without positive identification from at least one digestion.

All possible overlaps between chymotryptic and tryptic peptides were observed except in the region C2CB, C3/T5A, T7, due to the non-isolation of the expected tryptic peptide T6. In this case the peptide order was made clear from a consideration of the appropriate region in other cytochromes c.

All the acidic and amide residues indicated in the sequence were placed, with the exception of glutamine-16, from the pH 6.5 electrophoretic mobilities of intact or partially degraded peptides taken from either, or both, of the chymotryptic and tryptic digestions. Glutamine-16 was placed from a comparison of the mobilities of the equivalent tryptic peptide from the Asterias sequence and indirectly from the Locusta heme region, where the amide content was unequivocally determined. The electrophoretic alternatives of peptides C13, T7, T14A and T15 that were isolated were attributed to the deamidation of amide groups during the preparation of the peptides.

The observed enzyme specificities were consistent with those expected (Smyth, 1967) except that only partial chymotryptic cleavage was observed at tyrosine-67 and tyrosine-97. In addition, full chymotryptic cleavage was observed at histidine-26 and leucine-85 and partial chymotryptic cleavage at histidine-33, methionine-80 and leucine-94. Partial tryptic cleavage was observed at lysine-5, and lysine-73, together with tyrosine-48 and tyrosine-97.

11. The Amino Acid Sequence of Asterias Cytochrome c

The amino acid sequence of Asterias cytochrome c was determined from the evidence of a chymotryptic and a tryptic digestion, together with a chymotryptic digestion of cyanogen bromide fragments, using a total of 2 μmol (24 mg) of protein. The sequence is given in Figure 34 showing the points of enzyme cleavage together with the overlapping peptides from which it was deduced. A list of the chymotryptic peptides, together with mobility and sequence data, is given in Table 31 and a similar list of tryptic peptides is given in Table 32. The amino acid composition of Asterias cytochrome c was obtained from three duplicate 50 μg samples hydrolysed for 24, 48 and 72 h respectively, and this is shown in Table 33.

Digestion

The oxidised protein was readily denatured by 80% ethanol (v/v) and adequately digested by both trypsin and chymotrypsin.

FIGURE 34.

The Amino Acid Sequence of *Asterias* Cytochrome *c*.

Residues which were identified by dansyl-Edman analysis are indicated by $\longrightarrow$; those identified by other means, e.g. by amino acid composition data, are indicated by $\dashrightarrow$; those identified by digestion with carboxypeptidase-A are indicated by $\longleftarrow$; arrows $\Longrightarrow$ indicate that the C-terminal amino acid was identified as the free amino acid. The arrows $\downarrow\uparrow$ indicate points of complete enzymic cleavage, upwards for trypsin and downwards for chymotrypsin; $\vdots$ indicate points of partial cleavage, and $\Downarrow$ indicates points of cleavage using chymotrypsin to digest the heme peptides.

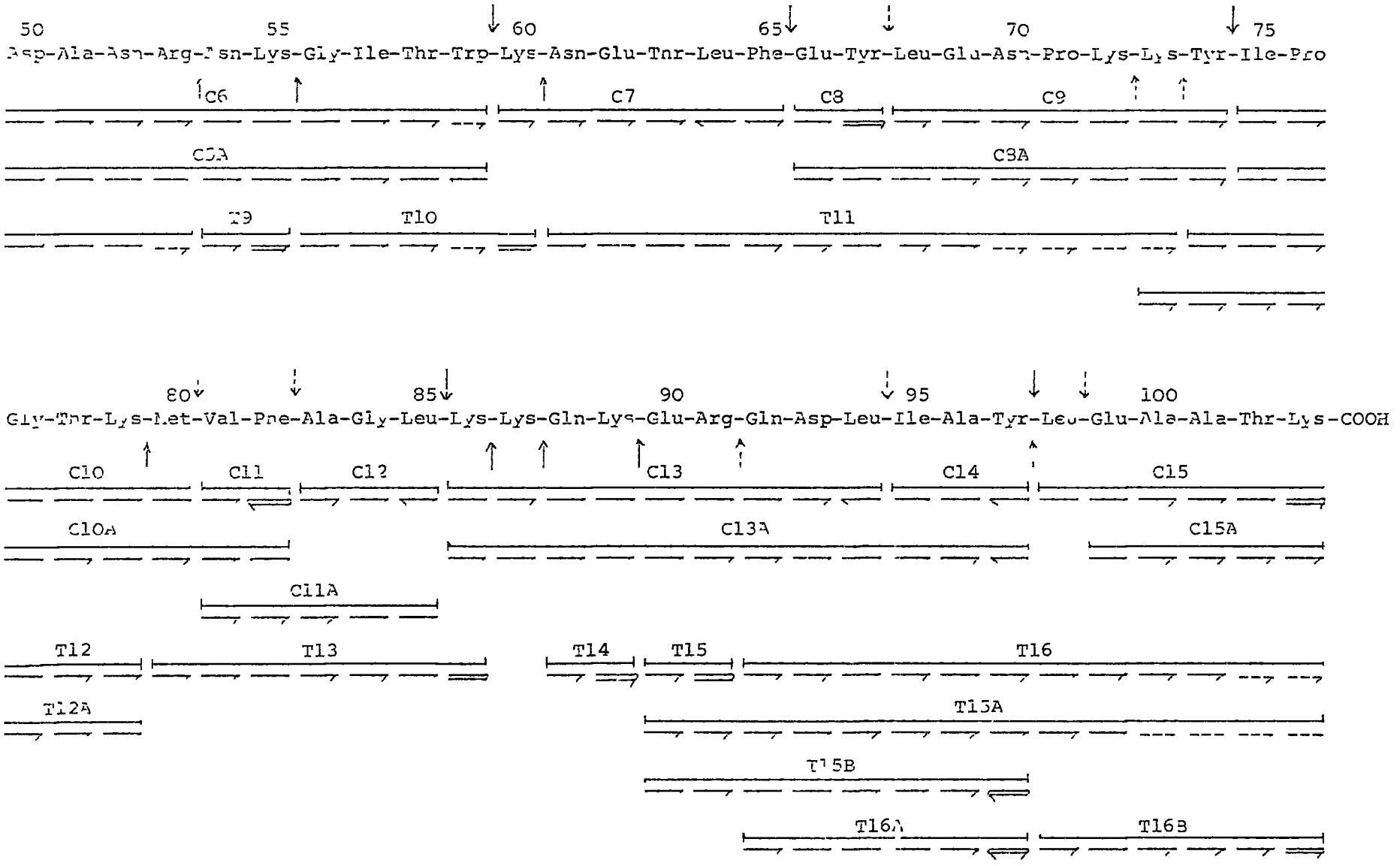

50
55
60
65
70
75
Asp-Ala-Asn-Arg-Asn-Lys-Gly-Ile-Thr-Trp-Lys-Asn-Glu-Thr-Leu-Phe-Glu-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr-Ile-Pro
C6
C7
C8
C9
C6A
C8A
T9
T10
T11
80
85
90
95
100
Gly-Thr-Lys-Met-Val-Phe-Ala-Gly-Leu-Lys-Lys-Gln-Lys-Glu-Arg-Gln-Asp-Leu-Ile-Ala-Tyr-Leu-Glu-Ala-Ala-Thr-Lys-COOH
C10
C11
C12
C13
C14
C15
C10A
C13A
C15A
C11A
T12
T13
T14
T15
T16
T12A
T13A
T15B
T16A
T16B

TABLE 31.

Chymotryptic Peptides for Asterias Cytochrome c

Peptide/ Position	M (pH 6.5)	M (pH 1.9)	RDNS-ARG BAWP	Dansyl-Edman Results
C1 (1-10)	0	0.85		Gly-Glx-Val-Glx-Lys-Gly-Lys-Lys-Ile-Phe
C2 (11-26)	0.20	0.47	0.15	Val- (Heme peptide; see text)
C2CA (11-18)		0.38		Val-Glx-Arg-$CySO_3$-Ala-Glx-($CySO_3$)-His
C2CB (19-26)	1.10	1.08		Thr-Val-Glx-Lys-Ala-Gly-Lys-His
C3 (27-36)	Not isolated.			
C3A (27-32)	0.95	0.76		Lys-Thr-Gly-Pro-Asx-Leu
C3B (33-36)	0	0.46		Asx-Gly-Ile-Leu
C4 (37-46)	1.10	0.85		Gly-Arg-Lys-Thr-Gly-Glx-Ala-Ala-Gly-Phe
C4A (40-46)	0	0.37	0.50	Thr-Gly-Glx-Ala-Ala-Gly-Phe
C5 (47-48)	0	0.70		Ser-Tyr
C5A (47-59)	0.56	0.73	0.24	Ser-Tyr-Thr-Asx-Ala-Asx-Arg-Asx-Lys-Gly-Ile-Thr-Trp
C6 (49-59)	0.56	0.73		Thr-Asx-Ala-Asx-Arg-Asx-Lys-Gly-Ile-Thr-(Trp)
C7 (60-65)	0	0.73	0.15	Lys-Asx-Glx-Thr-Leu-Phe
C8 (66-67)	-1.52	0.51		Glx-Tyr

TABLE 31 (Cont'd)

C8A (66-74)	0	0.80	0.46	Glx-Tyr-Leu-Glx-Asx-Pro-Lys-Lys-Tyr
C9 (68-74)	0.77	0.84		Leu-Glx-Asx-Pro-Lys-Lys-Tyr
C10 (75-80)	0.87	0.78	0.46	Ile-Pro-Gly-Thr-Lys-Met
C10A (75-82)	0.80	0.64		Ile-Pro-Gly-Thr-Lys-Met-Val-Phe
C11 (81-82)	0	0.67	0.98	Val-Phe
C11A (81-85)	0	0.49		Val-Phe-Ala-Gly-Leu
C12 (83-85)	0	0.68	0.83	Ala-Gly-Leu
C13 (86-94)	1.20	0.99		Lys-Lys-Glx-Lys-Glx-Arg-Glx-Asx-Leu
C13A (86-97)	1.09	0.93	0.17	Lys-Lys-Glx-Lys-Glx-Arg-Glx-Asx-Leu-Ile-Ala-Tyr
C14 (95-97	0	0.56	0.94	Ile-Ala-Tyr
C15 (98-103)	0	0.66	0.21	Leu-Glx-Ala-Ala-Thr-Lys
C15A (99-103)	0	0.81	0.14	Glx-Ala-Ala-Thr-Lys

TABLE 32.

Tryptic Peptides from Asterias Cytochrome c

Peptide/ Position	M (pH 6.5)	M (pH 1.9)	RDNS-ARG BAWP	Dansyl-Edman results
T1 (1-5)	0	0.77	0.44	Gly-Glx-Val-Glx-Lys
T2 (6-7)	2.05	1.55		Gly-Lys
T3 (9-13)	0.87	0.75	0.46	Ile-Phe-Val-Glx-Arg
T4 (14-22)	0	0.30	0.23	(Heme peptide; see text)
T4CA (14-18)		0.41		$CySO_3$-Ala-(Glx,$CySO_3$)-His
T4CB (19-22)		0.96		Thr-Val-Glx-(Lys)
T5 (23-25)	1.60	1.25	0.36	Ala-Gly-Lys
T6 (26-27)	2.30	1.62		His-Lys
T7 (28-38)	0.54	0.49	0.58	Thr-Gly-Pro-Asx-Leu-Asx-Gly-Ile-Leu-(Gly,Arg)
T8 (40-53)	0	0.56	0.23	Thr-Gly-Glx-Ala-Ala-Gly-Phe-Ser-Tyr-Asx-Ala-Asx-(Arg)
T9 (54-55)	1.78	1.28		Asx-Lys
T10 (56-59)	0.35	0.72	0.43	Gly-Ile-Thr-(Trp)-Lys
T11 (60-73)	-0.45	0.62	0.69	Asx-Glx-Thr-Leu-Phe-Glx-Tyr-Leu-Glx-(Asx,Pro,Lys,Lys)
T11A (60-72)	Not isolated.			

TABLE 32 (Cont'd.)

T12 (74-79)	0.87	0.76	0.58	Tyr-Ile-Pro-Gly-Thr-Lys
T12A (73-79)	1.78	0.90	0.38	Lys-Tyr-Ile-Pro-Gly-Thr-Lys
T13 (80-86)	0.87	0.66		Met-Val-Phe-Ala-Gly-Leu-Lys
T14 (88-89)	1.78	1.26		Glx-Lys
T15 (90-91)	0	1.15		Glx-Arg
T15A (90-103)	-0.54	0.55	0.24	Glx-Arg-Glx-Asx-Leu-Ile-Ala-Tyr-Leu-Glx-(Ala,Ala,Thr,Lys)
T15B (90-97)	-0.83	0.58		Glx-Arg-Glx-Asx-Leu-Ile-Ala-Tyr
T16 (92-103)	0.58	0.51	0.39	Glx-Asx-Leu-Ile-Ala-Tyr-Leu-Glx-Ala-Ala-(Thr,Lys)
T16A (92-97)	-0.83	0.28		Glx-Asx-Leu-Ile-Ala-Tyr
T16B (98-103)	0	0.68	0.28	Leu-Glx-Ala-Ala-Thr-Lys

TABLE 33.

The Amino Acid Composition of *Asterias* Cytochrome *c*

	Mean Values 24 h hydrolysis	Mean Values 48 h hydrolysis	Mean Values 72 h hydrolysis	Mean Corrected Values	Amino Acid Analysis	Sequence Values
Asp	8.4	8.9	7.9	8.4	9	8
Thr	7.0	6.6	6.4	7.2*	7	8
Ser	3.8	3.3	3.3	4.0*	4	1
Glu	10.7	10.9	11.0	10.9	11	13
Pro	4.0	4.3	3.7	4.0	4	3
Gly	11.0	11.1	11.2	11.1	11	11
Ala	9.0	9.6	8.7	9.1	9	9
Cysteine	1.4	1.7	1.8	1.6	2	2
Val	3.8	3.4	3.7	3.6	4	4
Met	1.2	0.8	0.5	0.8	1	1
Ile	4.6	4.7	4.8	4.7	5	5
Leu	7.1	6.7	6.9	6.9	7	7
Tyr	3.6	3.5	3.3	3.4	3	4
Phe	3.6	3.5	3.8	3.7	4	4
His	1.8	1.8	1.9	1.8	2	2
Lys	18.9	18.5	17.2	18.2	18	16
Arg	3.6	3.6	3.6	3.6	4	4
Trp	-	-	-	-	+	1

The mean values were obtained from the analysis of 2 samples of 50 µg of cytochrome *c* at each hydrolysis time. The mean corrected values were observed as an average of the six determinations with corrections made for the destruction of certain amino acids.

*Calculated from 24 h and 72 h values assuming first order kinetics for destruction (Moore & Stein, 1963).

+Trp was not determined; the best spectral ratios of the cytochrome indicated one residue is present.

For the chymotryptic digests, 8 mg of cytochrome c was oxidised and denatured and equilibrated at 37°C and pH 8.0 under nitrogen. 2% (w/w) enzyme was added at zero time; a further 2% (w/w) after 60 min and the digestion was terminated after 80 min had elapsed. 8 mg of cytochrome c was similarly treated prior to digestion with trypsin. After equilibration at 37°C and pH 8.0, 2% (w/w) trypsin was added at zero time; a further 2% at 40 min and the digestion was terminated after 70 min. A cyanogen bromide cleavage was performed on 8 mg of cytochrome c. The oxidised protein was dissolved in 200 µl of 70% (v/v) formic acid and a 150 molar excess of cyanogen bromide was added in 400 µl of 70% (v/v) formic acid. Incubation was for 24 h in the dark at 26°C and the two resulting protein fragments were separated by passage through a column (1 cm x 100 cm) of Sephadex G-50 equilibrated in 70% (v/v) formic acid. The two fragments were lyophilised and each digested with 5% chymotrypsin at 37°C and pH 8.0 for 1 h.

Chymotryptic Peptides

Peptide C1 (1-10 (Gly-Gln-Val-Glu-Lys-Gly-Lys-Lys-Ile-Phe)

The peptide was initially purified as part of a larger fragment as judged by the amino acid composition and electrophoretic mobilities. Digestion with 5% chymotrypsin for 30 min at pH 8.0 and 37°C produced a peptide whose sequence is given above and whose pH 1.9 electrophoretic mobility was

1.05. A peptide aliquot taken after three Edman degradations had an electrophoretic mobility at pH 6.5 of -0.72, and this information, together with that from Peptide T1, placed position 2 as glutamine and position 4 as glutamic acid.

Peptide C2 (11-26) (Heme peptide)

The heme peptide was purified by electrophoresis and paper chromatography and was shown to have valine as its N-terminal amino acid. After mild performic oxidation it was digestion with 5% chymotrypsin at pH 8.0 and 37°C to produce peptides C2CA and C2CB which were separated by electrophoresis at pH 1.9.

Peptide C2CA (11-18) (Val-Gln-Arg-$CySO_3$-Ala-Gln-$CySO_3$-His)

The pH 1.9 guide-strips gave positive Pauly and phenanthraquinone reactions indicating the presence of arginine and histidine, and digestion of a 10% peptide aliquot with carboxypeptidase-A for 9 h followed by dansyl analysis indicated histidine as the C-terminal amino acid. The dansyl-Edman analysis gave only the first six positions of the peptide; the cysteic acid at position 17 was placed from the semi-quantitative amino acid composition. The placement of amides was not possible from the electrophoretic mobilities of C2CA due to the presence of cysteic acid and histidine, but position 12 could be placed as glutamine from the evidence of peptide T3 and a comparison with the mobility of the

equivalent peptide in the Locusta sequence indicated that position 16 was also glutamine.

Peptide C2CB (19-26) (Thr-Val-Glu-Lys-Ala-Gly-Lys-His)

The peptide gave a positive Pauly reaction and a 24 h digestion of a sample with carboxypeptidase-A yielded bis-dansyl-histidine after dansyl analysis. A further sample of the peptide had an electrophoretic mobility at pH 6.5 of 1.10 and together with the mobility evidence from peptide T4CB places residue-21 as glutamic acid.

Peptide C3A (27-32) (Thr-Gly-Pro-Asn-Leu)

A 3 h digestion with carboxypeptidase-A of a 10% aliquot of C3A gave dansyl-leucine after dansyl analysis, and the C-terminus was confirmed as leucine by dansylation without hydrolysis after the fifth Edman degradation. The electrophoretic mobility at pH 6.5 fixed residue-31 as asparagine.

Peptide C3B (34-36) (Asn-Gly-Ile-Leu)

Digestion with carboxypeptidase-A for 3 h gave dansyl-leucine after dansyl analysis. The peptide was neutral during pH 6.5 electrophoresis indicating position 34 to be asparagine.

Peptide C4 (37-46) (Gly-Arg-Lys-Thr-Gly-Gln-Ala-Ala-Phe)

The presence of arginine was indicated using the phenanthraquinone reagent and confirmed during the sequence analysis. A 3 h digestion with carboxypeptidase-A followed by dansyl analysis produced dansyl-phenylalanine whilst similar

treatment of a 24 h digestion produced large amounts of dansyl-phenylalanine and dansyl-glycine, together with significant amounts of dansyl-alanine. The electrophoretic mobility at pH 6.5, although slightly lower than expected, indicated that position 42 was glutamine.

Peptide C4A (40-46) (Thr-Gly-Gln-Ala-Ala-Gly-Phe)

This peptide was obtained in good yield and could only have resulted from anomalous chymotryptic cleavage at arginine—38. Digestion for 3 h with carboxypeptidase-A followed by dansyl analysis indicated that phenylalanine was the C-terminal amino acid and this was confirmed by dansylation without hydrolysis following the sixth Edman degradation step. Position 42 was placed as glutamine from the electrophoretic mobility at pH 6.5.

Peptide C5 (47-48) (Ser-Tyr)

Dansylation without hydrolysis following the first Edman degradation step confirmed tyrosine as the C-terminal amino acid.

Peptide C5A (47-59) (Ser-Tyr-Thr-Asp-Ala-Asn-Arg-Asn-Lys-Gly-Ile-Thr-Trp)

The peptide was purified by electrophoresis and paper chromatography and gave positive results with the Ehrlich and phenanthraquinone reagents. Digestion of a sample with carboxypeptidase-A for 3 h, followed by dansyl analysis, indicated tryptophan to be the C-terminal residue. A similar

9 h digestion yielded significant amounts of threonine, together with the tryptophan, as judged by dansyl analysis. During the trifluoroacetic acid stage of the first Edman degradation, the sample exhibited a pink coloration, indicating the presence of tryptophan (Uphaus et al., 1959). The electrophoretic mobility of pH 6.5 of the purified peptide, and of samples of the peptide after four and six Edman degradation steps, indicated that residue-50 was aspartic acid.

Peptide C6 (49-59) (Thr-Asp-Ala-Asn-Arg-Asn-Lys-Gly-Ile-Thr-Trp)

The peptide gave a positive result with both the Ehrlich and phenanthraquinone reagents indicating the presence of tryptophan and arginine. The presence of tryptophan as the C-terminus could not be confirmed although mobility considerations demanded that this was so, together with the evidence of peptide C5A. The placing of aspartic acid at position 50 was from the evidence for peptide C5A.

Peptide C7 (60-65) (Lys-Asn-Glu-Thr-Leu-Phe)

Digestion with carboxypeptidase-A followed by dansyl analysis gave dansyl-phenylalanine after 3 h and both dansyl-phenylalanine and dansyl-leucine after 9 h. The pH 6.5 electrophoretic mobility indicates the presence of one acid residue and the mobility of a peptide aliquot at pH 6.5 after two steps of the Edman degradation was -1.15 placing position 62 as glutamic acid.

Peptide C8 (66-67) (Glu-Tyr)

The peptide had a pH 6.5 electrophoretic mobility of -1.52 indicating position 66 to be glutamic acid. Dansylation of a peptide aliquot after a single Edman degradation yielded bis-dansyl-tyrosine, confirming tyrosine as the C-terminal amino acid.

Peptide 8A (66-74) (Glu-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr)

The peptide was neutral during pH 6.5 electrophoresis indicating the presence of two acidic residues placed at position 66 from peptide C7 and position 69 from peptide C9. Digestion with carboxypeptidase for 3 h followed by dansyl analysis confirmed tyrosine as the C-terminal amino acid.

Peptide C9 (68-74) (Leu-Glu-Asn-Pro-Lys-Lys-Tyr)

The electrophoretic mobilities at pH 6.5 after purification, and after two Edman degradation steps, placed residue-69 as glutamic acid.

Peptide C10 (75-80) (Ile-Pro-Gly-Thr-Lys-Met)

During the trifluoroacetic acid stage of the first Edman degradation, the peptide showed a rose-pink coloration characteristic of tryptophan (Uphaus et al., 1959). No other evidence could be found to support this observation.

Peptide C10A (75-82) (Ile-Pro-Gly-Thr-Lys-Met-Val-Phe)

Digestion with carboxypeptidase-A for 3 h gave phenylalanine as the C-terminal amino acid after dansyl analysis.

Peptide C11 (81-82) (Val-Phe)

Carboxypeptidase-A digestion for 3 h yielded dansyl-phenylalanine after dansylation and this was confirmed as the C-terminal residue by dansylation without hydrolysis following the first Edman degradation step.

Peptide C11A (81-85) (Val-Phe-Ala-Gly-Leu)

Peptide C12 (83-85) (Ala-Gly-Leu)

Digestion with carboxypeptidase-A for 3 h followed by dansyl analysis placed leucine as the C-terminal amino acid.

Peptide C13 (86-94) (Lys-Lys-Gln-Lys-Glu-Arg-Gln-Asp-Leu)

The peptide gave a positive reaction with the phenanthraquinone reagent and digestion with carboxypeptidase-A followed by dansyl analysis fixed leucine as the C-terminal amino acid. The two acidic residues expected from the pH 6.5 electrophoretic mobility data were placed from the mobilities of tryptic peptides with T14 fixing residue-88 as glutamine, T15 residue-90 as glutamic acid, and T15B residue-92 as glutamine and residue-93 as aspartic acid.

Peptide C13A (86-97) (Lys-Lys-Gln-Lys-Glu-Arg-Gln-Asp-Leu-Ile-Ala-Tyr)

The peptide gave a positive reaction with the phenanthraquinone reagent. Digestion with carboxypeptidase-A yielded tyrosine after 3 h, and tyrosine together with significant amounts of alanine and isoleucine after 24 h incubation, as shown by dansyl analysis. The pH 6.5 electrophoretic

mobility indicated the presence of two acidic residues and these were placed as in C13 from the appropriate tryptic peptides.

Peptide C14 (95-97) (Ile-Ala-Tyr)

Digestion with carboxypeptidase-A for 3 h followed by dansyl analysis indicated that tyrosine was the C-terminal amino acid.

Peptide C15 (98-103) (Leu-Glu-Ala-Ala-Thr-Lys)

Dansylation without hydrolysis after five Edman degradations gave bis-dansyl-lysine confirming lysine as the C-terminal amino acid. The mobility of C15 on pH 6.5 electrophoresis indicated that residue-99 was glutamic acid.

Peptide C15A (99-103) (Glu-Ala-Ala-Thr-Lys)

The peptide was neutral on pH 6.5 electrophoresis confirming residue-99 asglutamic acid.

Tryptic Peptides

Peptide T1 (1-5) (Gly-Gln-Val-Glu-Lys)

Dansylation without hydrolysis after four Edman degradations yielded bis-dansyl-lysine. Consideration of the pH 6.5 electrophoretic mobility of T1 indicates the presence of a single acid residue and this was placed as glutamic acid at position 4 from the pH 6.5 mobility of a peptide sample after three Edman degradations which was equal to -1.40. A peptide with a similar sequence was purified in low yield

having a pI 6.5 electrophoretic mobility equal to -0.96. This was assumed to be a deamidated form rather than an alternative form of peptide T1.

Peptide T2 (6-7) (Gly-Lys)

Dansylation without hydrolysis after the first Edman degradation confirmed position 7 as lysine.

Peptide T3 (9-13) (Ile-Phe-Val-Gln-Arg)

The peptide gave a positive reaction with the phenanthraquinone reagent and although the result of the final dansyl-Edman step was unclear, dansylation without hydrolysis at this stage confirmed arginine as the C-terminal amino acid. The pH 6.5 electrophoresis mobility data places residue-12 as glutamine.

Peptide T4 (14-22) (Heme peptide)

The heme peptide was purified by electrophoresis and paper chromatography, dehemed using a mild performic oxidation, and digested with 5% chymotrypsin at 37°C and pH 8.0 for 1 h. The two products, T4CA and T4CB, were separated by electrophoresis at pH 1.9.

Peptide T4CA (14-18) ($CySO_3$-Ala-Gln-$CySO_3$-His)

Digestion with carboxypeptidase-A followed by dansyl analysis indicated that histidinewas the C-terminal residue. The dansyl-Edman evidence beyond the second Edman degradation was inconclusive and residues-16 and 17 were placed from a semi-quantitative amino acid composition conducted at this

stage. Position 16 was placed as glutamine from the evidence of peptide C2CA.

Peptide T4CB (19-22) (Thr-Val-Glu-Lys)

The third dansyl-Edman step proved inconclusive but the lysine was placed from the results of dansylation without hydrolysis after three Edman degradations. A sample of the purified peptide was neutral on pH 6.5 electrophoresis indicating that position 21 must be glutamic acid.

Peptide T5 (23-25) (Ala-Gly-Lys)

Dansylation of T5 after two Edman degradations yielded bis-dansyl-lysine.

Peptide T6 (26-27) (His-Lys)

Identification of α-N-dansyl-histidine during the determination of the N-terminus of T6 proved inconclusive, but the presence of bis-dansyl-histidine was observed during the semi-quantitative determination of the amino acid composition of T6. The composition of T6 showed an absence of any histidine after the first Edman degradation, thus fixing residue-26 as histidine. Dansylation without hydrolysis at this point yielded bis-dansyl-lysine.

Peptide T7 (28-37) (Thr-Gly-Pro-Asn-Leu-Asn-Gly-Ile-Leu-Gly-Arg)

The peptide reacted positively with the phenanthraquinone reagent indicating the presence of arginine. The dansyl-Edman evidence was inconclusive after nine degradations,

but a semi-quantitative amino acid composition conducted on the peptide residue at this stage indicated the presence of glycine and arginine. The pH 6.5 electrophoretic mobility was lower than expected but still indicated that residues-32 and 33 were both asparagine.

Peptide T8 (40-53) (Thr-Gly-Gln-Ala-Ala-Gly-Phe-Ser-Tyr-Thr-Asp-Ala-Asn-Arg)

The peptide gave a positive result with the phenanthraquinone reagent. The dansyl-Edman results were inconclusive beyond thirteen degradation steps but residue-53 was placed from amino acid composition data. The presence of a single acidic residue was indicated by the pH 6.5 electrophoretic mobility and this was placed at position 50 as aspartic acid from the evidence of peptide C5A.

Peptide T9 (54-55) (Asn-Lys)

Dansylation without hydrolysis after the first Edman degradation step yielded bis-dansyl-lysine. Position 55 was placed as asparagine from consideration of the pH 6.5 electrophoretic mobility.

Peptide T10 (56-60) (Gly-Ile-Thr-Trp-Lys)

The peptide gave a positive reaction with the Ehrlich reagent indicating the presence of tryptophan, which was placed at position 59 from the evidence of a blank dansyl-Edman analysis at degradation step three between positive analyses at degradation steps two and four. The tryptophan could also be

placed as residue-59 from the evidence of peptide C5A. Dansylation without hydrolysis after four Edman degradation steps confirmed lysine as the C-terminus of T10.

Peptide T11 (61-73) (Asn-Glu-Thr-Leu-Phe-Glu-Tyr-Leu-Glu-Asn-Pro-Lys-Lys)

The dansyl-Edman evidence was inconclusive beyond nine degradation steps, but a semi-quantitative amino acid composition at this stage yielded significant quantities of proline, aspartic acid and lysine. Residues-71 to 73 were placed from amino acid composition data and from the evidence of C8A and C9. The pH 6.5 electrophoretic mobility of T11 indicated that three acid residues were present and these were placed at positions 62, 66 and 69 from the mobility evidence of peptides C7, C8, C8A and C9.

Peptide T12 (74-79) (Tyr-Ile-Pro-Gly-Thr-Lys)

Peptide T12A (73-79 (Lys-Tyr-Ile-Pro-Gly-Thr-Lys)

Peptide T13 (80-86) (Met-Val-Phe-Ala-Gly-Leu-Lys)

Dansylation without hydrolysis after six Edman degradation steps yielded α-dansyl-ε-PTC-lysine.

Peptide T14 (88-89) (Gln-Lys)

Dansylation without hydrolysis after the first Edman degradation confirmed residue-89 as lysine whilst a consideration of the pH 6.5 electrophoretic mobility fixed residue-88 as glutamine.

Peptide T15 (90-91) (Glu-Arg)

The peptide reacted positively with the phenanthraquinone reagent and arginine was confirmed in position 91 by dansylation without hydrolysis after a single Edman degradation step. T15 was neutral on pH 6.5 electrophoresis, indicating that residue-90 was glutamic acid.

Peptide T15A (90-103) (Glu-Arg-Gln-Asp-Leu-Ile-Ala-Tyr-Leu-Glu-Ala-Ala-Thr-Lys)

The peptide gave a positive reaction with the phenanthraquinone reagent. The dansyl-Edman evidence was inconclusive beyond ten degradations, and residues 100-103 were placed from semi-quantitative amino acid compositions conducted on the purified peptide and the peptide residue after ten degradation steps. Consideration of the pH 6.5 electrophoretic mobility indicated the presence of three acid residues and these were placed as glutamic acid-90, aspartic acid-93 and glutamic acid-99 from peptides T15, T15B and T16B respectively.

Peptide T15B (90-97) (Glu-Arg-Gln-Asp-Leu-Ile-Ala-Tyr)

Digestion of the peptide for three hours with carboxypeptidase-A followed by dansyl analysis yielded bis-dansyl-tyrosine, indicating that T15B arose from an anomalous tryptic cleavage at tyrosine-97. The mobility of the purified peptide on pH 6.5 electrophoresis was -0.83 and this became -1.0 after three Edman degradation steps indicating residue-93 to be aspartic acid, given that residue-90 was glutamic acid from the evidence of peptide T15.

Peptide T16 (92-103) (Gln-Asp-Leu-Ile-Ala-Tyr-Leu-Glu-Ala-Ala-Thr-Lys)

The dansyl-Edman analysis was inconclusive beyond ten degradation steps so that residues-102 and 103 were placed from amino acid composition data and from peptides T16B, C16 and C16B. The two acidic residues expected from mobility considerations were placed at positions 93 and 99 from the evidence of peptides T15B and T16B respectively.

Peptide T16A (92-97) (Gln-Asp-Leu-Ile-Ala-Tyr)

Digestion with carboxypeptidase-A for 3 h followed by dansyl analysis yielded bis-dansyl-tyrosine indicating that T16A resulted from anomalous tryptic cleavage at tyrosine-97. Consideration of the pH 6.5 electrophoretic mobility indicated the presence of a single acidic residue and this was placed as aspartic acid at position 93 from the evidence of T15B.

Peptide T16B (98-103) (Leu-Glu-Ala-Ala-Thr-Lys)

Dansylation without hydrolysis after five Edman degradation steps confirmed lysine as the C-terminal residue of T16B. The peptide was neutral on pH 6.5 electrophoresis indicating that position 99 was occupied by glutamic acid.

Amino acid Sequence

The overlapping chymotryptic and tryptic peptides gave the sequence of Asterias cytochrome c as shown in Figure 34. The sequence agreed with the amino acid composition data shown in Table 33 with a number of exceptions. In all

determinations, values obtained for aspartic acid, serine and lysine were routinely high but no evidence for an impurity associated with the purified cytochrome could be found from the sequence analysis. The low value for glutamic acid may have been due to the incomplete deamidation of glutamine under the conditions of the acid hydrolysis. Tryptophan was not determined as part of the amino acid composition, but the best spectral rates of purified Asterias cytochrome c indicated that one residue was present.

From Figure 34 it can be seen that all the residues were positively identified by sequence analysis in both chymotryptic and tryptic digests, except at positions 16, 17, 22, 37, 38, 53, 70, 71 and 72. These residues were placed from only one of the digestions, from amino acid composition data, or from a combination of the two. The lysine residues at positions 8, 39 and 84 could not be positively identified in the tryptic digest because of the course of tryptic cleavage at (-X-Lys-Lys-X-) sequences, and tryptophan-59 was not observed in the tryptic digest because of the destruction of dansyl-tryptophan during the acid hydrolysis stage of dansyl analysis. Only cysteine-17 was not positively identified in either the chymotryptic or tryptic digest and this residue could only be placed from amino composition data.

All possible overlaps between chymotryptic and tryptic peptides were observed except in the region C3B-C4/T7-T8

where the placement of peptides is unambiguous from an inspection of the sequences of C3B and T7 and C4, C4A and T8.

All the amides indicated in the sequence could be placed, with the exception of glutamine-16, from the electrophoretic mobilities at pH 6.5 of the complete, or partially degraded, peptides taken from either, or both, of the chymotryptic and tryptic digests. Glutamine-16 is placed from a comparison of mobilities with the equivalent peptide in Locusta cytochrome c where the amide content of the heme region was unequivocally determined. Evidence for a proportion of T1, having a more acidic pH 6.5 electrophoretic mobility, was attributed to deamidation of glutamine-2 during peptide preparation, rather than heterogeneity at residue-2. No alternative forms of peptide C1 were observed.

The observed enzyme specificities were consistent with those expected (Smyth, 1967) except that only partial cleavage was observed at phenylalanine-46, tyrosine-48 and tyrosine-67. In addition, full chymotryptic cleavage was observed at histidine-26, leucine-32, leucine-36 and leucine-86, whilst partial cleavage was observed at methionine-80, leucine-94 and leucine-98. Partial tryptic cleavage was observed at lysine-72 and lysine-73, and arginine-91, together with tyrosine-97.

12. Preliminary Sequence Investigation of Arenicola Cytochrome c

Approximately 3 mg (0.25 μmol) of oxidised and denatured Arenicola cytochrome c was equilibrated at 37°C and pH 8.0 under nitrogen. At zero time 2% (w/w) chymotrypsin was added, and a further 1% (w/w) was added after 60 min. The digestion was terminated at 90 min.

The peptides were separated on pH 6.5 electrophoresis, but gave only weak reactions with the ninhydrin reagent. A small number of the stronger peptides were further purified on pH 1.9 electrophoresis. Overall, the digestion was very poor. Only a fraction of the sequence was characterised, and these peptides were placed from comparisons with other cytochromes c. The peptides, together with mobility and sequence data, and tentative placements, are shown in Table 34.

Chymotryptic Peptides

Peptide C1 (11-26) (Heme peptide)

The heme peptide showed characteristic streaking on pH 6.5 electrophoresis, having an overall mobility of about zero, in common with other invertebrate chymotryptic heme peptides. Determination of the N-terminal amino acid by dansyl analysis proved inconclusive through lack of material.

Peptide C2 (47-48) (Ala-Tyr)

Peptide C3 (49-59)

The peptide gave a positive result with the Ehrlich

TABLE 34.

Chymotryptic Peptides from Arenicola Cytochrome c

Peptide/ Position	M (pH 6.5)	M (pH 1.9)	Dansyl-Edman results
C1 (11-26)	0	0.40	(Heme peptide; see text)
C2 (47-48)	0	0.68	Ala-Tyr
C3 (49-59)	0.62	0.68	Thr- (Trp peptide; see text)
C4 (66-74)	0	0.86	Glx-Tyr-Leu-Glx-Asx-Pro-(Lys, Lys,Tyr)
C5 (68-74)	0.86		Leu-Glx-Asx-Pro-Lys-Lys-Tyr
C6 (75-80)	1.10		Ile-Pro-Gly-Thr-Lys-Met
C7 (81-82)	0	0.62	Val-Phe
C8 (83-85)	0	0.60	Ala-Gly-Leu

reagent and digestion of a 40% aliquot of C3 with carboxypeptidase-A for 6 h yielded tryptophan as judged by dansyl analysis. The N-terminal amino acid was determined to be threonine by dansyl analysis, but the dansyl-Edman analysis was inconclusive beyond the N-terminus because of lack of material.

Peptide C4 (66-74) (Glu-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr)

The dansyl-Edman analysis was inconclusive beyond five degradation steps and residues 72-74 were placed from the semi-quantitative amino acid composition. The pH 6.5 electrophoretic mobility of C4 indicated that two acidic residues were present and these were placed at positions 66 and 69 from the assumed homology of Arenicola cytochrome c with other invertebrate cytochromes c.

Peptide C5 (68-74) (Leu-Glu-Asn-Pro-Lys-Lys-Tyr)

The pH 6.5 electrophoretic mobility indicated that a single acidic residue was present and this was placedat position 69 from the assumed homology of Arenicola cytochrome c with other invertebrate cytochromes c.

Peptide C6 (75-80) (Ile-Pro-Gly-Thr-Lys-Met)

Peptide C7 (81-82) (Val-Phe)

Peptide C8 (83-85) (Ala-Gly-Leu)

13. The Amino Acid Composition of Arenicola Cytochrome c

The amino acid composition of Arenicola cytochrome c was determined from three duplicate 50 µg samples hydrolysed for 24, 48, and 72 h respectively, and is shown in Table 35. The molar ratios of the amino acid residues were based on an arbitrary sequence length of 106 residues. Tryptophan was not determined, but the best spectral ratios of the purified cytochrome c indicated that one residue was present.

14. The Amino Acid Composition of Nereis Cytochrome c

The amino acid composition of Nereis cytochrome c was determined from three duplicate 50 µg samples hydrolysed for 24, 48 and 72 h respectively, and is shown in Table 36. The molar ratios of the amino acid residues were based on an arbitrary sequence length of 106 residues. Tryptophan was not determined, but the best spectral ratios of the purified cytochrome c indicated that one residue was present.

15. The Amino Acid Composition of Loligo Cytochrome c

The amino acid composition of Loligo cytochrome c was determined from three duplicate 50 µg samples hydrolysed for 24, 48 and 72 h respectively, and is shown in Table 37. The molar ratios of the amino acid residues were based on the assumption of two histidine residues in the sequence.

TABLE 35.

The Amino Acid Composition of *Arenicola* Cytochrome *c*

	Mean Values 24 h hydrolysis	Mean Values 48 h hydrolysis	Mean Values 72 h hydrolysis	Mean Corrected Values	Amino Acid Analysis
Asp	10.9	11.7	10.3	10.9	11
Thr	7.6	7.3	7.2	7.8*	8
Ser	4.7	4.3	3.8	5.0*	5
Glu	8.9	10.0	8.9	9.3	9
Pro	6.1	6.7	6.0	6.3	6
Gly	12.8	12.5	12.0	12.4	12
Ala	8.6	9.6	9.1	9.1	9
Cysteine	1.4	1.8	2.0	1.7	2
Val	5.6	6.2	6.1	6.0	6
Met	0.9	0.7	0.6	0.7	1
Ile	4.2	4.3	4.4	4.3	4
Leu	6.1	6.7	6.3	6.4	6
Tyr	3.4	3.5	3.3	3.4	3
Phe	3.4	3.5	3.8	3.6	4
His	2.7	2.8	2.5	2.7	3
Lys	15.1	15.7	14.4	15.1	15
Arg	3.4	3.2	3.2	3.3	3
Trp	-	-	-	-	+

The mean values were obtained from the analysis of 2 samples of 50 µg cytochrome *c* at each hydrolysis time. The mean corrected values were obtained as an average of the six determinations with corrections made for the destruction of certain amino acids.

+Tryptophan was not determined; the best spectral ratios of the cytochrome indicate one residue is present.

*Calculated from 24 h and 72 h values assuming first order kinetics for destruction (Moore & Stein, 1963).

TABLE 36.

The Amino Acid Composition of Nereis Cytochrome c

	Mean Values 24 h hydrolysis	Mean values 48 h hydrolysis	Mean values 72 h hydrolysis	Mean Corrected Values	Amino Acid Analysis
Asp	11.1	10.7	11.5	11.1	11
Thr	7.2	7.0	6.6	7.6*	8
Ser	3.4	3.0	3.0	3.7*	4
Glu	8.4	8.1	8.0	8.2	8
Pro	4.3	4.6	4.4	4.4	4
Gly	13.7	13.3	13.3	13.4	13
Ala	7.7	7.6	8.3	7.9	8
Cysteine	1.4	1.9	1.2	1.5	2
Val	7.9	7.5	8.4	7.9	8
Met	0.9	0.7	0.8	0.8	1
Ile	4.3	4.0	4.5	4.3	4
Leu	6.6	6.5	6.2	6.4	6
Tyr	3.6	3.4	3.7	3.6	4
Phe	3.7	3.8	3.6	3.7	4
His	2.0	2.1	2.3	2.1	2
Lys	18.0	18.0	18.2	18.1	18
Arg	3.4	3.3	3.3	3.3	3
Trp	-	-	-	-	+

The mean values were obtained for the analysis of 2 samples of 50 µg of cytochrome c at each hydrolysis time. The mean corrected values were obtained as an average of the six determinations with corrections made for the degradation of certain amino acids.

*Calculated from 24 h and 72 h values assuming first order kinetics for destruction (Moore & Stein, 1963).

+Tryptophan was not determined; the best spectral ratios of the cytochrome indicated one residue is present.

TABLE 37.

The Amino Acid Composition of *Loligo* Cytochrome *c*

	Mean Values 24 h hydrolysis	Mean Values 48 h hydrolysis	Mean Values 72 h hydrolysis	Mean Corrected Values	Amino Acid Analysis
Asp	10.8	11.2	11.6	11.2	11
Thr	5.8	5.6	5.6	5.9*	6
Ser	4.5	4.2	3.8	4.7*	5
Glu	9.2	8.9	9.3	9.1	9
Pro	8.3	8.4	7.6	8.1	8
Gly	16.3	16.0	16.4	16.2	16
Ala	10.3	10.9	9.8	10.3	10
Cysteine	1.6	1.8	1.4	1.6	2
Val	4.9	5.1	5.5	5.2	5
Met	0.9	1.1	0.9	1.0	1
Ile	5.4	5.7	5.9	5.7	6
Leu	5.0	5.4	5.5	5.3	5
Tyr	4.4	4.1	4.5	4.3	4
Phe	3.7	4.0	4 0	3.9	4
His	2.0	2.0	2.0	2.0	2
Lys	15.6	16.5	15.9	16.0	16
Arg	3.1	3.2	3.3	3.2	3
Trp	-	-	-	-	+

The mean values were obtained from the analysis of 2 samples of 50 µg of cytochrome *c* at each hydrolysis time. The mean corrected values were obtained as an average of the six determinations with corrections made for the destruction of certain amino acids.

\+ Tryptophan was not determined; the best spectral ratios of the cytochrome indicate one residue is present.

* Calculated from 24 h and 72 h assuming first order kinetics for determination (Moore & Stein, 1963).

Tryptophan was not determined, but the best spectral ratios of the purified cytochrome c indicated that one residue was present.

16. Calculations based on a matrix of Amino Acid Differences

The numbers of amino acid differences between a phylogenetically representative sample of forty out of seventy determined sequences of cytochrome c, are given in Appendix I and the sequences are shown in Appendices II and III. The mean variations between the members of certain taxonomic groups are given in Table 38.

The "unit evolutionary period" (UEP), has been defined as the average time taken for a single variation to arise between two diverging lines of descent (Zuckerkandl & Pauling, 1965; Margoliash & Smith, 1965). Using a palaeontological estimated time period of 280 million years for the divergence of the mammalian and avian lines of descent (Simpson, 1964) together with the matrix of differences (Appendix I), the UEP has been calculated to be equal to 26.9 million years for a single amino acid substitution. Linear extrapolation from a fixed time point, corrected for the probability of multiple changes at the same codon (Feller, 1950; Margoliash & Smith, 1965), provide estimates for the times of divergence of the major invertebrate phyla from the vertebrate line of descent, and for the divergence of the major kingdoms of organisms (see Whittaker, 1969).

TABLE 38.

Times of divergence calculated for selected groups of organisms.

Groups compared	Average number of variant residues	Times of* divergence (million years)	Corrected times* of divergence (million years)
(A) Animal groups:			
Mammalia-Aves	10.4 ± 1.3 (8)	280+	280+
Vertebrata-Echinodermata (Asterias)	23.5 ± 2.8 (14)	632	680
Vertebrata-Mollusca (Helix)	26.2 ± 2.6 (14)	705	780
Vertebrata-Crustacea (Macrobrachium)	27.4 ± 3.9 (14)	737	800
Vertebrata-Insecta	26.1 ± 3.1 (84)	702	780
Vertebrata-Annelida (Eisenia)	34.0 ± 2.8 (14)	914	1080
(B) Kingdoms:			
Animalia-Plantae	49.2 ± 2.9(150)	1323	1960
Animalia-Fungi	46.9 ± 2.8(125)	1261	2200
Plantae-Fungi	51.6 ± 2.4 (40)	1388	2380
Protista-Plantae	57.4 ± 3.3 (24)	1544	2900

The species considered are those given in the matrix of amino acid differences (Appendix I). For each group of comparisons, the average amino acid difference together with the standard deviation is given.

*Times of divergence were calculated by a linear extrapolation using the "unit evolutionary period" of the mammal/bird divergence. Corrected times of divergence were derived using the formula of Feller (1950) as recommended by Margoliash & Fitch (1968).

+Dating obtained direct from the fossil record.

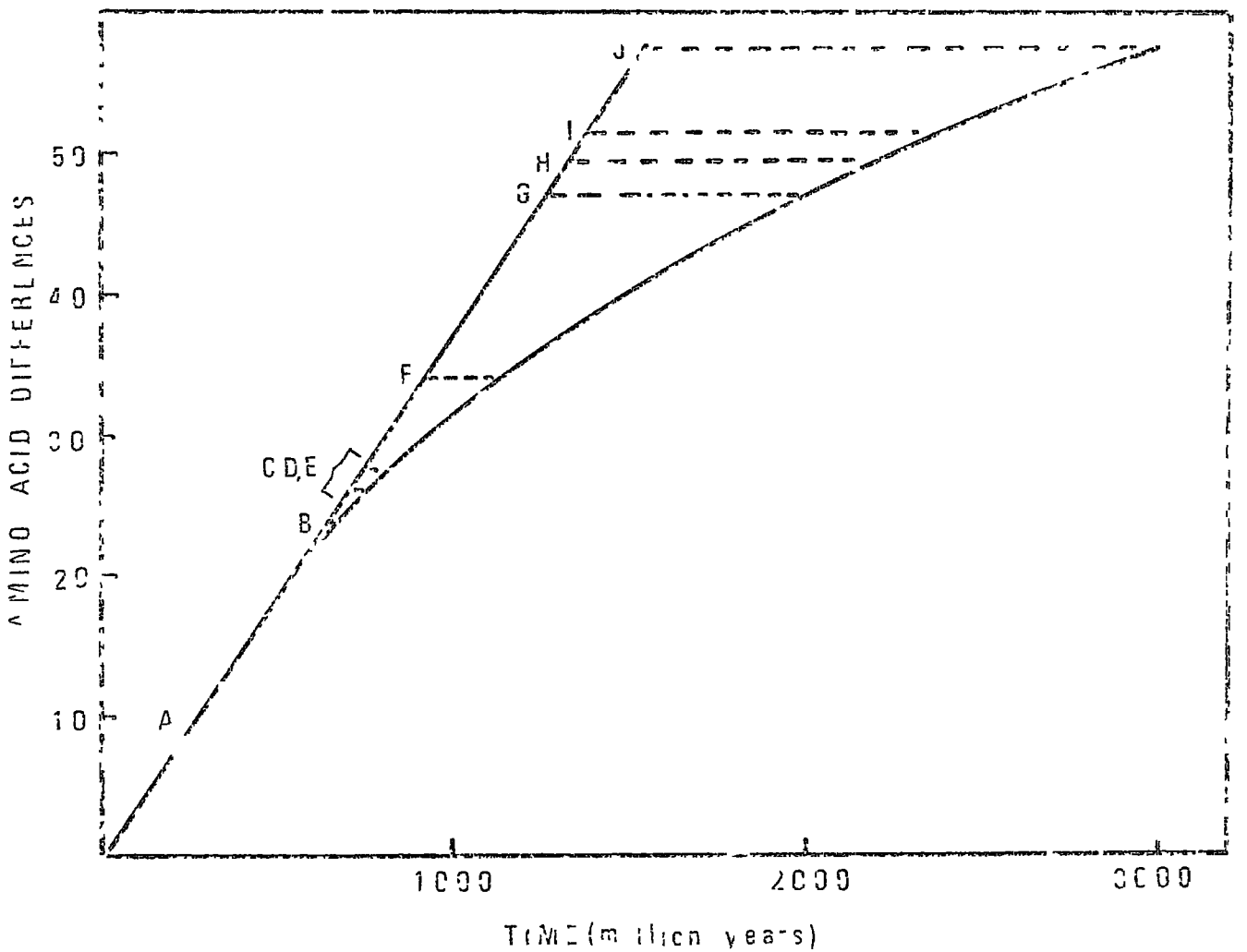

AMINO ACID DIFFERENCES
10
20
30
40
50
A
B
C,D,E
F
G
H
I
J
1000
2000
3000
TIME (million years)

Figure 35 and Table 38 show the Echinodermata to be the most recent invertebrate phylum and the Annelida to be the most primitive, with reference to the species used in this study. The difference in times of divergence of the classes Insecta and Crustacea from the vertebrate line may not be indicative of a diphyletic origin of the Arthropoda, in view of the small number of Vertebrate-Crustacea comparisons available.

17. Molecular Phylogenies constructed using Ancestral Sequence Methods

The ancestral sequence method produced topologies which could be assessed by the total number of substitutions required to relate the amino acid sequences included in a particular topology. From the phylogenetic point of view, the best topology was considered to be that for which this total was a minimum but, because the number of substitutions was an integral value, it was possible that no unique minimum existed, i.e. several phylogenies may be obtained having a particular minimum value.

The time required to build a phylogeny relating the seventy known cytochrome c sequences makes such a computation unjustified. A better approach was to compute phylogenies on the basis of intra-kingdom differences and to relate the kingdoms in a similar fashion using representative sequences from each kingdom (see Whittaker, 1969; Dayhoff, 1972;

McLaughlin & Dayhoff, 1973).

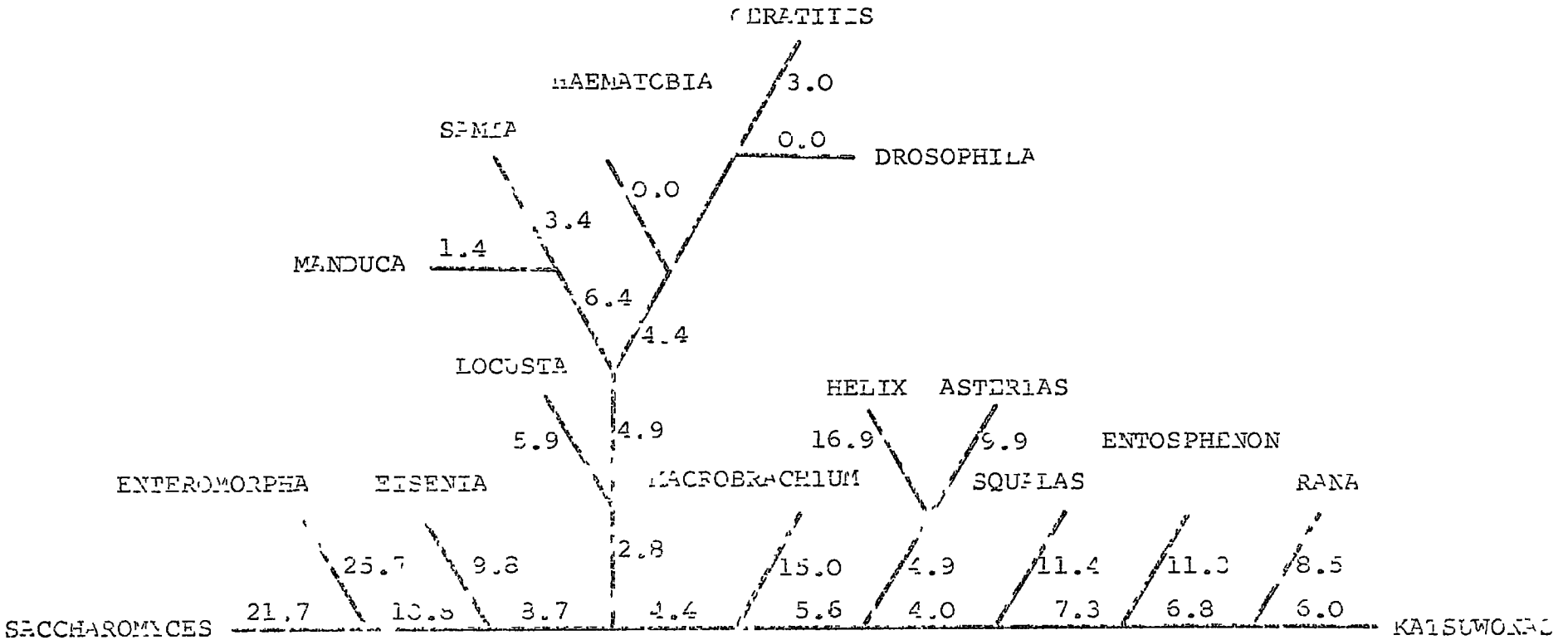
The aim of the present study was to relate the known invertebrate cytochromes c with each other, and with selected "fixing" species taken from the Fungi, Plantae and vertebrate members of the Animalia. A total of ten complete invertebrate cytochrome c sequences were used in the study, together with a sequence from each of the Fungi and Algae and four sequences from the Vertebrata. These are listed in Table 39. The treatment of the differences in length, due to the N-terminal "tails" in the standard cytochrome c alignment was not always the same (see below, Figures 36-39). Absent residues at the C-terminus of the alignment were regarded as deletions, and were computed as differences.

Figure 36 shows the best phylogeny obtained relating the amino acid sequences listed in Table 39 and Appendix II. The minimum of amino acid substitutions relating the sequences was 228 and no equal alternatives to this figure were found. All the insect sequences were grouped together on a single line of descent diverging from the main vertebrate line. The arrangement of the Orders Diptera (Haematobia, Ceratitis and Drosophila), Lepidoptera (Samia and Manduca) and Orthoptera (Locusta) was clear-cut, but the position of the crustacean, Macrobrachium, was suggestive of at least a diphyletic origin of the Arthropoda. The position of the Eisenia sequence indicated that the Annelida were the most primitive of the invertebrate phyla used in this study. The grouping of the

TABLE 39.

The Source Species of the Cytochromes *c* used in the Building of Molecular Phylogenies (see Figs. 36-42)

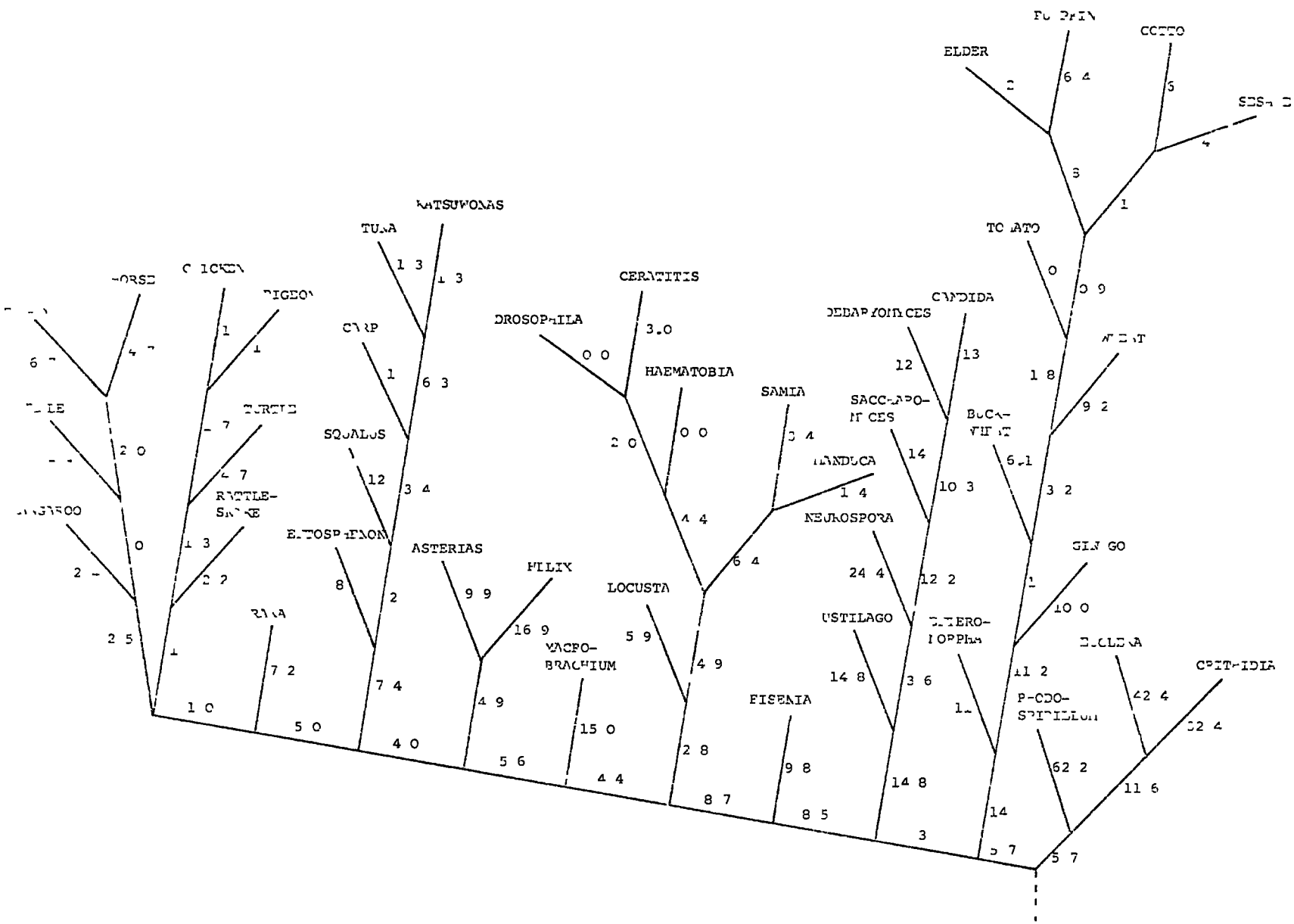
Species	Common Name	Kingdom	Phylum	Class	Order
Saccharomyces oviformis	bakers yeast	Fungi	-	Ascomycetes	-
Enteromorpha intestinalis	-	Plantae	Chlorophyta	Chlorophyceae	Ulvales
Eisenia foetida	brandling worm	Animalia	Annelida	Oligocheata	Tericola
Manduca sexta	tobacco horn-worm moth	"	Arthropoda	Insecta	Lepidoptera
Samia cynthia	silkworm moth	"	"	"	"
Drosophila melanogaster	fruit fly	"	"	"	Diptera
Haematobia irritans	screw-worm fly	"	"	"	"
Ceratitis capitata	Mediterranean fruit fly	"	"	"	"
Locusta gregaria	common locust	"	"	"	Orthoptera
Macrobrachium malcomsonii	king prawn	"	"	Crustacea	Decapoda
Helix aspersa	garden snail	"	Mollusca	Gastropoda	Stylommatophora
Asterias rubens	common starfish	"	Echinodermata	Asteroidea	Forcipulata
Entosphenon tridentatus	Pacific lamprey	"	Chordata	Agnatha	Petromysontiformes
Squalus suckli	puget sound dog-fish	"	"	Chondrichthyes	Squaliformes
Katsuwonas vagrans	bonito	"	"	Osteichthyes	Perciformes
Rana catesbiana	bullfrog	"	"	Amphibia	Anura

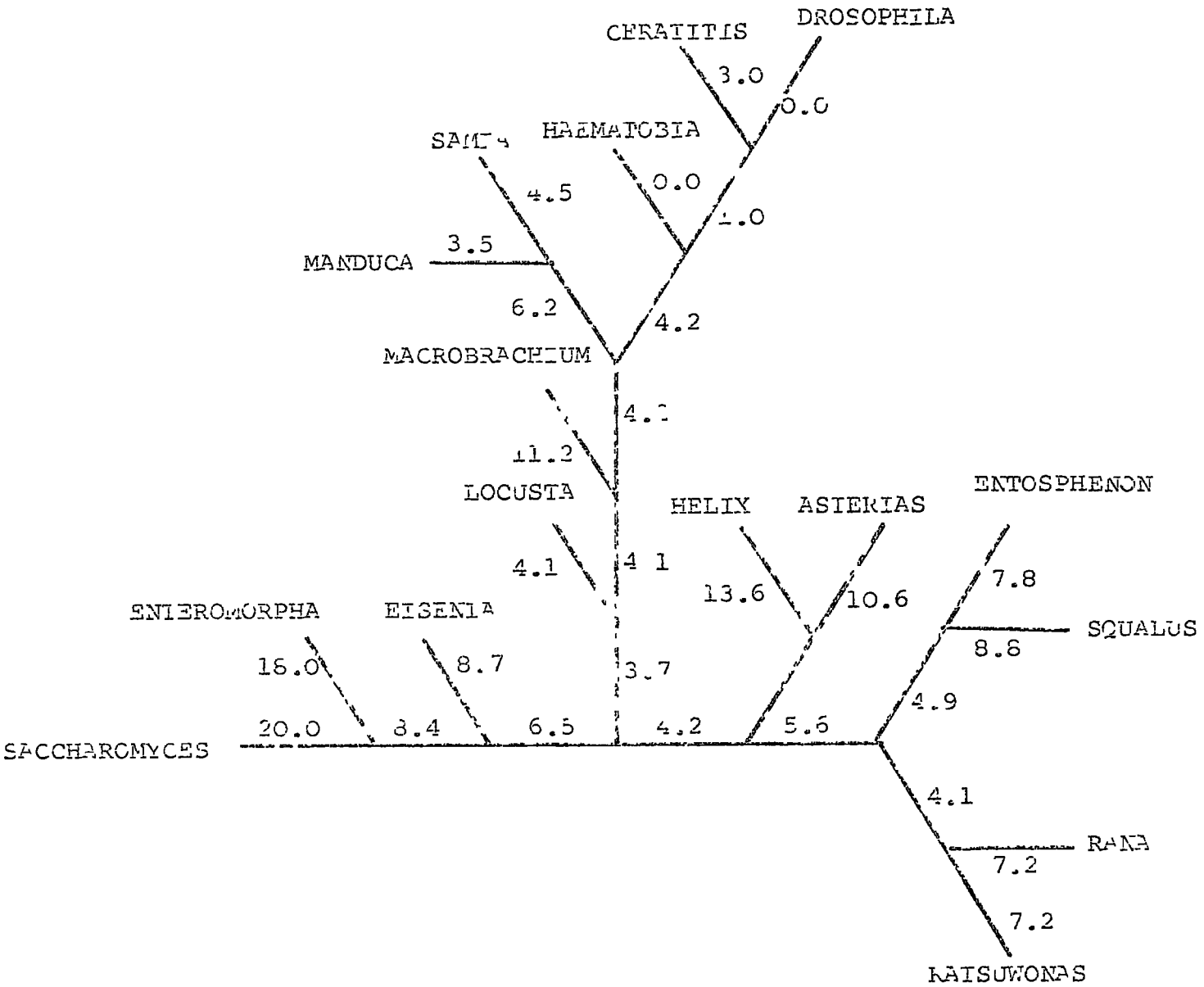

CERATITIS
HAEMATOBIA
3.0
SEMIA
0.0
DROSOPHILA
0.0
3.4
MANDUCA
1.4
6.4
4.4
LOCUSTA
HELIX
ASTERIAS
5.9
4.9
16.9
9.9
ENTOSPHENON
ENTEROMORPHA
EISENIA
MACROBRACHIUM
SQUILAS
RANA
2.8
25.7
9.8
15.0
4.9
11.4
11.0
8.5
SACCHAROMYCES
21.7
10.8
8.7
4.4
5.6
4.0
7.3
6.8
6.0
KATSUWONUS

sequences of Helix and Asterias on a common node was unexpected, being members of the phyla Mollusca and Echinodermata, respectively. However, this may have been due to the lack of sequence information from these phyla.

Figure 37 shows an overall phylogeny relating the cytochrome c sequences listed in Appendix II. The tree relating the invertebrates (Figure 36) was grafted on to a tree constructed by Meatyard (1974) using the intra-kingdom/inter-kingdom approach previously described (see above). Similar trees have been produced by Dayhoff (1972) and McLaughlin & Dayhoff (1973).

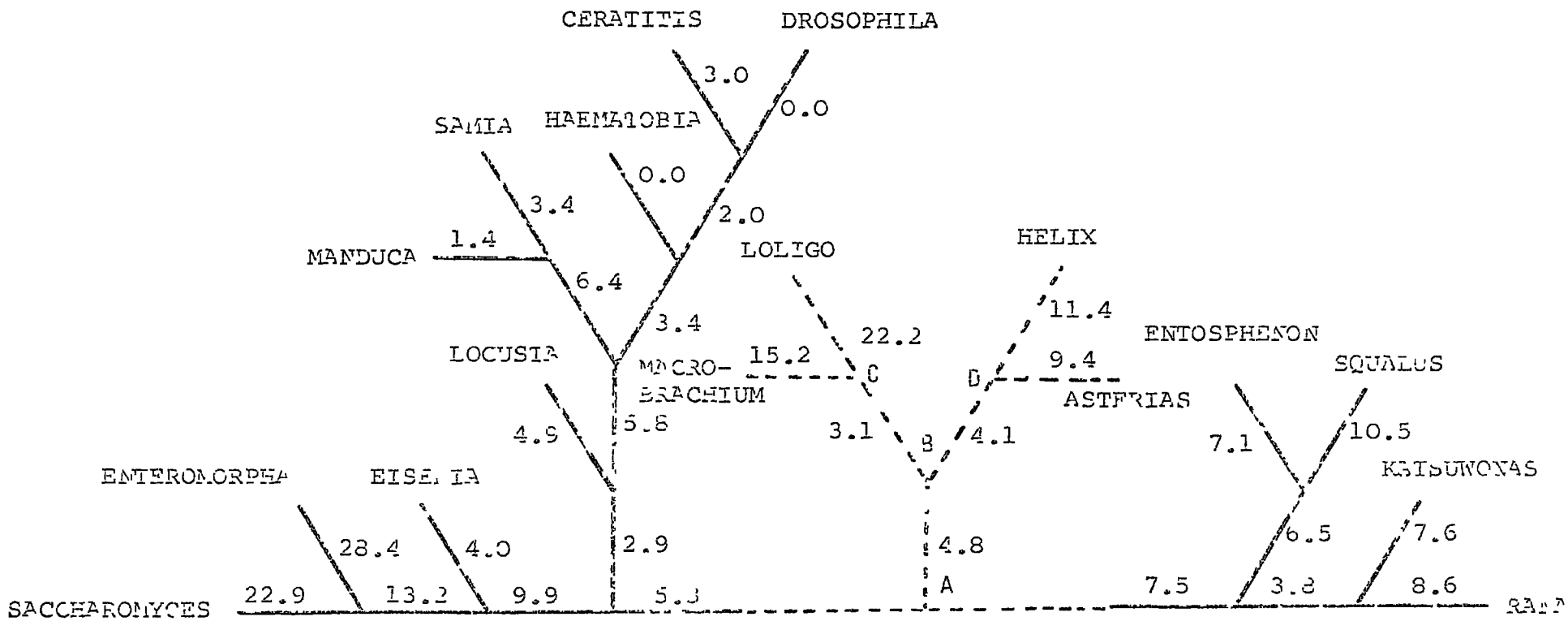
Figure 38 shows the best phylogeny obtained relating the sixteen cytochrome c sequences listed in Table 39, taking no account during the computer analysis, of the residues and/or blank positions in the sequences occurring prior to the invariant glycine-1 in the standard alignment. The minimum number of amino acid substitutions relating the sequences in this phylogeny was 207 and no equal alternatives to this figure were found. A comparison with Figure 36 demonstrates that the N-terminal "tail" was critical only in the fixing of the Macrobrachium sequence to the tree and that the clear-cut grouping of the annelid and insect sequences in Figure 36 was not merely due to the possession of similar N-terminal "tails". The computer analysis of regions of the cytochrome c molecule common to all the species, also served to reduce

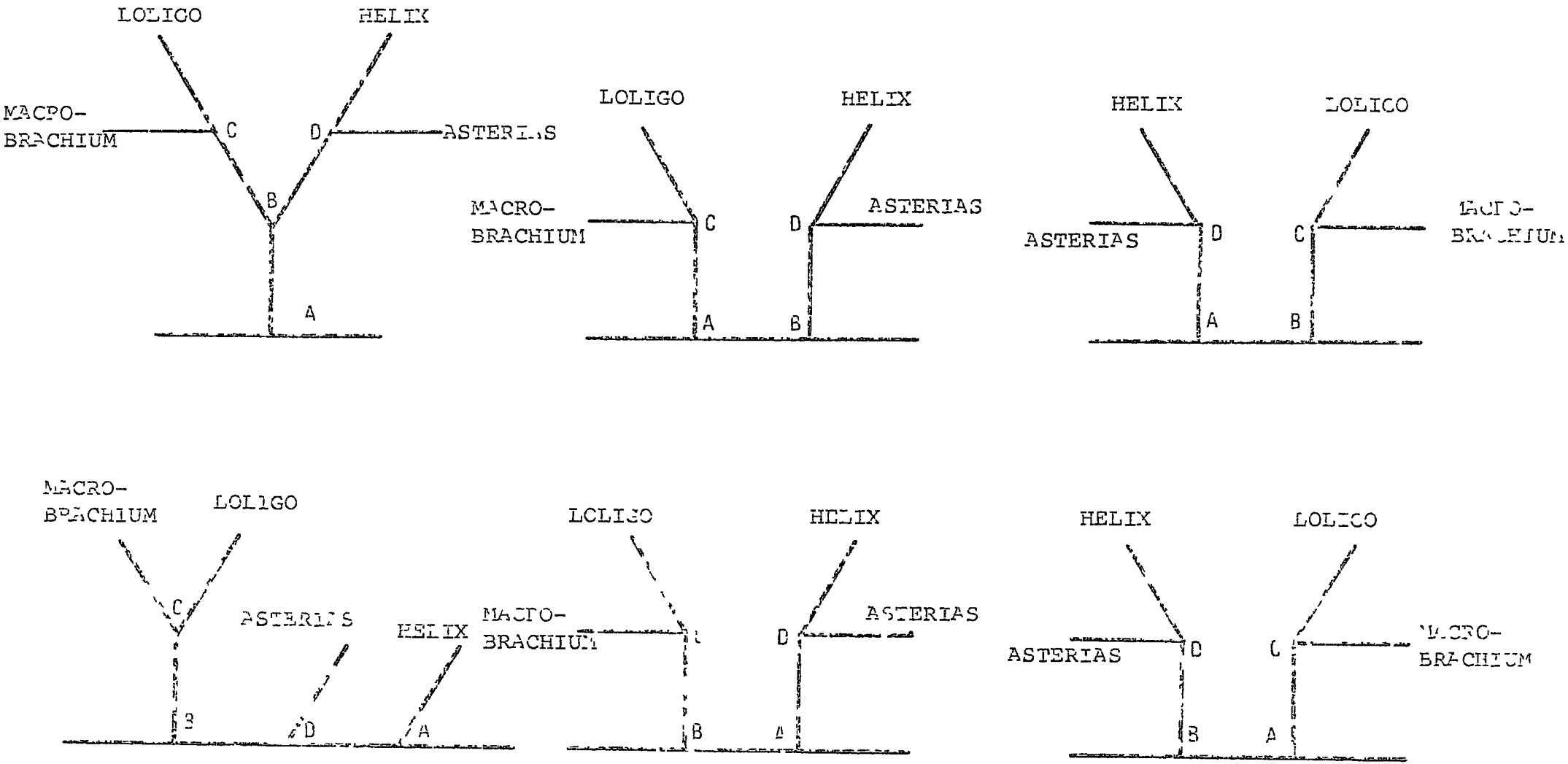

HORSE
CHICKEN
PIGEON
TUNA
KATSUWONAS
CARP
SQUALUS
RATTLE-SNAKE
ELOPS PHALON
ASTERIAS
HILIX
RANA
VACFO-BRACHIUM
CERATITIS
DROSOPHILA
HAEMATOBIA
SAMIA
LOCUSTA
FISENIA
CANDIDA
DEBARYOMYCES
SACCHARO-MYCES
NEUROSPORA
USTILAGO
BOTERO-TOPPHA
BUCA-RINAT
TOMATO
GINKGO
P-CDO-SPITILLUM
CRITHIDIA
ELDER


CERATITIS
DROSOPHILA
3.0
0.0
SAM[illegible]
HAEMATOBIA
4.5
0.0
1.0
MANDUCA
3.5
6.2
4.2
MACROBRACHIUM
4.1
11.2
LOCUSTA
HELIX
ASTERIAS
ENTOSPHENON
4.1
4 1
13.6
10.6
7.8
ENTEROMORPHA
EISENIA
SQUALUS
8.8
16.0
8.7
3.7
4.9
20.0
8.4
6.5
4.2
5.6
SACCHAROMYCES
4.1
RANA
7.2
7.2
KAISUWONAS

the majority of the branch-lengths shown in Figure 38 as compared to Figure 36. Because of the nature of the data input which gave rise to Figure 36, Asterias was considered to be ten amino acid substitutions away from Enteromorpha just because of the latter's N-terminal tail, and this may represent a distortion of the true relationship between these two sequences (see Discussion V).

Figure 39 shows a phylogeny relating the sixteen complete sequences listed in Table 39 and Appendix II, together with the partial sequence of Loligo opalescens cytochrome c (Boulter, M. E., unpublished experiments). The undetermined portion of the Loligo sequence was computed as amino acid differences (see Appendix III). The phylogeny shown was related by a minimum of 252 amino acid substitutions, but five other equal alternatives were obtained differing only in the arrangement of Helix, Macrobrachium, Asterias and Loligo sequences and the nodes A-D in Figure 39. The alternatives are summarised in Figure 40. The arrangement of the sequences in Figure 39 was essentially identical to that obtained without the Loligo sequence and shown in Figure 36. The Loligo sequence unexpectedly shared a node with Macrobrachium and not with the other mollusc, Helix, the Helix-Asterias group remaining constant throughout all the ancestral sequence methods (Figures 36-40). A consideration of the large branch-lengths and the small number of common residues upon


CERATITIS
DROSOPHILA
3.0
0.0
SAMIA
HAEMATOBIA
0.0
3.4
2.0
1.4
MANDUCA
LOLIGO
HELIX
6.4
11.4
3.4
22.2
ENTOSPHENON
LOCUSTA
15.2
9.4
SQUALUS
MACRO-
BRACHIUM
C
D
ASTERIAS
4.9
5.8
3.1
B
4.1
7.1
10.5
ENTEROMORPHA
EISENIA
KRISUWONAS
28.4
4.0
2.9
4.8
6.5
7.6
22.9
13.2
9.9
5.3
A
7.5
3.8
8.6
SACCHAROMYCES
RAT


LOLIGO
HELIX
MACRO-
BRACHIUM
C
D
ASTERIAS
B
A
LOLIGO
HELIX
MACRO-
BRACHIUM
C
D
ASTERIAS
A
B
HELIX
LOLIGO
ASTERIAS
D
C
A
B
MACRO-
BRACHIUM
LOLIGO
C
ASTERIAS
HELIX
B
D
A
LOLIGO
HELIX
MACRO-
BRACHIUM
C
D
ASTERIAS
B
A
HELIX
LOLIGO
ASTERIAS
D
C
B
A

which the method selected this phylogeny, suggested that the lack of mollusc and echinoderm sequence data, together with the computation as differences of the undetermined Loligo residues, may well account for the instability of this area of the tree.

The sixteen cytochrome c sequences listed in Table 39 and Appendix II, were used to investigate the efficiency of the shuffling procedure of the ancestral sequence computer method (procedure 3, see Materials and Methods IV,2.). A number of topologies were built using different orders of sequences and they were assessed in the usual manner at the end of procedures 1 and 2 and at the end of procedure 3. The efficiency of procedure 3 was judged by its ability to sort out the best phylogeny (i.e. that for which the number of amino acid substitutions was a minimum), regardless of the building order of the original phylogeny. The results shown in Table 40 indicate that procedure 3 was at least 60% efficient in obtaining the best phylogeny relating a particular group of sequences of cytochrome c.

18. Molecular Phylogenies constructed using Numerical Methods

Phylogenies were constructed from a matrix of amino acid differences using the cytochrome c sequences listed in Table 39 and Appendix II, together with a partial sequence of Loligo cytochrome c (see Appendix III). The method used was

TABLE 40.

The Building Orders of Sequences used to Estimate the Efficiency of Computer Procedure 3 of the Ancestral Sequence Method.

Build order	Sequence build order (8 expts.)							
	1	2	3	4	5	6	7	8
1	ENT	ENT	CER	TBC	LCS	CER	DGF	SCC
2	SCC	SCC	SQD	PRN	CER	DGF	STF	ENT
3	BNT	WMN	BNT	SCR	BLF	ENT	SMN	SQD
4	WMN	LCS	WMN	BNT	WMN	LMP	BLF	WMN
5	STF	CER	TBC	FRT	DGF	SNL	ENT	SNL
6	SNL	FRT	DGF	BLF	PRN	TBC	TBC	PRN
7	SQD	SCR	ENT	CER	FRT	PRN	LCS	LMP
8	PRN	TBC	FRT	DGF	ENT	BNT	PRN	STF
9	LMP	SMN	SNL	LCS	LMP	FRT	BNT	SMN
10	DGF	SQD	PRN	LMP	STF	BLF	SNL	DGF
11	BLF	SNL	LCS	WMN	BNT	LCS	WMN	TBC
12	FRT	LMP	SMN	SNL	SCC	SCC	LMP	FRT
13	SMN	DGF	BLF	SCC	TBC	SCR	SCR	BNT
14	SCR	BLF	SCC	SQD	SCR	SMN	CER	CER
15	TBC	BNT	SCR	ENT	SMN	SQD	FRT	SCR
16	LCS	PRN	LMP	SMN	SNL	STF	SCC	BLF
17	CER	STF	STF	STF	SQD	WMN	SQD	LCS
*Tm 1,2 =	252	252	264	253	253	256	257	255
*Tm 3 =	252	252	252	252	252	255	254	254

Abbreviations

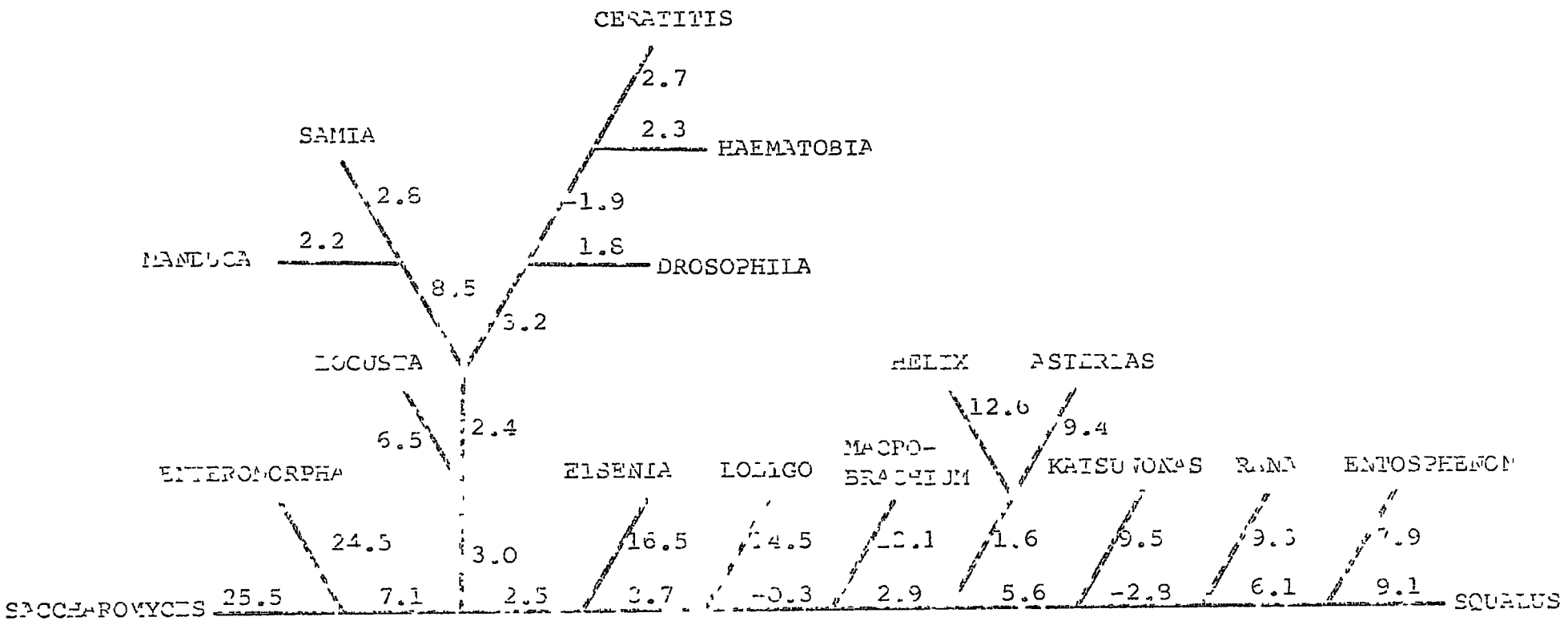
ENT = Enteromorpha
SCC = Saccharomyces
WMN = Eisenia
TBC = Manduca
SMN = Samia
FRT = Drosophila
SCR = Haematobia
CER = Ceratitis
LCS = Locusta
PWN = Macrobrachium
SNL = Helix
SQD = Loligo
STF = Asterias
LMP = Entosphenon
DGF = Squalus
BNT = Katsuwonas
BLF = Rana

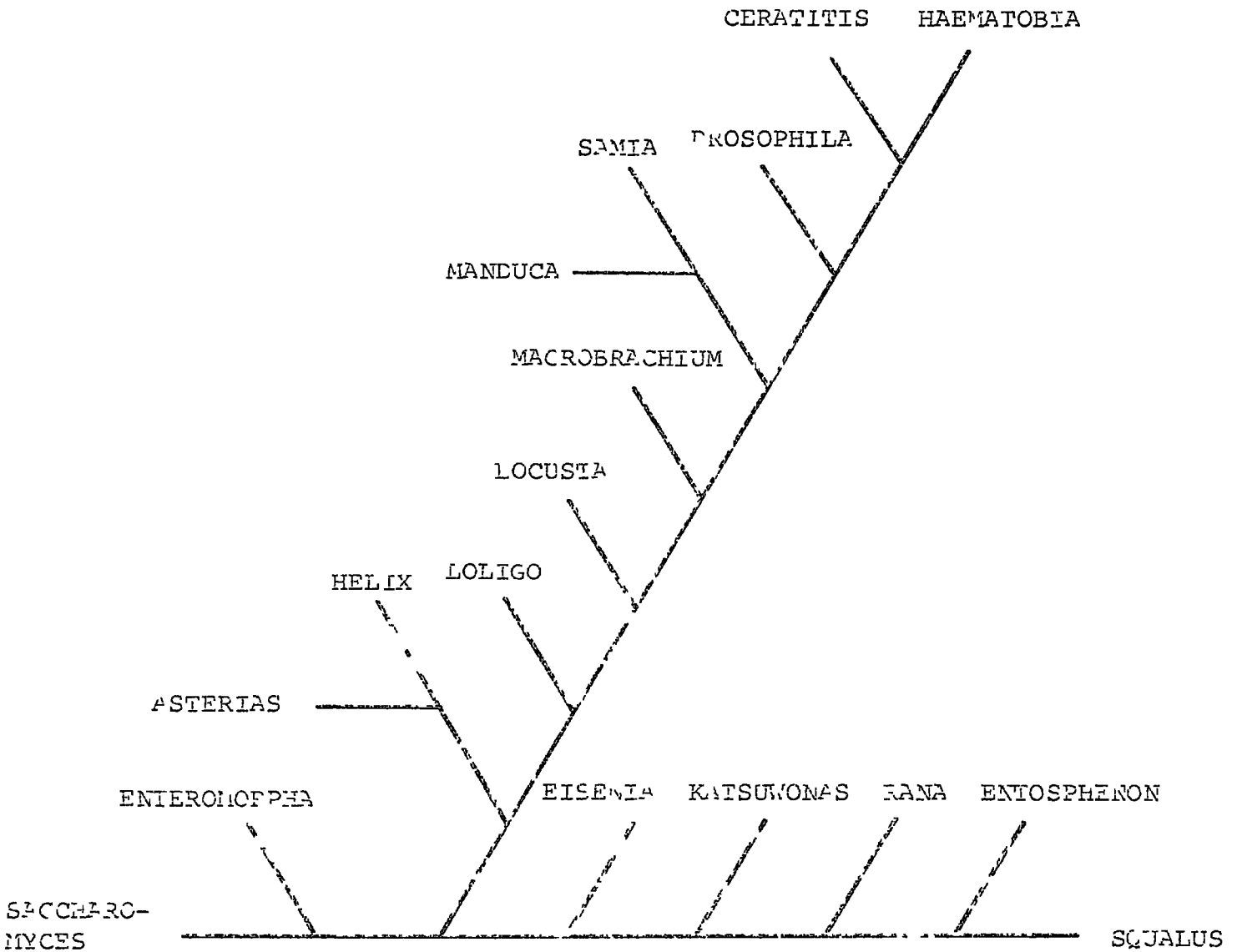
*Tm 1,2, Tm 3 refer to the total minimum number of amino acid substitutions relating the phylogenies produced by the building and shuffling procedures.

that of Moore et al. (1973) developed from the approaches of Fitch & Margoliash (1967a) and phylogenies were selected for which the Moore Residual Coefficient (MRC) was a minimum (Moore et al., 1973). Differences in the length of sequences, due to the N-terminal "tails" in the standard alignment, were treated in two ways (see below), whilst absent residues at the C-terminus of the alignment were regarded as deletions and were computed as differences. An advantage of the matrix method over the ancestral sequence method, is that sequences are computed regarding undetermined residues as unknowns, of which no account is taken, rather than as blanks which are computed as differences. The undetermined portion of the Loligo sequence was treated in this manner.

Figure 41 shows a minimum phylogeny relating the sixteen sequences and the partial sequence of Loligo. The grouping of the insect sequences and those of Helix and Asterias were identical to that consistently obtained using the ancestral sequence method (see Figures 36-40), but Macrobrachium was fixed more distantly from the insects than in the ancestral sequence method. The matrix method failed to group the mollusc sequences Helix and Loligo and placed Eisenia as being less primitive than the insect line of divergence.

Figure 42 shows a minimum phylogeny relating the sixteen sequences listed in Table 39, together with the partial sequence

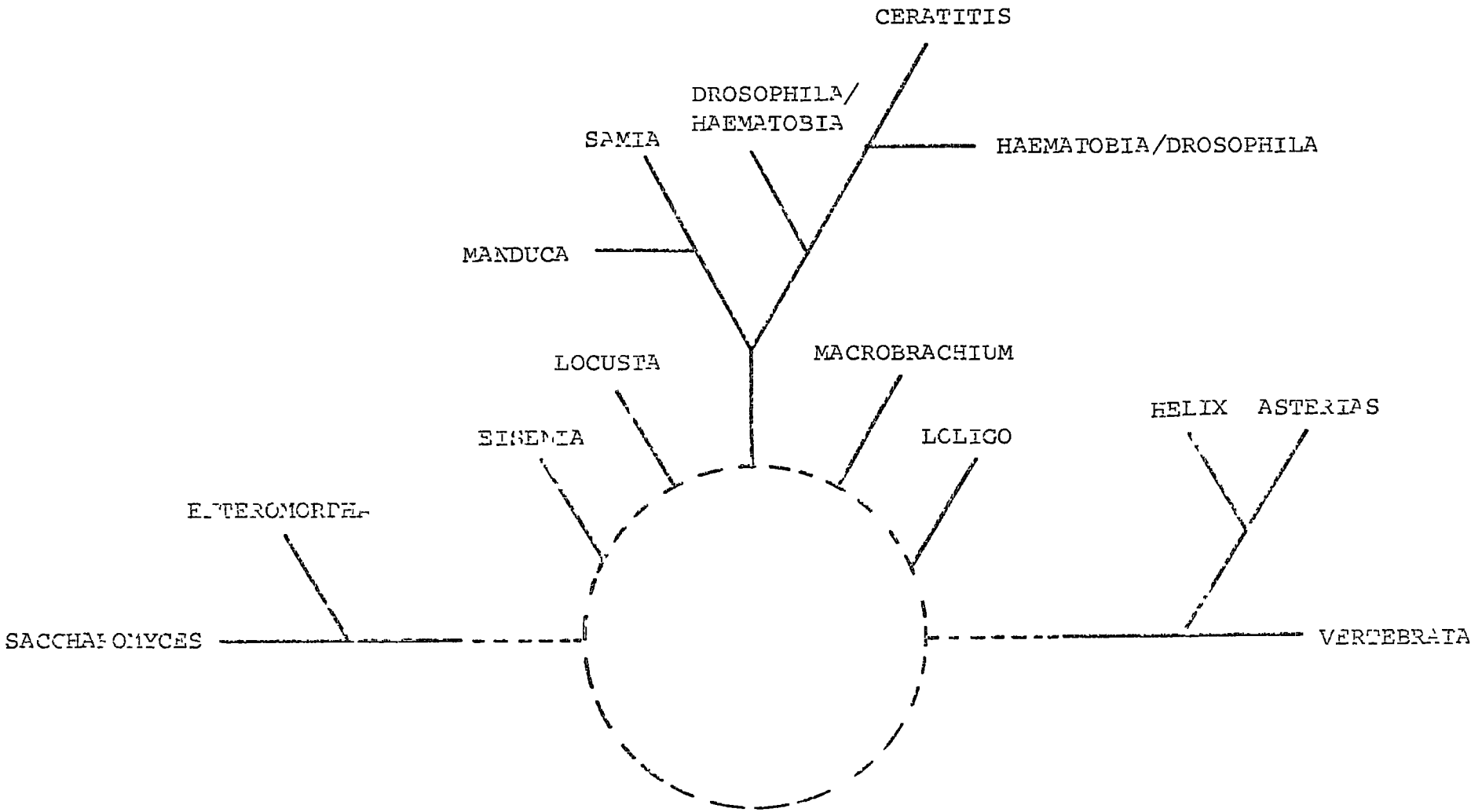

CERATITIS
2.7
2.3
HAEMATOBIA
SAMIA
2.8
1.9
2.2
1.8
MANDUCA
DROSOPHILA
8.5
3.2
LOCUSTA
HELIX
ASTERIAS
12.6
9.4
2.4
6.5
MACRO-
BRACHIUM
KATSUWONUS
RANA
ENTOSPHENON
ENTEROMORPHA
EISENIA
LOLIGO
16.5
24.5
3.0
1.6
9.5
9.3
7.9
25.5
7.1
2.5
3.7
-0.3
2.9
5.6
-2.8
6.1
9.1
SQUALUS


CERATITIS
HAEMATOBIA
SAMIA
DROSOPHILA
MANDUCA
MACROBRACHIUM
LOCUSTA
HELIX
LOLIGO
ASTERIAS
ENTEROMORPHA
EISENIA
KATSUWONAS
RANA
ENTOSPHENON
SACCHARO-
MYCES
SQUALUS

of Loligo cytochrome c. No account was made of the residues and/or blank positions occurring prior to the invariant glycine-1 in the standard alignment (see Appendices II and III). The phylogeny showed a distinctive invertebrate line of descent with only the Eisenia sequence standing apart. The positions of Macrobrachium in Figures 41 and 42 paralleled those obtained from the ancestral sequence method treating the N-terminal "tails" in the same way (see Figures 36 and 38). The sequences of Helix and Asterias were related by a common node, and the Loligo sequence was fixed close to the Arthropod line of descent.

19. A Summary of the Results of Phylogenetic Analysis

The results obtained from the various approaches are summarised in Figure 43. Only the Diptera and Lepidoptera of the insect sequences and the Helix-Asterias grouping maintained stable relationships throughout all of the analytical procedures. The relative merits of these treatments and the selection of the most significant phylogenies are discussed later (see Discussion V-VII).


CERATITIS
DROSOPHILA/
HAEMATOBIA
SAMIA
HAEMATOBIA/DROSOPHILA
MANDUCA
LOCUSTA
MACROBRACHIUM
HELIX
ASTERIAS
SINEMIA
LOLIGO
E_TEROMORPH-
SACCHAROMYCES
VERTEBRATA

DISCUSSION

I. The Methods of Extraction and Purification of Invertebrate Cytochrome c

Pilot experiments with Holothuria, Mytilus, Arenicola and Nereis using the extraction procedure of Richardson et al. (1970; 1971), showed that the yields of cytochrome c obtainable from the majority of invertebrate sources were likely to be as low as those reported for germinating seedlings (Richardson, M., personal communication).

Previously, cytochromes c from invertebrate sources have been extracted by a number of different methods. Ghirretti (1956; 1959) studied the terminal respiration chain in the squid, Aplysia, and the octopus, Octopus vulgaris, by obtaining a particulate preparation using a dilute salt extraction. Yamanaka and co-workers purified cytochrome c from the protochordate, Styela plicata, from the larvae, pupae and adults of the housefly, Musca domestica, from the molluscs, oyster and squid, from the crustacean, king prawn, and the marine worm, Dendrostomum zostericolum, by using a method that included homogenisation in phosphate buffer, followed by ammonium sulphate fractionation and ion-exchange chromatography (see Yamanaka, 1966).

Claims of yields of 3 mg from 15 g of Dendrostomum, 2-3 mg/kg for oyster, squid and prawn muscle, 4 mg/kg for Styela internal organs, and 128 mg/kg for Musca domestica were made,

but no evidence was given of the purity of the preparations (Yamanaka et al., 1963; 1964a,b; Yamanaka & Kamen, 1967). Pettigrew (1972) and Hill et al. (1971) purified cytochrome c from species of protozoan trypanosomatids and the nematode Ascaris lumbricoides, by using a method which included homogenisation in dilute salt solution, precipitation with ammonium sulphate and gel filtration on Sephadex G-75. This method gave yields of 5-40 mg/kg wet weight of starting material.

Chain & Margoliash (1966) reported an extraction of cytochrome c from the moth, Samia cynthia, by using homogenisation in aluminium sulphate, adsorption on Amberlite IRC-50 resin and ammonium sulphate fractionation which gave a yield of 220 mg/kg with an $E_{550}^{C2+}/E_{280}^{C3+}$ ratio of 1.25. Brown et al. (1972) applied the method developed for germinating seedlings (Richardson et al., 1971a) to the snail, Helix aspersa, and obtained a yield of 20 mg/kg with an $E_{550}^{C2+}/E_{280}^{C3+}$ ratio of 1.14. Margoliash & Walasek (1967) homogenised moth thoraxes, Drosophila and screw-worm flies in aluminium sulphate, concentrated the cytochrome c on Amberlite resins, precipitated it with ammonium sulphate and further purified the protein by ion-exchange chromatography and crystallisation. Diano & Martinez (1971) purified cytochrome c directly from animal mitochondria isolated using zonal centrifugation.

The methods outlined above generally used less than 5 kg of starting material, but for this study it was necessary to design methods that could best handle the large amounts of

starting material and homogenate generated in the early stages of the purification. Viewing the reduced spectrum of cytochrome c with a hand, direct-vision spectroscope is only a sensitive assay at concentrations greater than 1 mg/l, and this fact, together with a consideration of the time factor, made the reduction of batch sizes impracticable. In the end, a compromise was made between batch size and time factors in the application of a method which combined points taken from those of Margoliash & Walasek (1967) and Richardson et al. (1970; 1971a). Speed of handling and reduced temperatures were recommended by Margoliash & Walasek (1967) as the best practical precautions to mitigate against yield reductions due to thermal instability and the action of degradative enzymes, so that the major problems encountered with 5-30 kg batches of material during this study were largely technological.

Previous investigations of cytochrome c have involved rather drastic methods of protein extraction, including the use of organic solvents (Goddard, 1944; Hagihara et al., 1958; 1959; Wasserman et al., 1963; Fridman et al., 1968) or with hot acid (Goddard, 1944), and the majorities of inhomogeneities in preparations have been shown to result from such treatments (Margoliash, 1954a,b; Yamanaka et al., 1959). The use of organic solvents can cause denaturation (Kaminsky & Davison, 1969), extremes of pH and temperature may lead to denaturation, deamidation and polymer formation and, generally

speaking, cytochrome c has been shown to be less stable than was previously thought (Margoliash & Schejter, 1966; Lemberg & Barrett, 1973). The homogenisation of plant material in dilute phosphate buffer has been shown to give an efficient extraction of cytochrome c (Morita & Ida, 1968; Richardson et al., 1970), and this method has been applied to invertebrate materials (Brown et al., 1972). The method employed in this study was to homogenise the material in 10 mM--aluminium sulphate solution (after Margoliash & Walasek, 1967) and generally this produced a less slimy extract than test extractions using dilute phosphate buffer.

The point frequently overlooked in a discussion of the relative advantages of extraction media, is the degree of tissue comminution and cell breakage obtained. Meatyard (1974) reported that microscopic examination of the green alga Enteromorpha after homogenisation in a 5 l Waring blendor, run at maximum speed for 3 min, showed less than 5% cell breakage. A similar investigation in this study gave a value of less than 20% cell breakage in an Eisenia homogenate. It was thought that the exoskeletons of the arthropods and echinoderms would be advantageous in acting as an abrasive agent, as had been reported for the snail shells during that cytochrome c extraction (Brown, R. H., personal communication), and to this end the material was homogenised in two volumes of aluminium sulphate and one volume of ice, which was to act as an abrasive agent and a buffer against localised temperature

increases due to prolonged running of the Waring blendor.

Except for Macrobrachium, the whole organism was homogenised because of the practical difficulties involved in selecting and dissecting out particular tissues. The extraction of Macrobrachium was performed on the muscular tails. The larger specimens such as the starfish were not pre-minced or cut up in any way because of the size of the batches, and the Waring blendor was capable of producing a fairly standard homogenate, regardless of the size of the starting material.

The majority of the materials were stored at -20°C after collection, but contrary to the view of Margoliash & Walasek (1967), it was found that the more quickly frozen material was used, the better the yield of cytochrome c. The freezing of material led to a thawing problem prior to homogenisation. For example, a 20 kg batch of frozen Eisenia thawed over a period of 8 h at room temperature, and then subsequently extracted, contained only negligible quantities of cytochrome c. This low yield was attributed in part to the prolonged thawing period, during which the proteases of these animals would be expected to be active, and subsequently, extractions of cytochrome c were performed using fresh, unfrozen samples of Eisenia. It is felt to be significant that the only non-insect invertebrate to yield more than 20 mg/kg in this laboratory was Helix aspersa, where the extraction was performed on fresh, unfrozen materials (Brown et al., 1972).

The starting material for the other extractions consisted of blocks of frozen animals which were broken into small pieces with a hammer and chisel and thawed rapidly over warm water at 4°C.

All the homogenates, except those of Asterias, were stirred at pH 4.5 and 4°C for 2 h, before filtration through two layers of muslin and refiltered by using a Broadbent 21" centrifuge. Because of the high concentration of calcium carbonate, the acid extraction of Asterias was avoided, and the extraction was performed at pH 6.5. Tests showed that the re-extraction of residues obtained from these steps was not worthwhile in terms of yield increase. In general, all but the Locusta and Macrobrachium extracts filtered very slowly on the Broadbent centrifuge, the slimy extracts having a 'waterproofing' effect on the Terylene centrifuge bag, such that only frequent bag changes could maintain the filtration rates. The extracts of Asterias and Eisenia required further centrifugation at pH 8.0.

The large volumes and low cytochrome c content of the extracts caused a major problem of concentration in the early stages of the purification. Yamanaka et al. (1964a) showed that the ability of cation exchangers to adsorb cytochrome c provided an efficient method of concentrating very dilute cytochrome c solutions obtained from low yield sources. Amberlite CG-50 resin in the NH_4^+ form was used both in columns and batchwise in this study; the method chosen being dependent

on the efficiency of the previous filtration step. The most efficient and rapid method was to pack a sintered glass funnel with Amberlite CG-50 and to pull the solution through this bed of resin using the suction of a water pump. Tests on the filtrate of an ox heart homogenate gave estimates of 95% retention of cytochrome c by this method at flow rates of 5 l per h. Clearly, there is a danger of impurity competition in total organism preparations, and this was shown to occur by the use of the batch method, where a second batch performed on an extract was found to yield more cytochrome c than the first batch. To offset these difficulties, filtrates were routinely recycled through fresh Amberlite resin. It was observed that precipitation of materials occurred on and within the resin, so that the elution step was performed batchwise following extensive washing of the resin with distilled water The screening of these washings showed that they contained negligible quantities of cytochrome c. The pH was maintained at a value of 8.0 during the elution to prevent irreversible binding of the cytochrome c to the resin (Boardman & Partridge, 1953; 1954; 1955), and it was found that a concentration of 2 M-sodium chloride was necessary to elute the majority of the cytochrome c from the resin, although 0.5 M-salt was considered sufficiently strong enough for plant cytochromes c (Richardson et al., 1970; 1971a). The use of such a strong eluant created a dialysis problem, but no side effects were

apparent from using running tap water for the first 8 h, followed by the appropriate pH 8.0 buffer. Dialysis residues generally collected a heavy white precipitate as the ionic strength decreased, and this was removed by centrifugation.

Batchwise elution of the cation-exchange resin still gave relatively large volumes of eluate, and CM-Sephadex proved the most satisfactory method of concentration, giving a good adsorption efficiency at good flow rates. The invertebrate preparations invariably formed a thick white precipitate on the top of the CM-50 Sephadex columns and cytochrome c was occasionally seen to be associated with this deposit. Elution and dissociation of the cytochrome c was slow using 0.5 M-sodium chloride, but stronger salt tended to break up the column of resin. The step was repeated with a smaller column to produce a manageable volume of cytochrome c solution.

Ammonium sulphate fractionation procedures have been applied at some stage in the preparations of invertebrate, vertebrate and plant cytochrome c (Brown & Boulter, 1966; Yamanaka, 1966; Margoliash & Walasek, 1967), although early workers had considered that the method may cause deamidation (Flatmark, 1966). Margoliash & Schejter (1966) suggested that the inhomogeneities attributed to such methods were due to poor pH control during the step, or too drastic initial extraction procedures. The use of ammonium sulphate fractionation in this study gave a substantial purification

of cytochrome c as judged by the quantity of a white copious precipitate discarded during the method, and no evidence of large-scale deamidation or modification resulting from the method were observed in the later purification stages. All the invertebrate cytochromes c purified for sequence analysis have been reported to precipitate in 60-100% saturated ammonium sulphate solution (Tuppy, 1957; Chan & Margoliash, 1966; Chan, 1970; Brown et al., 1972; Fernandez-Sousa et al., 1974). Yamanaka (1966) reported the precipitation of prawn, housefly and Styela cytochromes c and the non-precipitation of Dendrostomum, oyster and squid, together with certain vertebrate cytochromes c reported to precipitate in saturated ammonium sulphate solution (Margoliash & Walasek, 1967). Of the cytochromes c purified during this study, all except Loligo showed no precipitation at saturations of ammonium sulphate up to 100%, so that fractionations were terminated at about 80% saturation when the precipitation of impurities ceased. Loligo cytochrome c was soluble in 60% saturated ammonium sulphate and showed partial precipitation at 100% saturation. The supernatant from the ammonium sulphate fractionation was exhaustively dialysed, and concentrated on a short column of CM-52 cellulose.

From this point in the purification method, concentration steps were performed using CM-52 cellulose or ultra-filtration, and dialysis was avoided because at high concentrations, cytochrome c has a tendency to bind to dialysis tubing in low

ionic strength buffers (Laycock, M. V., unpublished experiments, cited in Meatyard, 1974). Cytochrome c was routinely handled in 10 mM-phosphate buffer as Wasserman et al. (1963) reported that wheatgerm cytochrome c was unstable at low ionic strengths and only the purest fractions from a particular step were selected for further purification. Less pure fractions were pooled for recycling.

Molecular exclusion chromatography was shown by Flatmark (1964) to provide further purification of protein preparations and the technique was applied to this study. However, a consideration of the purity ratios in Table 41 show that the step rarely increased the purity ratio by very much, but that this increase was critical to the purification. The evidence suggested that this method was more applicable to purification yielding larger quantities of protein than were obtained in this study.

Final purification was achieved by adsorbing the cytochrome c to a column of CM-52 cellulose, oxidising the sample with potassium ferricyanide and eluting the protein by means of linear ionic or pH gradients. Some of the preparations of Asterias and Eisenia required both types of gradient. The pH gradient was applied with a constant cation concentration (Boardman, 1959; Margoliash, 1962; Margoliash & Lustgarten, 1962), and fractions were collected in tubes containing pH 7.2 buffer to minimise the time of exposure to conditions of high pH. The linear ionic gradient was applied at a

TABLE 41.

Estimates of the yield and purity of invertebrate cytochrome c at different stages of the purification

Source	$(NH_4)_2SO_4$ Fractionation Yield/Purity	Biogel P-30 Yield/Purity	CM-52 Ionic grad. Yield/Purity	CM-52 pH gradient Yield/Purity	Sephadex G-75 Yield/Purity	Final step Yield/Purity
Locusta	36/2.6	26/3.7	20/4.0	-	-	20/4.0
Macrobrachium I	1.8/2.6	1.6/2.6	1.1/4.1	-	-	1.0/4.1
Macrobrachium II	2.0/1.6	1.2/2.6	0.8/3.6	-	0.6/4.3	0.5/4.1
Eisenia I	<1.0/<0.2	0.6/<0.5	0.4/2.1	0.3/3.9	-	0.3/4.0
Eisenia II	<1.5/<1.0	1.0/1.4	0.6/4.0	0.8/2.6	-	0.6/4.0
Eisenia III	<1.0/<0.5	0.8/<1.0	0.4/3.6	-	-	0.3/3.5
Asterias I	1.8/0.7	1.1/0.9	0.6/3.8	0.8/2.1	-	0.5/3.8
Asterias II	<2.0/<0.4	1.0/0.6	0.8/4.1	-	-	0.6/3.8
Asterias III	<1.0/<0.2	0.5/0.4	0.3/3.8	-	-	0.2/4.0
Loligo	1.1/<0.8	-	0.4/3.9	-	-	0.3/3.8

The yields are expressed as mg/kg starting material.

The purity is expressed as the $E^{C3+}_{410}/E^{C3+}_{280}$ ratio.

constant pH (Margoliash, 1962; Margoliash & Lustgarten, 1962), and this technique was capable of resolving deamidated and modified forms from native cytochrome c. The quantities of observed deamidated and modified forms of the protein were small. Novotny (1971) concluded that fractionations were not always improved by the use of a shallower gradient, especially when the protein elutes at a low point in the gradient. This conclusion was confirmed by the results of varying the steepness of the gradient and the column length, the most ideal combination being a 1 cm x 30 cm column with a gradient of 0.5 mmol/ml.

The ampholine isoelectric focusing method used by Flatmark & Vesterberg (1966) for the purification of beef heart cytochrome c, was not investigated during this study. The recent availability of commercial ampholine carriers in the pH range 9-11 may well overcome difficulties experienced with this method as a result of the generally high IpH values for cytochrome c (see Richardson et al., 1970).

The crystallisation of cytochrome c was not attempted, although this has been reported for invertebrate, vertebrate and plant cytochromes (see Lemberg & Barrett, 1973). Crystallisation is not reliable evidence for purity, e.g. crystalline whale heart cytochrome c has been reported to contain myoglobin (Minikami et al., 1958) and crystalline beef heart cytochrome c has been reported to contain deamidated forms of the protein

(Flatmark, 1964).

The invertebrate cytochromes c used in this study are spectrally related to other mitochondrial cytochromes c and although the $E^{C2+}_{416}/E^{C2+}_{550}$ spectral ratio is reported to be higher for plants and algae than for animals (Meatyard, 1974), values obtained for invertebrate samples are comparable with the plant figures (see Table 42). Apparently equally pure samples of the same cytochrome c have been shown to vary in their spectral ratios possibly due to phenolic impurities, demonstrating that caution is required when applying this method to assess purity. Differences in the tryptophan, phenylalanine and tyrosine content can cause variations in the 280 nm absorbance and the soret band (410 nm) is particularly susceptible to modifications in tertiary structure, increasing as the protein unfolds (Stellwagen, 1968), so that no absolute purity criteria based on spectral ratios can be established for all species. The use of iron content as an estimate of purity has inconsistencies because the method cannot distinguish between most modified forms and the native protein. Analytical column chromatography on weak cation exchanges, in combination with the inability of cytochrome c fractions to bind carbon monoxide, is considered a good criteria for purity (Margoliash, 1962; Margoliash & Schejter, 1966) as is analytical and polyacrylamide gel electrophoresis, but these methods were not investigated.

TABLE 42.

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

Species	$E^{C2+}_{550}/E^{C3+}_{280}$	$E^{C2+}_{416}/E^{C3+}_{410}$	$E^{C3+}_{410}/E^{C3+}_{280}$	$E^{C2+}_{416}/E^{C2+}_{550}$	Yield mg cytochrome/ kg starting material
Maize (seeds)(1)	1.1	1.25	4.9	5.7	0.55
Cauliflower (inflorescences)(2)	1.2	1.22	4.9	5.0	0.20
Spinach (leaves)(3)	0.9	-	4.0	4.7	0.07
Enteromorpha (4)	1.0	1.2	4.1	5.2	0.13
Locusta	0.9	1.2	3.9	5.3	20.00
Macrobrachium	1.0	1.2	4.1	4.8	0.75
Eisenia	1.1	1.2	4.0	4.4	0.40
Asterias	1.1	1.2	3.9	4.6	0.40
Loligo	0.9	1.2	3.8	4.6	0.30
Horse heart (5)	1.2	1.2	4.6	4.4	250.00

(1) Stevens et al. (1967)

(2) Richardson et al. (1971a)

(3) Asada & Takahashi (1971).

(4) Meatyard (1974)

(5) Margoliash & Walasek (1967)

The increasing range of cytochromes from a variety of structures is making apparent the great range of structures and properties within the mitochondrial cytochrome c group, and this means that a single set of purity criteria cannot be devised to cover the entire group. Similarly, no single purification method can be applied to obtain cytochrome c from the variety of sources available. The best approach appears to be the establishment of a basic method to follow in principle and modify as necessary, together with an excess of starting material to be handled in 5-10 kg batches.

II. The Methods of Protein Sequence Determination

Previous sequence determinations of cytochrome c have tended to use large quantities of protein starting material, e.g. Samia cynthia, 1.30 g (Chan & Margoliash, 1966); rabbit, 1.70 g (Needleman & Margoliash, 1966); dog, 0.82 g (McDowall & Smith, 1965) and wheat, 0.26 g (Stevens et al., 1967). The quantities of pure cytochrome c available in this study, were generally less than 30 mg and required a micro-sequencing method different as opposed to the direct or subtractive Edman methods previously applied. The method chosen was based on that used by Boulter and his co-workers for sequence studies on higher plant cytochrome c where only limited amounts of protein were also available, e.g. Abutilon and cotton (12 mg each) (Thompson et al., 1971) and sesame and castor (18 mg) (Thompson et al., 1970). The method was devised by Gray &

Hartley (1963b) and involves the proteolytic enzyme digestion of the protein, the separation and purification of the resulting peptides by high-voltage paper electrophoresis, and their sequence analysis by the dansyl-Edman method.

The proteolytic enzymes used in this investigation were chymotrypsin and trypsin, being the most specific of those available and having been used successfully in the structural analysis of other cytochromes c (see Chan & Margoliash, 1966; Stevens et al., 1967; Boulter et al., 1970). The set of peptides obtained from the digestion of cytochrome c by each of these enzymes normally gives sufficient overlapping peptide sequences after analysis to logically establish the total protein sequence by inspection.

The invertebrate cytochromes c were readily denatured in 95% ethanol and were normally adequately digested with chymotrypsin and trypsin. However, similar digests were difficult to obtain, particularly those using chymotrypsin, despite the standardisation of conditions. The degree of digestion was estimated from the titration curve of added alkali required to maintain the pH at a value of 8.0 using an autotitrator fitted with a recorder. The specificities shown by the enzymes were those expected (see Smyth, 1967; Kasper, 1970). In addition to the major points of cleavage, α-chymotrypsin showed some activity at peptide bonds involving the carbonyl groups of methionine, histidine and asparagine, and in a

chymotryptic digestion of Asterias cytochrome c cleavage at lysine-39 was observed. The activity of trypsin was much more specific, but in all digests cleavage of the peptide bond involving the carboxyl group of tyrosine-97 was observed, and occasional cleavage was observed at the tyrosine residues in positions 38 and 67, and the phenylalanine at 82. The activity of trypsin at tyrosyl peptide bonds has been reported (Carpenter & Baum, 1962; Matsubara et al., 1967; Ramshaw, 1972) and was thought to be due to a chymotrypsin contamination, but Inagami & Sturtevant (1960) demonstrated that trypsin preparations could hydrolyse N-acetyl-L-tyrosine ethyl ester (ATEE) and that this activity was due to intrinsic tryptic activity. Kostka & Carpenter (1964) claimed that TPCK-treated trypsin showed a small degree of activity towards ATEE but not against tyrosyl peptide bonds, and Keil-Dlouha et al. (1971) showed that only TPCK-treatment and extensive SE-Sephadex chromatography could remove the 'chymotryptic-like' activity from trypsin preparations. There are no reports of TPCK-treated trypsin cleaving phenylalanyl peptide bonds (see Kasper, 1970; Inagami, 1972).

Cyanogen bromide cleavage, used in the determination of several protein sequences, has been applied previously to cytochrome c (Heller & Smith, 1966; Kasper, 1970; Brown et al., 1972). The technique owes its success to the specificity of cyanogen bromide for methionyl-peptide bonds and the rarity

of methionine in proteins, thereby increasing the probability of obtaining large peptide fragments. These may provide valuable information on the correct sequence of peptides from enzyme digests, and also confirm sequences already established. A 96% cleavage of susceptible bonds has been reported using a 30-100-fold molar excess of cyanogen bromide dissolved in 70% (v/v) formic acid (Steers et al., 1965). In this study, a cyanogen bromide cleavage was performed on a preparation of Asterias cytochrome c apparently contaminated with a "glycine rich" impurity. After the reaction, the products and impurities were separated by molecular exclusion chromatography and the two fragments of cytochrome c were digested normally with chymotrypsin.

Peptides derived from enzymic digestions were routinely separated using high-voltage electrophoresis at pH 6.5 and pH 1.9 and paper chromatography. Resolution of chromatograms was improved by running the paper in the same dimension twice with a drying step at room temperature in between, and this approach was particularly effective for purifying heme peptides. It was found that the elution of electrophoresis and chromatography papers was more efficient when carried out twice, with a drying step at room temperature in-between.

The inclusion of the phenanthraquinone reagent in the scheme of location agents for electrophoresis and chromatography guide strips provided valuable information for the sequencing

of arginine-containing peptides, and it was found that the arginine test could be performed before the ninhydrin and Ehrlich tests. The Pauly reagent was also useful but was best applied on its own, rather than in the combinations recommended by Easley (1956).

The dansyl-Edman method for sequencing peptides is an extremely sensitive technique and has been used in a micro-form to determine the sequence of as little as 10 pmol of peptide (Bruton & Hartley, 1970). When the technique was combined with the method of Woods & Wang (1967) to separate the dansyl-derivatives of amino acids using polyamide chromatography, it had the advantage of providing an excellent resolution of all normally occurring protein amino acid derivatives. Resolution of the dansyl-derivatives of more unusual amino acids such as hydroxy-proline and ε-N-trimethyl-lysine, the separation of dansyl-isoleucine and dansyl-leucine, and the resolution of the dansyl derivatives of the basic amino acids, can also be achieved. The main difficulty associated with this and other end group techniques was the identification of amino acids whose dansyl derivatives were labile during the acid hydrolysis of the dansylated peptide. Asparagine and glutamine were deamidated to the corresponding acids and bis-dansyl-histidine was degraded to α-N-dansyl-histidine. The internal lysine residues of a peptide, after a number of Edman degradation steps, react their ε-amino function with PITC to form ε-PTC-lysine, the dansyl derivative

of which was rather unstable to acid hydrolysis, although α-dansyl-ε-PTC-lysine could be resolved with care from dansyl-leucine and dansyl-phenylalanine. Dansyl-proline was degraded on prolonged acid hydrolysis so that when a proline residue was suspected at the N-terminus of an intact or partially degraded peptide, the dansylated peptide was hydrolysed for only 6 h. The use of the 5 cm x 5 cm polyamide sheets allowed the rapid analysis of samples and resolution of the dansyl derivatives of proline, isoleucine, leucine, ε-PTC-lysine, phenylalanine, and various dipeptides was improved greatly by repeating the second dimension step. The method allowed an accurate chromatographic analysis of a hydrolysate in less than 50 min. The dansyl method can only be qualitative when used at each step of the Edman degradation, but the dansylation of a peptide hydrolysate can provide a semi-quantitative estimate of the amino acid composition of the peptide by an assessment of the relative intensities of the spot fluorescence under u.v. light following polyamide sheet chromatography. During the degradation of a peptide, re-dansylation of the sample, following hydrolysis of the labelled peptide in the normal way, provided information on the composition of the degraded peptide and was useful for confirming histidine in a sequence by the presence or absence of bis-dansyl-histidine. Dansylation without hydrolysis was capable of identifying free amino acids as impurities, the free C-terminal amino acid of a fully degraded peptide and amino acids released from a peptide by

timed digestions with carboxypeptidases.

In the Edman method, it was found that the use of nitrogen at the PITC and TFA stages improved the degradation yields (Percy & Buchwald, 1972), and only one extraction with butyl acetate, instead of the three recommended by Thompson et al. (1970), served to cut down peptide losses. When sequencing small quantities of peptides, it was found that the silylation of Durham tubes allowed a minimum aliquot to be taken for dansyl-analysis after each degradative step as the sample was prevented from sticking around the glass surrounds of the tube.

The purification of a peptide from Macrobrachium cytochrome c with a cyclised glutamine residue as the N-terminus demonstrated the ease with which peptides may become modified whilst undergoing manipulation, and the usefulness of the pyrrolidonyl-peptidase preparation of Armentrout & Doolittle (1969) for removing such residues.

In the dansyl-Edman method, the identification of amide residues presented the major problem. The method of Offord (1966) using the pH 6.5 electrophoretic mobility could assign the majority of amide residues, but certain residues could not be placed unambiguously, e.g. in sequences of mixed acids and amides. Limited use was made of the pH 6.5 electrophoretic mobilities of partially degraded peptides, but care had to be exercised when analysing lysine-containing peptides because of the formation of ε-PTC derivatives on exposure to PITC

and the loss of positive charges at pH 6.5.

The problems encountered with the sequencing of the heme peptide were similar to those reported by Meatyard (1974). Both chymotryptic and tryptic heme peptides were frequently purified in low yield only, and it has been suggested that the heme moiety may interfere with nearby susceptible bonds, thereby decreasing the yields of this and adjacent peptides (Margoliash & Schejter, 1966). This was supported in part by the observation in this study of apparently undigested heme-coloured material, bound irreversibly to the origin during electrophoresis at pH 6.5. Neither the direct performic oxidation or mercuric chloride methods were found to be satisfactory for the removal of the heme moiety for, even following the mildest treatments, the subsequent sequence analysis was generally poor. The de-heming of the total protein prior to digestion would have surmounted this problem, but the small quantities of cytochrome c available for this study would not allow this approach. The ideal solution to this problem for cytochromes c without an N-terminal blocking group is to use an automatic sequencer which can degrade the protein with the heme moiety intact, as was demonstrated with the information obtained for Locusta cytochrome c. The automatic sequencer is an ideal tool for the sequence determination of unblocked cytochromes c when used in conjunction with manual methods, for it can analyse the troublesome heme region where manual

analysis frequently breaks down. The identification systems for the phenylthiohydantoin (PTH)-derivatives of amino acids obtained by the sequencer has been developed such that the application of gas chromatography, thin-layer chromatography and conversion to the parent amino acid and subsequent analysis allows at least two independent methods for identifying a particular residue (see Niall, 1973).

The technical problems associated with the sequence determination of proteins and the likely sources of inaccuracy and error when using limited quantities of protein, have been discussed by Ambler & Wynn (1973).

III. Comparisons of Cytochrome c Sequences

The amino acid sequences of the invertebrate cytochromes c show several characteristics which can be considered as being typical of the cytochrome c group. All consist of a single polypeptide chain of 103-108 residues with a covalently bound heme group. They are generally longer than vertebrate cytochromes c (103-104 residues) but are shorter than higher plant, fungal, algal and protozoan cytochromes c (111-113, 108-111, 112-114 and 103-110 residues respectively), whose amino acid sequences have been determined.

It is possible to align the cytochrome c sequences with reference to the two cysteine residues involved in the thioether linkages with the heme prosthetic group, except for

the Euglena and Crithidia sequences which have a common -Ala-Ala-Gln-Cys- sequence in this region. It is common practice to number the residues in the sequence from the invariant glycine which is the N-terminus of the vertebrate cytochromes c, so that the cysteine residues occupy positions 14 and 17. The alignment also demonstrates that the differences in length of various cytochromes c are due almost entirely to additional residues occurring at the N terminus prior to the invariant glycine designated residue-1 in the standard alignment. These N-terminal tails vary from 1-10 residues in length, whilst differences in length at the C-terminus are restricted to 1-3 residues. The vertebrate cytochrome c sequences commence with an acetylated glycine residue at the N-terminus, whilst higher plants and the green alga, Enteromorpha, have an 8-residue N-acetylated "tail". The fungal cytochromes c have an N-terminal "tail" of 1-8 residues, whilst, of the protozoan species, Euglena has no tail and glycine-1 is not acetylated and Crithidia carries an 8-residue "tail". All the invertebrate cytochromes c analysed, prior to and during this study, have a free α-amino group at the N-terminus except for Macrobrachium, where the nature of the blocking group was not determined. All the insects have a similar 4-residue "tail" differing only at the minus-1 position, and Eisenia has a 5-residue N-terminal "tail", which is clearly related to the N-terminal region of the insect cytochromes c. N-terminal determinations

on the cytochromes c of the annelids, Arenicola and Nereis, and the mollusc, Loligo, showed them all to be unblocked and all to possess a glycine N-terminus, but no determination was made of the presence or absence of an N-terminal "tail".

The cytochromes c of Asterias, Macrobrachium and Helix all have the invariant glycine-1 of the standard alignment as their N-terminus. All the higher plant cytochromes c analysed to date carry ε-N-trimethyllysine in positions 72 and 86, whilst this residue occurs in position 72 in Enteromorpha and is either absent or occurs at either position 72 or 86 in the fungi. Crystallographic studies with vertebrate cytochromes c have shown that residues occupying positions 72 and 86 occur at the surface of the molecule (Dickerson et al., 1971) and it is thought that ε-N-trimethyllysine residues result from the post-protein synthesis methylation of lysine residues (Scott & Mitchell, 1969). No vertebrate or invertebrate cytochrome c sequences have been found to contain ε-N-trimethyllysine, although it has been reported in the sequences proposed for Euglena and Crithidia cytochromes c (Pettigrew, 1972; 1973) and also in the amino acid compositions of cytochrome c from various protozoan trypanosomatids (Hill et al., 1971).

The absence of ε-N-trimethyllysine in vertebrate and invertebrate cytochromes c may be due either to the absence of the lysine-methylating enzymes, or to conformational changes resulting from amino acid substitutions affecting the specificity of these enzymes. However, the latter cause is

not indicated as Neurospora crassa cytochrome c contains ε-N-trimethyllysine at position 72, whilst the surrounding amino acids are identical to those found in this normally invariant region in animal cytochromes c (Delange et al., 1969). It would appear, therefore, that the methylating enzymes of plants and the fungi are not present in animals.

The determination of further invertebrate cytochrome c sequences has not reduced the number of invariant residues found in all the known cytochromes c. If no distinction is made between lysine and ε-N-trimethyllysine because of the post-synthesis origins of the latter, twenty eight residues are found to be invariant throughout seventy known cytochrome c sequences. Boulter & Ramshaw (1972) reported thirty three invariant residues, and suggested that because only three of these residues were not specifically correlated with a structural aspect of horse-heart cytochrome c (Dickerson et al., 1971), the statistical estimate of thirty two invariant residues (Fitch & Markowitz, 1970) was expected to be very close to the numbers expected from structural considerations. The publishing of further data has, however, reduced the number of observed invariant residues to twenty eight (Brown et al., 1972; Pettigrew, 1972; 1973) and the difference between this figure and statistical and structural estimates (Fitch & Markowitz, 1970) may be due to the occurrence of conservative amino acid substitutions. These may fulfil

structural constraints whilst reducing the level of residue invariance.

No positive evidence for heterogeneity in the invertebrate cytochromes c understudy was obtained, although this has not always been the case. For example, heterogeneity at a small number of residues has been observed in the cytochromes c from castor and sesame (Thompson et al., 1971), from rape (Richardson et al., 1971b) and from carp (Gutler & Horstmann, 1970), but in all cases predominant forms were observed for each residue. Observed heterogeneity in a protein sequence may be due to heterozygosity at one or more loci, polymorphism due to inter-cistronic differences of the gene specifying the protein either within or between populations, translational ambiguity at the messenger RNA template (Carbon et al., 1966; Rifkin et al., 1966) or merely as an artefact due to modification of the protein during extraction, purification or sequencing procedures. The heterogeneity reported for carp and plant cytochromes c was probably a result of polymorphism. During this study a number of peptides were isolated from a given source, and found to possess identical sequences except for acid/amide differences at particular residues. Such observations were attributed to deamidation during protein and peptide purifications, and where an amide-containing form of the peptide was isolated, it was considered to represent the in vivo situation. With the exception of the iso-

cytochromes c-1 and -2 isolated from Baker's yeast (Slonimski et al., 1965), little evidence has been found for different forms of cytochromes c within an individual, although mules and hinnies carry 50% horse and 50% donkey cytochrome c which differ by a single residue (Walasek & Margoliash, 1969). Matsubara & Smith (1962) reported that using a mixed population of seventy individuals, 10% were found to carry leucine instead of methionine at position 65 in human cytochrome c. However, attempts to demonstrate polymorphism in cytochrome c from populations of horses and humans proved unsuccessful (Margoliash, 1969a) and Stewart & Margoliash (1965) showed that cytochrome c preparations from different organs of the same species were identical. Margoliash (1969b) estimated that the sequence methods employed in his work could detect no less than 5% heterogeneity for a given residue. In view of the small quantities of starting material and the nature of the dansyl-Edman method, it is clear that the invertebrate cytochromes c investigated in this study would have to carry more than 25% heterogeneity for a given residue in order to be detected and recorded as significant.

The sequences of invertebrate cytochromes c show remarkable similarities among themselves and with other cytochromes c from eukaryotic sources. The mitochondrial cytochromes c are all chemically and physically similar, functioning as electron carriers in the mitochondrial respiratory electron

transport chain (see Keilin, 1966; Margoliash & Schejter, 1966; Lemberg & Barrett, 1973). These similarities are underlined by the reports that cytochromes c from other species can successfully replace the cytochrome c of a given species in the intra-mitochondrial reaction with the cytochrome c oxidase of that species (Jacobs & Sanadi, 1960; Margoliash et al., 1971). However, this interchangeability is not unlimited. Yamanaka and others (Yamanaka & Kamen, 1965; Yamanaka, 1966; 1967; Yoshida, 1966; Yamanaka et al., 1969) have attempted to make evolutionary deductions from the relative rates of interactions of a variety of cytochromes c with mammalian and Pseudomonas cytochrome oxidases. The results obtained for invertebrate cytochromes c as compared to those obtained for vertebrates, plants, fungi and microorganisms, were generally as expected from classical phylogenetic considerations. However, the degree of resolution of such a method is rather poor, and even the apparent resolution of the various cytochromes c reported (Yamanaka, 1966; 1969) is probably due to the use of photosynthetic c-type cytochromes for the algae, rather than the mitochondrial c-type cytochromes homologous with those applied for higher organisms.

Little is known of the relations of mitochondrial cytochrome c with the cytochromes which function in anaerobic energy-yielding reactions and which occur in chemosynthetic bacteria, not of those which function in the photoreduction processes of photosynthetic organisms (Kamen et al., 1971).

All these proteins contain heme-c as the prosthetic group, but sequence analysis has been largely confined to eukaryotic mitochondrial cytochrome c (see Lemberg & Barrett, 1973). The data of Dus et al. (1968) for the cytochromes cc' c_2 of Chromatium and Rhodospirillum respectively, show strong evidence for homology with eukaryotic mitochondrial cytochrome c. Also, Dickerson (1971) has found strong evidence for homology of Pseudomonas cytochrome c-551 with horse-heart cytochrome c, if a deletion of sixteen residues in a hairpin loop of the sequence is assumed. A photosynthetic cytochrome c from the alga, Monocystis lutheri, has also been shown to have some sequence homology with mitochondrial cytochrome c (Laycock, 1972). A review of the various c-type cytochromes has been given by Lemberg & Barrett (1973).

IV. Structure and Function in Invertebrate Cytochrome c

The structure of crystalline horse-heart and bonito cytochromes c has been determined by x-ray methods to a resolution of 2·8 Å (Dickerson et al., 1971). The tertiary structure of both molecules was identical, apart from minor changes associated with the side-chain substitutions shown in the differences of primary structure. This invited the conclusion that the essential structure of the protein had been conserved for at least four hundred million years (Dickerson et al., 1971).

Cytochrome c is a good example of the oil-drop model of a protein having buried hydrophobic side-chains, and polar side-chains at the surface of the molecule. The main features of the molecule, as described by Dickerson et al. (1971), are the binding of the heme together with its interactions with the polypeptide chain; the hydrophobic environment around the heme group; the two hydrophobic channels into the interior on either side of the heme group; the role of the aromatic side-chains and the basic surface patches; the involvement of residues in hydrogen bonding and the importance of the invariant glycine residues. Such features are seen to be broadly representative of all the eukaryotic mitochondrial cytochromes c, as judged by the amino acid sequences published to date, and are implicated in the functional interactions of cytochrome c with cytochrome oxidase, cytochrome reductase and the mitochondrial membranes.

Boulter & Ramshaw (1972) used the α-carbon diagrams of Dickerson et al. (1971) to construct a three-dimensional model of horse-heart cytochrome c and applying the amino acid sequences of fourteen plant cytochromes c together with that of horse-heart cytochrome c to the model, they were able to make predictions regarding the structure and function of higher plant cytochrome c. In the absence of x-ray crystallography data for plant cytochrome c, this is the only approach that can provide information on the protein three-dimensional

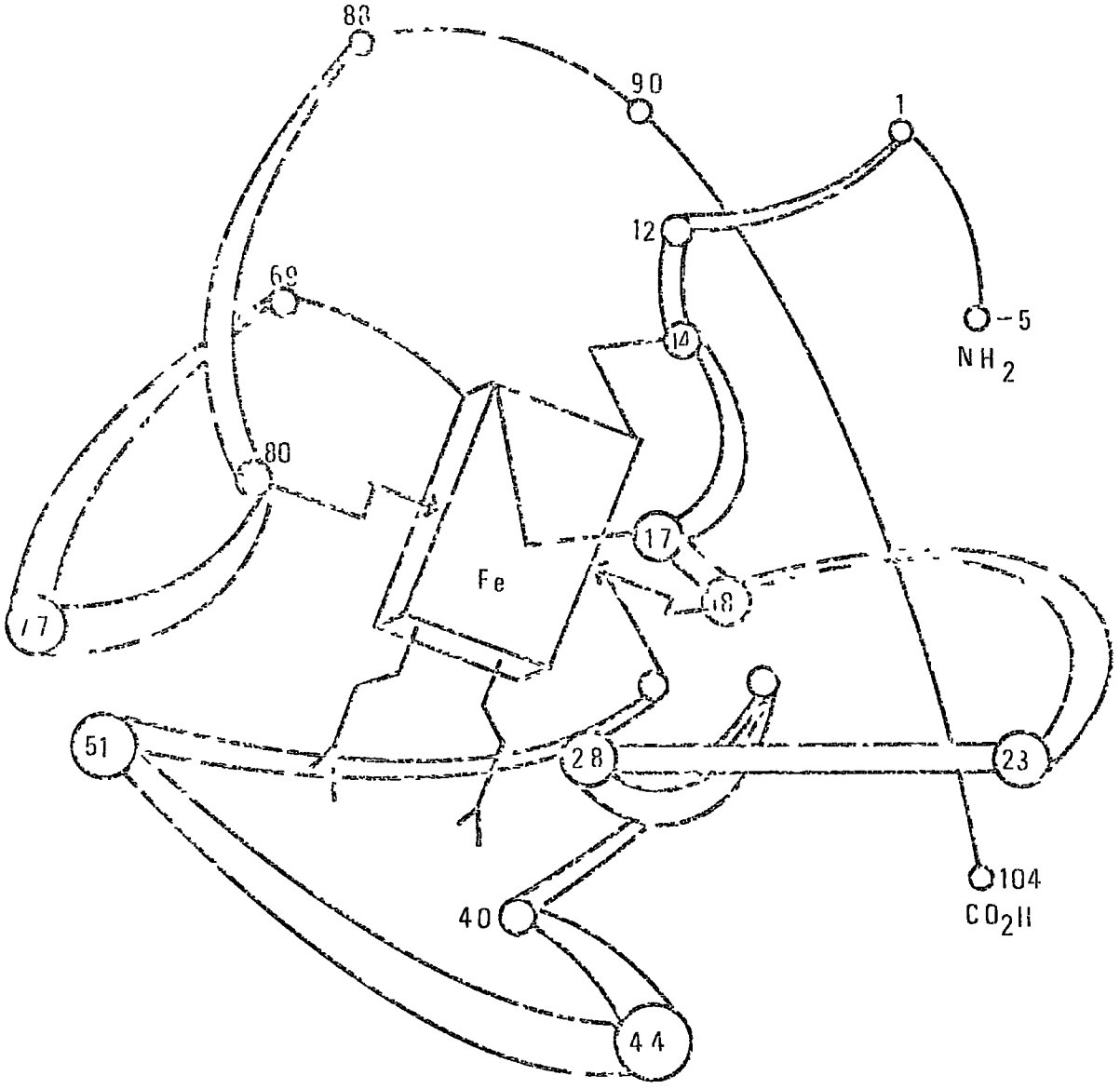
structure. This similarity of both the primary and tertiary structures suggests that these cytochromes c are homologous and provides a sound basis for the use of sequence data from higher plant cytochrome c in molecular phylogenies.

In a similar fashion, predictions may be made about the structure of invertebrate cytochrome c using the ten amino acid sequences now known from this multi-phyletic group. Invertebrate cytochromes c, in common with the majority of other cytochromes c, have the residues cysteine-14 and cysteine-17 covalently linked to the heme prosthetic group via the vinyl side-chains of its α-carbon atoms. Evidence that histidine-18 and methionine-80 are liganded to the heme, as shown for the horse-heart molecule (Dickerson et al., 1971), follows from the invariance of these residues in all invertebrate cytochromes c, and from the similarities of absorption spectra shared with those of horse-heart cytochrome c.

Horse-heart cytochrome c has sixteen hydrophobic residues which point in towards, and pack around, the heme group. Invertebrate cytochromes c carry identical residues in these positions, or replacements which are chemically and sterically very similar. Such replacements have been observed in vertebrate, plant and fungal cytochromes c. In horse-heart cytochrome c the hydrophobic channel to the right of the heme is bordered by residues 6-20 and the α-helix of residues 92-101 (see Figure 44) and in the invertebrates, a similar

TOP

BACK RIGHT



FRONT LEFT

BOTTOM

arrangement probably exists involving isoleucine or leucine in position 9, phenylalanine-10, leucine-94, isoleucine-95, tyrosine-97 and leucine-98, these last five residues being invariant in invertebrate and all other known cytochromes c, with the exception of certain fungi. The invertebrates, in common with the vertebrates and higher plants, have an invariant leucine at position 32, and at position 35 they may carry leucine, isoleucine or phenylalanine. The side-chains of the residues in these two positions point inwards towards the heme cavity. In horse-heart cytochrome c, the hydrophobic channel to the left of the heme is enclosed by residues 52-74 and the invertebrate sequences have invariant residues isoleucine-57, tryptophan-59, leucine-64, tyrosine-67 and leucine-68, in a similar fashion to the vertebrates. In particular, the region 63-68 appears to require a certain hydrophobic stability in invertebrates. For example, all the insects have a sequence -Thr-Leu-Phe-Glu-Tyr-Leu- except for Locusta which has -Thr-Leu-Phe-Ile-Tyr-Leu-; Macrobrachium has -Thr-Leu-Asp-Val-Tyr-Leu-, whilst Eisenia has -Thr-Leu-Tyr-Glu-Tyr-Leu-. Thus, there are always at least four hydrophobic residues in this short sequence.

In horse-heart cytochrome c there are nine aromatic residues and only those in positions 36 and 46 show any variability in the other vertebrate sequences (phenylalanine/isoleucine and phenylalanine/tyrosine, respectively). In

the invertebrates there are eight invariant aromatic residues in addition to which residues 35-36 are either Leu-Phe (Macrobrachium), Phe-Tyr (Samia), Phe-Phe (Manduca) and Ile-Leu (Asterias), whilst phenylalanine or tyrosine occupies position 65, except in Macrobrachium which has aspartic acid in this position. Thus, the invertebrate tyrosine/phenylalanine ratio varies from $^{6}/4$ (Manduca) to $^{5}/5$ (Samia), $^{4}/4$ (Macrobrachium and Helix), and $^{5}/4$ (Locusta).

In horse-heart cytochrome c, the basic residues are non-randomly distributed and their side-chains generally point outwards. On the left-hand side of the molecule, residues 86, 87 and 88, 72, 73 and 79, 53, 39 and 55 form three basic strips at the surface of the molecule and these groups are shown in part within the invertebrate sequences. Residue-39 is not always basic in invertebrate sequences (e.g. Eisenia and Macrobrachium) and no invertebrate sequence to date carries a basic residue-88, although the region 86-91, with the exception of Helix, always contains three basic residues. Residue-86 is an invariant-lysine in invertebrate cytochrome c. On the right-hand side of the horse-heart molecule, residues 5, 7, 8, 13, 22, 25, 27, 99 and 100 are all basic and in invertebrates residues 5, 7, 13 and 27 are invariantly basic, whilst residues 8, 22, 25, 99 and 100 are basic in certain cases. In general, the invertebrate sequences are more similar to the vertebrates, rather than the fungi and

higher plants, with regard to the basic amino acid content and the degree of base excess. Brown et al. (1972) reported that Helix cytochrome c had a greater electrophoretic mobility at pH 8.7 than horse-heart cytochrome c, and it has been generally observed, in this study and elsewhere, that animal cytochromes c require a stronger ionic concentration to effect their elution from cation exchangers than do plant cytochromes c. Thus, the pI of invertebrate cytochrome c may lie more closely to the value of pH 10.0 determined for mammalian cytochrome c (Barlow & Margoliash, 1966) than that of pH 9.5 determined for mungbean cytochrome c (Laycock, 1971). All cytochromes c sequenced to date have contained 2-3 histidine residues and in addition, vertebrate sequences contain sixteen invariant basic residues (arginine and lysine) and five invariant acidic residues, whilst invertebrate sequences contain ten invariant basic residues and two invariant acidic residues. In general, vertebrate cytochrome c contains an 8-10 basic residue excess, invertebrate cytochrome c a 7-9 basic residue excess, and plant cytochrome c a 4-6 basic residue excess.

The acidic residues in horse-heart cytochrome c are also not randomly distributed (Dickerson et al., 1971), and the acidic patch at residues 2-4 is duplicated in six of the ten invertebrate cytochromes c. Residue-21 is an invariant acidic amino acid in all the vertebrates and higher plants and in all the invertebrates, with the exception of Macrobrachium. Groups of acidic residues occur in vertebrate

cytochrome c in the regions 60-69 and 89-93, and these are also seen in invertebrates, plants and fungi. Acidic residues are not normally found elsewhere in the sequences, with the exception of aspartic acid-50 in six of the ten invertebrate sequences and the glutamic acid residues close to the C-terminus (see Figure 45).

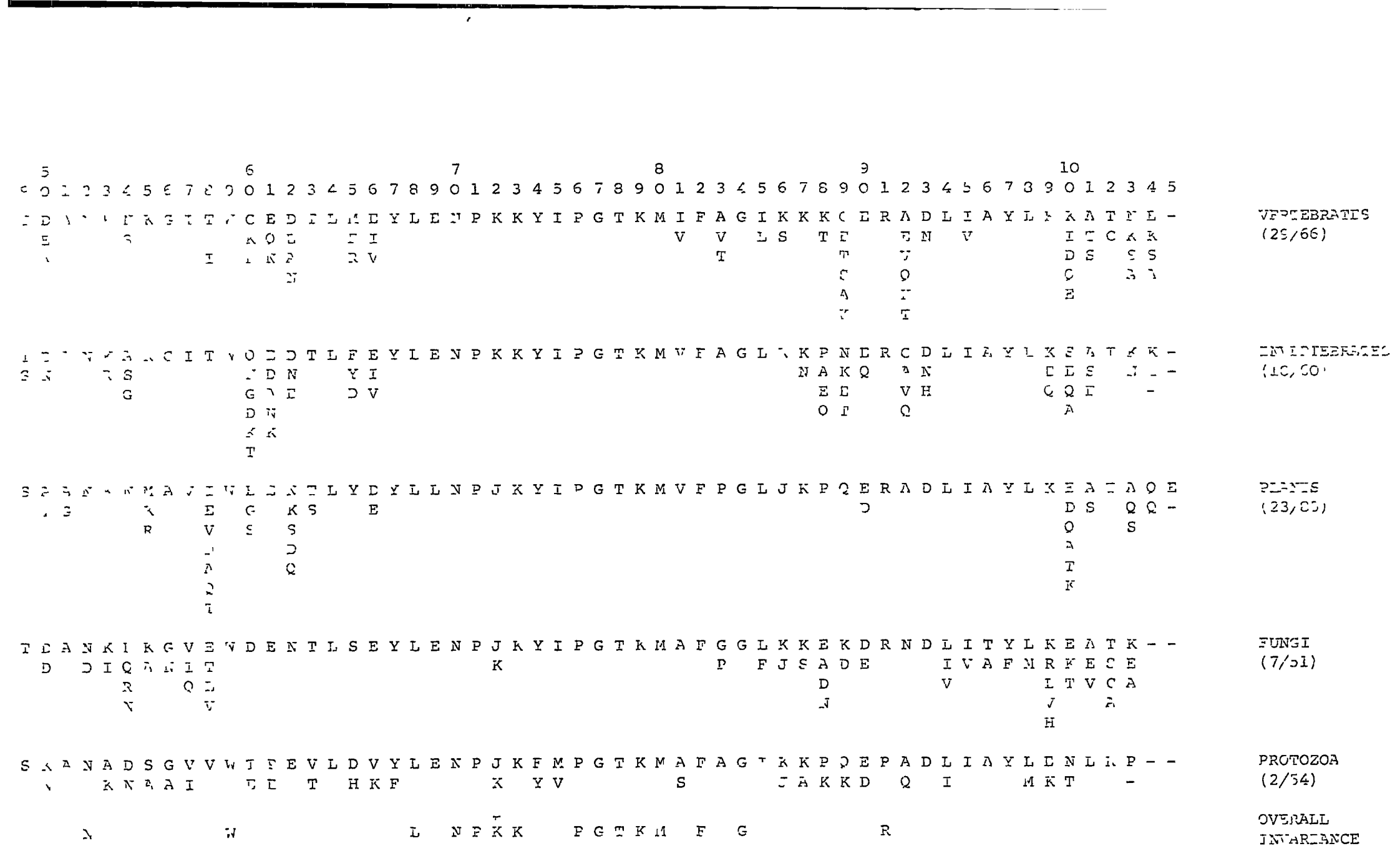
Invertebrate cytochrome c has fewer residues capable of forming hydrogen bonds than is the case for higher plant cytochrome c (Boulter & Ramshaw, 1972) and generally appears more close to the vertebrate condition. Many of the important hydrogen bonding residues tend to be invariant and a consideration of hydrogen-bonding potential shows fourteen such invariant residues in the higher plants, ten in the vertebrates and eleven in the invertebrates. In horse-heart cytochrome c, the δ nitrogen atom of histidine-18 is hydrogen-bonded to the carbonyl oxygen of proline-30; the hydroxyl of tyrosine-67 is hydrogen bonded to threonine-78; and hydrogen bonding occurs between tyrosine-48, tryptophan-59 and the carbonyl group of threonine-40 with the buried propionic group of the heme, whilst the other heme propionic group is probably hydrogen-bonded to threonine-49 (Dickerson et al., 1971). All these residues are invariant in invertebrate cytochrome c, except for residue-49 which can be either serine or threonine, and it would seem likely that the hydrogen-bonding is similar in both vertebrates and invertebrates. Threonine-40 in horse-

FIGURE 45.

A summary of the variability in the sequences of vertebrate, invertebrate, plant and fungal cytochromes c discussed in the text (see Discussion IV). The data is taken from the sequences listed in Appendices II and III.

The single-letter abbreviations for amino acids are given on page iv; dashes indicate true deletions and undetermined residues.

The figures in parentheses represent the number of sequences compared and the number of invariant residues in those sequences respectively. *In the overall invariance no difference is made between lysine and ε-N-trimethyllysine.


VERTEBRATES (29/66)
INVERTEBRATES (10/60)
PLANTS (23/85)
FUNGI (7/51)
PROTOZOA (2/54)
OVERALL INVARIANCE

heart cytochrome c connects two polypeptide chains by hydrogen bonding with the main chain carbonyl group of residue-55, a state likely to be duplicated in the invertebrates, where residue-40 can be threonine or glutamine and residue-55 is an invariant lysine. In horse-heart, only a single amide residue (asparagine-52) is not situated at the surface of the molecule, and this is thought to hydrogen bond with threonine-49 and the outer heme propionic group (Dickerson et al., 1971). Residue-52 is an invariant asparagine in the invertebrate sequences. Dickerson et al. (1971) reported that the glutamine-16 side-chain pointed outwards in horse-heart cytochrome c, and that the glutamine residues at positions 12 and 42 interact with arginine residues at positions 91 and 38 respectively, thus closing the top and bottom of the heme crevice (see Figure 44). These glutamine and arginine residues occur invariantly in the invertebrate sequences and are presumed to form identical interactions.

In the known vertebrate cytochromes c, there are eleven glycine residues, eight of which occur in the first fifty residues of the protein, and in most cases the steric reasons for their occurrence have been made clear (Dickerson et al., 1971). In the invertebrate sequences, ten glycines are invariant, all in identical positions to those in the vertebrates. The only exception is residue-24 which is glycine in all the invertebrates except for Macrobrachium,

where residue-24 is leucine.

Estimations of α-helical regions in cytochrome c have not been very satisfactory. In horse-heart, only residues 91-101 exist in an α-helical conformation, largely due to the distortion of the rest of the molecule by the dominance of the heme prosthetic group (Dickerson et al., 1971). It is likely that similar conditions would affect potential α-helical regions in invertebrate cytochrome c in the same manner.

A consideration of all the known sequences of cytochrome c, shows that twenty nine vertebrate sequences have sixty six invariant residues, ten invertebrate sequences have sixty invariant residues, twenty three higher plant sequences have eighty five invariant residues, seven fungal sequences have fifty one invariant residues and the two protozoan sequences have fifty five invariant residues. Overall, the cytochromes c published to date show common possession of twenty eight invariant residues (see Figure 45 and Appendices II and III).

Figure 45 indicates a number of differences in the construction of cytochrome c molecules from vertebrates, invertebrates, higher plants, fungi and protozoa. The vertebrates show much variability at residues 33, 44, 67, 89, 92, 100, 103 and 104, and this variability is not always represented by conservative amino acid substitutions. The invertebrates show a similar variability at residues 89, 92 and 100, but in addition show a characteristic extensive variance

at residues 36, 60, 61 and 68. Regions of variance at residues 60-66 and 100-104 are similar in both vertebrates and invertebrates, but the latter approach more the characteristics of the fungal sequences in the variability observed at residues 19-25 and 88-93 which are less variable in plant and vertebrate sequences. In general, however, the invertebrate structure is close to that of the vertebrates than to any other group.

At first sight, it may appear that there has been a progressive shortening of the cytochrome c molecule from the N-terminus of the standard alignment during evolution (Smith, 1968). However, there is strong evidence for the synthesis of mammalian and fungal cytochrome c on 80S cytoplasmic ribosomes (Boulter, 1970) and it has been demonstrated that the initiation mechanism for synthesis on these particles involves $tRNA_{F^*}^{MET}$ (Smith & Marcker, 1970). This implies that the N-terminus of all proteins so synthesised is N-formyl-methionine and that proteins are "trimmed" by a hydrolytic enzyme(s), to give the active forms normally observed. This postulated enzyme system has yet to be isolated, but the observed differences in length of known cytochromes c reflect the specificity of the postulated enzyme rather than the length of the gene responsible for coding for cytochrome c. The invariance of glycine at the N-terminus of vertebrate cytochromes c and the fact that these are the shortest cytochromes c found, may also reflect the specificities of

the postulated enzyme. The variable length at the C-terminus of the standard alignment, however, is almost certainly due to the occurrence of deletions and/or mutations to a chain terminating triplet. Thus, Ginkgo cytochrome c terminates with glutamic acid at position 105, whilst Crithidia terminates with a lysine at position 102 in the alignment.

The limited variation in the invertebrate sequences substantiates the view that cytochrome c has been a conservative molecule during evolution (Dayhoff, 1969; 1972) and an overall consideration of the amino acid sequences suggests that the essential three-dimensional structure of cytochrome c has been conserved since the divergence of the major invertebrate phyla from the main vertebrate line of descent.

V. Aspects of Molecular Evolution

Similarities between amino acid sequences could have arisen in a number of ways. They could have arisen by chance as the result of parallel or coincidental changes, by the convergence of proteins related by a common biological function (analogy), as a result of the energy constraints on the folding of polypeptides or by the descent from a common ancestor (homology). The sequence evidence available points to a common ancestor for all mitochondrial cytochromes c and thus a divergent, homologous evolution is inferred rather than a convergent, analogous process. Investigations into the molecular evolution

of animal cytochrome c have been carried out by Margoliash and coworkers (Margoliash, 1963; 1964; Margoliash & Smith, 1965; Margoliash & Fitch, 1968) and a similar approach to higher plant cytochrome c has been made by Boulter and his colleagues (Boulter, 1972; 1973; Ramshaw et al., 1972). The whole field of the evolutionary aspects of protein primary structure has been reviewed by Nolan & Margoliash (1968), and more recently by Lemberg & Barrett (1973).

Fitch (1970) and others (Fitch, 1966a,b; Cantor & Jukes, 1966) demonstrated a statistical approach that was capable of distinguishing between similarities resulting by chance, and those resulting from analogous or homologous processes. The method involved a comparison of peptide segments of a fixed length (usually thirty residues) throughout pairs of protein sequences, and the frequency of the minimum mutation distances relating these segments was compared to that expected from a purely random situation. Dickerson (1971) pointed out the inbuilt error in basing a method of evolutionary homology determination on minimum mutation distance data, resulting from the degeneracy of the genetic code. A change in the third nucleotide base of a particular codon changes the amino acid coded for in only about half the possible cases. Changes in the first and second nucleotide bases of a codon are most influential in changing the amino acid coded for, although those resulting from first-base changes are frequently of a

chemically and sterically conservative type (see Crick et al., 1961; Crick, 1966).

When a departure from the random situation for pairs of sequences has been shown (Margoliash et al., 1969), it still remains to distinguish between convergence and homology. Fitch & Margoliash (1967b) presented an approach for use with a set of proteins containing a number of invariant residues statistically determining the degree of invariance expected to remain, irrespective of the number of sequences determined. Two sequences were considered to be ancestrally related if the number of shared identical residues exceeded the statistical estimate of invariance, for the protein set to which the pair belonged. Fitch & Margoliash (1967b) calculated the expected invariance for eukaryotic cytochromes c to be 27-29 residues, although more recently, Fitch & Markowitz (1970) have calculated an expected invariance of 32 residues. The observed invariance in cytochrome c became twenty eight residues as a result of the publication of the sequences of Helix (Brown et al., 1972) and Crithidia and Euglena (Pettigrew, 1972; 1973). The sequences determined in this study have not changed this figure. The observed degree of similarity between any pair of the known cytochrome c sequences is always greater than the calculated and observed invariance (see Appendices I-III), and it is thus concluded that they are ancestrally related.

A further problem with sequence comparisons arises where additions and/or deletions of residues have occurred, but a

number of methods are capable of demonstrating a departure from randomness without the need to consider such events (Fitch, 1969; 1970; Gibbs & McIntyre, 1970). In the mitochondrial cytochromes c, the numbers and sizes of these are small and may be readily identified. Thus, the additional residues in the plant, fungal, protozoan and certain of the invertebrate sequences are clearly located at the N-terminus, when the two cysteine residues involved in the covalent heme linkage are aligned (see Appendices II and III; Dayhoff, 1972). Even the protozoan Crithidia and Euglena cytochromes c, which have only one cysteine involved in the binding of the heme (Pettigrew, 1972; 1973), align at the Gln-Cys sequence of the heme attachment region, and this places the extra residues at the N-terminus. In a comparison of more distantly related c-type cytochromes, however, the aligning of sequences and thus the identification of additions and deletions is more difficult. The sequence of Pseudomonas fluorescens c-551 (Ambler, 1963) and Rhodospirillum rubrum c-2 (Dus et al., 1968) have been aligned with mitochondrial cytochrome c sequences, but the proposed arrangements disagree (see Cantor & Jukes, 1966; Dus et al., 1968; Needleman & Blair, 1969; Dickerson, 1971). In addition, Ambler and co-workers have aligned various bacterial cytochrome sequences demonstrating some similarity in this group (see Ambler et al. in Dayhoff, 1972; Ambler, 1971; Ambler & Wynn, 1973).

The validity of such comparisons can only be rigorously

tested by x-ray crystallographic analysis of the three-dimensional structure of the proteins under study, and this approach offers solutions to the differentiation between convergence and homology. The demonstration of similarity of three-dimensional structure points to homologous origins, although it is clear that the minimal sequence constraints necessary for function may not require an identical tertiary structure throughout the molecules under study. It should be noted, however, that tertiary structural similarities may reflect only the energetic constraints associated with polypeptide chain folding rather than a true homology, although this seems unlikely for cytochrome c as judged by the degree of sequence and structural similarities.

Dickerson et al. (1971) have determined the three-dimensional structure of horse-heart and bonito cytochromes c and similar studies have been undertaken on the cytochromes c of spinach and rice (Morita & Ida, 1972; Morita et al., 1973). The data obtained indicates that these cytochromes c are very similar in terms of their tertiary structure. The usefulness of tertiary studies has been shown by Dickerson (1971) in justifying the postulated location of deletions in the sequence of Pseudomonas fluorescens c-551 necessary to align this protein with other cytochromes c. Despite only 25% similarity between the sequences of Pseudomonas c-551 and horse-heart cytochrome c, a comparative x-ray crystallographic study

clearly justifies the proposed areas of deletion.

The invertebrate cytochromes c, in company with other eukaryotic cytochromes c, show a clear resemblance to the structural constraints determined for the horse-heart and bonito proteins. Thus, many points of similarity existing between the sequences have been shown to be basic structural or functional requirements for horse-heart cytochrome c (Dickerson, 1971). These similarities manifest themselves as identical residues or residues linked by chemical and steric similarity in structurally important regions of the molecule. A semi-rigorous analysis of the structure of invertebrate cytochrome c has been carried out by comparing the sequences to that of horse-heart cytochrome c, noting the identical residues and investigating the chemical, physical and steric nature of the observed substitutions (see Discussion IV). From this, it is concluded that the tertiary structures of invertebrate and vertebrate cytochromes c are essentially identical and that this is due to a common ancestry for the two groups. A similar conclusion has been made for the higher plant cytochromes c (Boulter & Ramshaw, 1972), for algal cytochromes c (Meatyard, 1974), and could be made for cytochromes c from the fungi.

Despite the large quantities of evidence, it is still necessary to make an assumption of homology between invertebrate and other eukaryotic cytochromes c, since it is impossible to rigorously rule out the possibility of convergence. However, assuming that homology has been sufficiently demonstrated, it is then possible to examine the evolution of the

cytochrome c molecule and thus by inference, the evolution of the gene coding for cytochrome c. Such an analysis could relate directly to the relationships between the species from which the cytochromes c of known sequence have been purified, as well as the kingdoms to which these species belong.

The early sequence data for cytochrome c was mainly collected from vertebrate sources (see Dayhoff, 1972) and showed that the number of amino acid differences between sequences from members of a taxonomic class was less than those between sequences taken from members of different taxonomic classes, the number of differences increasing with the remoteness of the class relationships. Within a given class, the number of amino acid differences between the sequences of members lies within a range, contrasting with the more or less constant differences found in interclass comparisons, irrespective of which members of the classes are compared (see Appendix I). The constancy of difference in interclass comparisons is interpreted to imply that elapsed time is the important parameter in determining the number of mutations which accumulate in the cytochrome c structural gene along a line of evolutionary descent (Zuckerkandl & Pauling, 1965; Nolan & Margoliash, 1968). Thus, over a long period of evolutionary history, such as the minimum of 1500 million years suggested for cytochrome c, other factors relating to the rate of fixation of amino acid substitutions in a population have either cancelled or averaged themselves out. No account

need to be made of the mechanisms through which the differences may have arisen (Margoliash & Schejter, 1966) and on acceptance of the relationship enables a direct correlation to be made between elapsed time and the rate of change within cytochrome c (Margoliash, 1963).

Times of divergence of the major classes of vertebrates have been established from the fossil record (see Romer, 1966; Colbert, 1969) and may be used to estimate the average time for a single difference to become fixed between the sequences of different lines of descent. This time is called the "unit evolutionary period" and refers to a given protein set for which homology has been demonstrated (Nolan & Margoliash, 1968). Fitch & Margoliash (1970) distinguished between "parologous" and "orthologous" genes when establishing gene or species phylogenies. Both classes of genes are homologous, but parologous genes, although originally identical, may have duplicated and subsequently diverged to the extent that they code for proteins with different functions. The parologous condition may relate the mitochondrial cytochrome c group and some cytochromes f (c_6) in examples where multiples of molecular weight and the numbers of heme groups clearly relate to those of cytochrome c. Orthologous genes, however, are those which have remained identical in function during evolution. All the cytochrome c sequences constitute an homologous set with regard to the calculations of Fitch & Margoliash (1967b) and, apart from isocytochromes c_1 and c_2

of Saccharomyces, there is no evidence for gene duplication in the cytochrome c gene. The cytochromes c are thus both homologous and orthologous in the sense discussed by Fitch & Margoliash (1970) and may be used legitimately to calculate "unit evolutionary periods".

The documentation of the emergence of the invertebrate phyla in the fossil record is inferior to that for the various classes of vertebrates, largely as a result of the age differences of the deposits involved (see Shrock & Twenhofel, 1953; Cowie, 1967). Estimates of the times of divergence of the major invertebrate phyla from the main vertebrate line of descent have been made using a "unit evolutionary period" calculated from the fossil-based dating of the Mammalia-Aves divergence and the matrix of amino acid differences (see Figure 35, Table 38 and Appendix I). Linear extrapolation of such calculations will always give an underestimate of time values because of the occurrence of multiple mutations at a single codon and parallel and back mutations, but these may be partially corrected for using the statistical method of Feller (1950) (see Margoliash & Fitch, 1968; Figure 35). A further correction is possible by using a matrix of minimum mutation distances rather than amino acid differences, but was not applied to this study. Using this correction, an account may be taken of multiple mutations at a single codon (Margoliash et al., 1963; Fitch & Margoliash, 1967a), but not of parallel and back mutations which have been estimated at

20-30% and 1-6% respectively in cytochrome c (Fitch & Margoliash, 1969; Boulter et al., 1972). Statistical corrections assuming the acceptance of amino acid substitutions at all variable sites in a protein molecule (Margoliash & Smith, 1965) are not strictly applicable to cytochrome c where many positions in the sequence are clearly not randomly variable (see Fitch & Margoliash, 1967b; Discussion IV).

The justification of using a "unit evolutionary period" derived from a specific, recent animal example with more remote problems of divergence is largely theoretical, but the method has proved itself when applied to the vertebrate divergence as judged by the evidence of the fossil record (Young, 1962; Romer, 1966; Colbert, 1969). Estimates of the times of divergence obtained for the invertebrate phyla from the main vertebrate line of descent, largely agree with what is known from Pre-Cambrian and Cambrian fossils (Shrock & Twenhofel, 1953; Harland et al., 1967). All the groups in this study were shown to diverge in Pre-Cambrian times, the Annelida having the most primitive and the Echinodermata the most recent origins, assuming that a single sequence is sufficiently representative of its group. Within the errors of the estimated mean number of variant residues within the groups, both the Mollusca and Arthropoda diverged from the main vertebrate line about 710 million years ago. Similarly, the estimates obtained for the Insecta and Crustacea neither directly supports nor refutes a multi-phyletic origin for the

Arthropoda.

The correlation of estimated times of divergence with the evidence of 700 million years of the fossil record provides support for an approximately constant rate of acceptance and fixation of amino acid substitutions in cytochrome c. The estimated "unit evolutionary period" indicates that cytochrome c has accepted changes more slowly than most other known proteins possibly as a result of the severe structural constraints evident from tertiary studies and the minimum of three interacting functions with oxidases, reductases, and the mitochondrial membranes (see Dickerson et al., 1971; Dayhoff, 1972; McLaughlin & Dayhoff, 1973).

From the evolutionary point of view, the potentially most useful application of homologous protein sequence data is in a computer analysis leading to the construction of molecular phylogenies. The major problem in this approach is the evaluation of all possible phylogenies relating a given group of sequences and methods must be found to reduce the number of evaluations whilst obtaining an optimum solution. An example of this is the intra-kingdom, inter-kingdom phylogeny method used by Dayhoff and others (see Dayhoff, 1972; McLaughlin & Dayhoff, 1973; Meatyard, 1974), and applied to this study. Two types of approach to molecular phylogeny construction were used in this study; the "ancestral sequence" method (Dayhoff & Eck, 1966; Dayhoff, 1972) and a numerical matrix method (see Lance & Williams, 1966;1967; Fitch &

Margoliash, 1967a; Gibbs & McIntyre, 1970; Moore et al., 1973).

Molecular phylogenies constructed by the "ancestral sequence" method consists of nodes, each of which has three branches which lead to determined amino acid sequences or to adjacent nodes for which the method constructs the most probable sequences. The phylogeny, therefore, contains the determined sequences of extant species together with computed ancestral sequences. However, the point of earliest time for the phylogeny may not be established from sequence and must be fixed from biological considerations. Sequence data was taken from the sources described in Appendices II and III and the computer strategy has been described (see Materials and Methods IV,2.).

The "ancestral sequence" method involves two assumptions. Firstly, evolution is considered to have taken place by a minimum number of amino acid substitutions, and secondly, the final phylogeny accepted is that for which the number of amino acid substitutions relating the phylogeny is a minimum for all possible phylogenies.

With regard to the first assumption, it is clear that convergent changes (parallel and back mutations) have occurred during the evolution of cytochrome c. Boulter et al. (1972) detected 27% parallel and 6% back mutations during the evolution of higher plant cytochrome c, and Fitch & Margoliash (1969) reported 20% and 1% respectively, for the same processes

during the evolution of animal (largely vertebrate) cytochrome c. The "ancestral sequence" method has provided estimates of 28% parallel mutation and less than 5% back mutation in the known invertebrate cytochromes c, but there is no certain way of estimating the extent of undetected convergent changes, particularly when the number of sequences in a given taxonomic unit is small. The characteristically slow acceptance of amino acid substitutions in the cytochrome c molecule (Dayhoff, 1972) reduces the number of convergent changes, and the likelihood of relating similar sequences as a consequence of these rather than common ancestry is small, since all positions in the sequences are considered at each comparison.

With regard to the second assumption outlined above, it is clear that there is no certainty that the selected phylogeny is related by the absolute minimum number of amino acid substitutions assumed to relate the true phylogeny. Computer procedure 3 (see Materials and Methods IV,2.) was designed specifically to reduce the number of comparisons necessary to obtain, with reasonable certainty, the absolute minimum tree. An estimation of the efficiency of this procedure was made using a phylogeny relating seventeen sequences for which a minimum number of substitutions had been determined with reasonable certainty. Using different orders of the sequences, a number of phylogenies were built in the normal way and assessed at the end of the building and shuffling procedures.

It was clear that the building order was crucial to the phylogeny produced by the building strategy. Thus, a building order based on biological grounds (e.g. Protista-Fungi-Plants-Animals) gave a phylogeny close to the final minimum, whilst more random orders produced phylogenies related by a greater number of amino acid substitutions (see Table 40). However, procedure 3 was shown to be 60% efficient in obtaining an absolute minimum phylogeny regardless of the "original build", when applied to cytochrome c data. Similar analysis of the amino acid sequences of plastocyanin which, overall, are known to carry 50% convergent changes, has been shown to be less than 10% efficient (Peacock, D., unpublished experiments).

Phylogenies constructed as a result of the application of the "ancestral sequence" method to taxonomically representative samples of known cytochromes c, have all shown a grouping of the sequences that parallels the kingdoms of organisms proposed by Whittaker (1969) (see Dayhoff, 1972; McLaughlin & Dayhoff, 1973; Meatyard, 1974). Differences between phylogenies have centred on the degrees of remoteness of these kingdoms, demonstrated by the application of the "unit evolutionary period" to the proposed branch-lengths. Discrepancies are attributed to differences in the assignment of values to blank residues in the ancestral node sequences, the numbers of which increase when few sequences are available to establish kingdom ancestral nodes. Where a blank exists,

the minimum number of changes that must have occurred is averaged out along the branches concerned, so that the existence of several blanks in one area of the tree may favour the over-estimation of inter-node distances. The blanks assigned to ancestral node sequences are frequently due to deletions or additions of more than one amino acid in the sequences used to compute the nodes, and these are considered as multiple mutational events by the ancestral sequence method. No method can determine whether these are multiple events, single events affecting a larger region of the cytochrome c gene, or, with regard to the N-terminal "tail" region, deletions of residues resulting from the differing specificities of the enzyme responsible for removing the N-formyl-methionine region following protein synthesis (see Discussion IV). Thus, the values assigned to branch-lengths involving the evaluation of blanks at the N-terminus of the ancestral sequences may be responsible for the large inter-node distances between the major kingdoms (see McLaughlin & Dayhoff, 1973). During this study, a number of strategies were adopted to attempt to reduce branch-length distortion resulting from particular treatments of the N-terminal "tails" (see Results 17 and 18; Figures 36-40).

The occurrence of blanks within ancestral node sequences is almost entirely due to the lack of sequence data from the more remote areas of the phylogeny.

The branch-lengths in Figures 36-39 are expressed in "accepted point mutations", or PAMs, calculated by the method

of Dayhoff (1972), and elapsed time may be calculated on the basis of 3 PAMs:100 million years (Dayhoff, 1972).

Molecular phylogenies were also constructed using a numerical matrix method. The simplest measure relating two sequences is the number of amino acid differences between them, and the branch-lengths of resulting phylogenies have a numerical meaning, clearly related to the sequences. A similar measure, using minimum mutation distances has been applied (Fitch, 1966a; Fitch & Margoliash, 1967a) but other similarity methods lack numerical links with the sequence data (Gibbs & McIntyre, 1970; Sackin, 1971). Despite hidden assumptions, amino acid difference or minimum mutation distance measures provide molecular phylogenies for vertebrate proteins in close agreement with trees based on biological and fossil data (Fitch & Margoliash, 1967a; Dayhoff, 1972).

Phylogenies are constructed using a species dissimilarity matrix and Ramshaw (1972) and Boulter *et al.* (1972) applied the fusion strategies of Fitch & Margoliash (1967a) and the "flexible" strategy of Lance & Williams (1967) in the construction of molecular phylogenies relating fifteen higher plant cytochromes *c*. Using both amino acid difference and minimum mutation distance they obtained similar trees to the absolute minimum obtained from ancestral sequence methods.

The iterative approach of Moore *et al.* (1973) was applied in this study, using a dissimilarity matrix of amino acid differences (see Results 17.). The method assumes the

ancestral relationships between amino acid sequences are proportional to their present-day dissimilarity (see Fitch, 1967) and has been tested with simulated data (Moore et al., 1973; Peacock & Boulter, 1975). Numerical methods assume evolution to have occurred as a minimum number of changes and take no account of parallel and back mutations (see Fitch & Margoliash, 1969). In addition, a phylogeny based on present sequences can only approach true evolutionary relationships, if the rate of change has been relatively constant along all lines of descent (Jardine et al., 1969). Boulter et al. (1972) concluded that for periods in excess of 200 million years, this was correct for mitochondrial cytochrome c so that numerical methods are of value for relating distant species, but less so for more recently diverged groups such as the higher plants. Peacock & Boulter (1975) used model data to demonstrate a decrease in accuracy of the numerical method of Moore et al. (1972) with an increase of present-day sequence - nearest ancestor sequence distance. They concluded that the ancestral sequence method was 3-4 times better than the numerical method over small distances (less than 7 amino acid substitutions/100 residues), but over larger distances the numerical method was slightly better. On the basis of this, it was felt that the numerical method would deal more efficiently with the large branch-lengths shown for certain of the more remote invertebrate sequences by the ancestral sequence method. The results were rather unsatisfactory,

possibly because the rates of change observed in invertebrate cytochrome c were clearly unequal as shown by the branch-lengths obtained in the ancestral sequence method (see Figures 36-39). Given an estimate of the earliest point of time on the phylogenies occurring on the branch linking the Eisenia and Enteromorpha nearest ancestral nodes, it appears that the Annelida have accepted changes at a much lower rate than the Insecta, which in turn have shown a lower rate of change than the other invertebrate cytochromes c (see Discussion VII).

VI. Assumptions and Errors in Computed Phylogenies

Dayhoff (1972) and McLachlan (1971) have estimated the relative frequencies by which one amino acid is substituted for any other, and Grantham (1974) applied the data to show a strong correlation of the substitution frequencies with chemical similarities. No similar correlation could be made with the minimum number of base substitutions required to change the amino acid, and it was inferred that amino acid substitution is largely controlled by the rate of selection and fixation of mutational events into a population, rather than by the rate of nucleotide changes. Peacock & Boulter (1975) used the substitution frequencies of Dayhoff (1972) in a computer model to control the rate of randomly generated mutations. The data for the model contained high numbers of convergent (parallel and back) substitutions, and was thus

an accurate reflection of real data. The degree of difference between calculated and true topologies was estimated as the total number of one step, nearest neighbour changes (ISNNC) needed for a conversion from one to another. Errors introduced into the methods as a result of increasing ancestral distance, the assumption of parsimony and the underestimation of convergent changes were all estimated from the true figures recorded in the computer data. Peacock & Boulter (1975) found that for a mean distance over four generations of 10 amino acid substitutions per 100 residues, the error in the calculated topology was less than 7%. In comparison with the numerical matrix methods, the ancestral sequence methods was 3-4 times better over small distances (less than 7 amino acid substitutions per 100 residues), but rather worse over larger distances. Boulter (1974) estimated 33% convergent changes in a data set of twenty four sequences of higher plant cytochrome c, but Peacock & Boulter (1975) showed that an accurate estimate of phylogeny may still be made, despite this high level of convergence. This is because the computer will choose a false topology only when the number of convergent residues exceeds that of linking residues in the ancestral sequence method. Peacock & Boulter (1975) concluded that the ancestral sequence method is capable of great accuracy over distances of 1-3 amino acid substitutions per 100 residues, whilst the matrix method achieves its greatest accuracy over larger distances (greater than 30 amino acid

substitutions per 100 residues).

Thus, care must be taken during the interpretation of the phylogenies produced during this study. An estimate of less than 30% convergent changes was made for the invertebrate data set so that it is unlikely that false phylogenies have been accepted because this only becomes likely at greater than 37% convergence. Using the error estimates of Peacock & Boulter (1975) and the branch-lengths observed in the phylogenies obtained (Figures 36-42), the expected error in these phylogenies was in the range 3-6 ISNNC. Generally speaking, the phylogenies represent a reasonable reflection of current views on invertebrate phylogeny but while the discrepancies fall within the estimated errors, no phylogenetic significance can be attached to them.

VII. Classical and Molecular Phylogenies

There have been numerous reviews of proposed phylogenies relating the invertebrates (see Hyman, 1940-1959; Kerkut, 1960; Dougherty et al., 1963; Hadzi, 1963; Clark, 1964; Barrington, 1965; Jagersten, 1972). The following is an account of the common ground between these and the molecular phylogenies constructed during this study, and as such ignores any consideration of the lower Metazoa (see Hadzi, 1953; Hardy, 1953; Jagersten, 1955; 1959; Marcus, 1958; Dougherty et al., 1963). The main points of discussion are the relationships among the Arthropoda, Annelida and Mollusca,

the proposed polyphyletic origins of the Arthropoda and the relationships between the Echinoderm and Chordate lines of descent.

Grobben (1908) proposed a scheme to relate the invertebrates which divided the major phyla into two groups, the Protostomia and the Deuterostomia, based on the fate of the blastophore in the developing embryo and subsequent embryological and larval characteristics (see Appendix IV). In the Protostomian embryo, the blastophore forms the mouth and anus, whilst in the Deuterostomia it forms the anus and the mouth develops in another position. Such a dichotomous phylogeny is not at all clear-cut (Kerkut, 1960), but the approach is still acceptable. Three of the principal phyla placed in the Protostomia are the Mollusca, Annelida and Arthropoda and the relationships between these have been discussed (Marcus, 1958; Lemeche, 1959; Vagvolgyi, 1967). Vagvolgyi (1967) suggested that the Annelida and Mollusca shared a common ancestor as judged by ontogeny, and that this ancestor was a non-segmented, acoelomate organism similar to a primitive flatworm. He concluded that the molluscs developed their characteristic shell, mantle, radula and ctenidium, and the annelids their characteristic coelomic segmentation following the divergence of the two phyla (Yonge, 1960; Morton & Yonge, 1964). Lemeche (1959), commenting on the true segmentation observed in the mollusc Neopilina galatheae, concluded that the Mollusca and Annelida were closely related and, together with

the Arthropoda, could be derived from a generalised turberllarian ancestor. Spiral cleavage occurs widely in the embryos of molluscs and annelids and to a lesser extent in arthropods, and Yonge (1957) has suggested a common ancestry for these three phyla based upon the proposed homology of the gills of Neopolina, polycheate parapodia and the generalised arthropod limb. Marcus (1958) suggested a common ancestry of the Annelida and Mollusca, but derived the Arthropoda from the Annelida, placing these two in the super-phylum Articulata (see Appendix V). Hadzi (1953; 1963) disregarded a Protostomia-Deuterostomia bifurcation and divided the metazoa into four phyla arranged in a straight line of descent. The most primitive of these was the Armeria (including the molluscs and turberllaria) followed by the Polymeria (e.g. arthropods and annelids), the Oligomeria (e.g. echinoderms, hermichordates, pogonophora) and the Chordonia (e.g. vertebrates and urochordates), and thus the molluscs were regarded as being more primitive than the annelids (see Appendix VI). Carter (1954) took issue with the straight line of descent proposed by Hadzi (1953), noting that evolution has normally been observed to follow a pattern of radiation. He disagreed with the view of Hadzi (1953) regarding the Protostomia-Deutoerostomia dichotomy and attacked his use of metamerism and the necessity of proposing a reduction of segmentation in the Oligomeria. Carter (1954) concluded

that the major invertebrate phyla arose more or less simultaneously during the course of radiation from primitive metazoan stocks, and that the characteristics of the Protostomia and Deuterostomia were established prior to this radiation. Two other phylogenies, proposed by Hanson (1961) and Hyman (1940), are shown in Appendices VII and VIII.

The estimated times of divergence of the relevant phyla from the vertebrate line of descent obtained during this study were in accord with the views expressed above; with the exception of those of Hadzi (1963). The divergence of the arthropods and molluscs from the vertebrate line were similarly placed in the Pre-Cambrian more than 700 million years ago, although the divergence of the annelid line was placed rather earlier than this (see Figure 35, Table 38). All the ancestral sequence methods suggested a shared common ancestry for the Annelida and Arthropoda, but the numerical methods fixed the annelid-vertebrate common ancestor more recently than the insect-vertebrate ancestor (see Figures 36-42). All of the methods showed a grouping of the cytochromes c of Helix and Asterias, despite the addition to the input data set of a partial sequence of the mollusc, Loligo opalescens, for certain of the computer strategies (Figures 39, 41 and 42). None of the classical phylogenies entertain such a mollusc-echinoderm relationship, and although the point of divergence of Asterias from the vertebrate line was correctly placed from classical considerations, the addition of the

Helix sequence proposed a far more recent origin for the molluscs than has been suggested previously and from the estimates of times of divergence in this study (see Table 38). However, the branch-lengths in this area of the molecular phylogenies were generally large, and Peacock & Boulter (1975) have pointed out certain errors associated with the ancestral sequence method when comparing sequences showing variation greater than 7 amino acid differences/100 residues. Both the ancestral sequence method, which treated undetermined residues as differences, and the numerical method, which treated undetermined residues as unknowns (i.e. no difference), maintained the Asterias-Helix grouping because the number of common residues between the sequences, although relatively small, was greater than that between Helix and the partial sequence of Loligo. The undetermined C-terminal residues of Loligo cytochrome c may well have been sufficiently similar to the equivalent region in Helix to provide strong enough positive evidence to alter the final phylogeny. Thus, because incomplete sequences can only yield negative or neutral evidence where positive evidence is required, the Loligo data was omitted from the building and assessment of other phylogenies (see Figures 36, 37 and 38). Only the determination of further mollusc and echinoderm sequences resulting in the reduction of branch-lengths between the sequences compared will fully resolve this area of the phylogeny.

Manton (1973) has provided evidence from comparative

functional morphology and functional embryology, that the Arthropoda have polyphyletic origins (see Appendix IX.) Arthropodisation is supposed to have occurred at least three times giving rise to the Crustacea, Chelicerata and Uniramia, the latter comprising the subphyla Onychophora, Myriapoda and Hexapoda differentiated by their jaw mechanisms, head structure and limb bases. Manton (1973) pointed out that comparative anatomical studies were frequently poorly conducted and cited examples of attempts to relate insects and crabs through the common possession of compound eyes, the proper investigation of which shows only a rough similarity. Manton (1973) stressed the importance of the studies of functional morphology and embryology in the construction of an arthropod phylogeny, and criticized the lack of functional considerations in the speculations on arthropod relationships existing in the literature. Dahl (1969), Tiegs & Manton (1958) and Cisne (1974) have all proposed a polyphyletic origin for the Arthropods, whilst the main opponents have been Siewing (1960) and Sharov (1966), who attempted to derive the Arthropoda monophyletically from the Annelida in the manner of Marcus (1958) (see Appendix V). Manton (1967) demonstrated the fundamental differences between the annelid parapodium, with its endoskeleton of setae to which the musculature is attached, and the arthropod limb with its exoskeleton and internal muscle attachment, and thus rejected the suggestion of Sharov (1966) that the polychaete Spinther, or a similar form, was the forerunner

of the Arthropoda. Sharov (1966) derived the Insecta from the Crustacea, but this has been rejected by Dahl (1969) and Manton (1973). Cisne (1974) performed x-ray studies on Trilobite fossils and concluded that these extinct arthropods had much in common with the Crustacea and Chelicerata. He proposed that the arthropod exoskeleton had evolved twice leading to a Trilobita, Crustacea, Chelicerata(TCC)-Uniramia dichotomy within the Arthropoda. Embryological evidence suggests that both these groups originated from the same soft-bodied, segmented, worm-like creatures in Pre-Cambrian times (Anderson, 1973). Further circumstantial evidence for a common origin is that the TCC-Uniramia dichotomy falls (with one exception) between the primitively marine and primitively terrestrial arthropods, suggesting that the differences arose during their divergent adaption to different environments. Arthropodisation was proposed to occur during this adaption and thus subsequent to the divergence of the two groups (Cisne, 1974).

The molecular phylogenies broadly suggest at least a diphyletic origin for the arthropods. The degrees of error outlined previously (see Discussion VI) were probably less in the arthropod segment of the phylogeny than suggested for elsewhere (see above), because of the shorter branch-lengths involved. However, only the determination of further cytochrome c sequences from the Chelicerata and Crustacea can provide a true indication of mono, di- or triphyletic origins.

Both the numerical and ancestral sequence methods placed Macrobrachium apart from the Insecta when using the whole of the cytochrome c molecule as the data input, whilst both methods placed Macrobrachium between the Orthoptera and the Diptera-Lepidoptera node, within the Insecta, when the residues prior to glycine-1 of the standard alignment were ignored (see Figures 36-42). It is clear that the absence of an N-terminal "tail" together with the possession of a blocked N-terminus sets the crustacean sequence apart from those of the Insecta. However, it was not clear whether the absent residues at the N-terminus of the alignment were due to a single or a multiple mutational event, or due to the differing specificities of the enzymes proposed to cleave the N-formyl-methionine region from the polypeptide chain following protein synthesis. This last explanation involves no expression of the cytochrome c gene and therefore precludes the use of those fragments of the N-terminal sequence not common to all the sequences under study. Until the exact reasons for the variations in length at the N-terminus are determined, difficulties will remain regarding the correct computer strategy to adopt for the analysis of small samples of highly variant sequences. The analyses with and without N-terminal "tails" applied in this study have inherited certain distortions associated with the ancestral sequence method. Thus, in Figure 36, the branch-lengths around the Arthropoda and the vertebrate line of descent must be distorted because of the computing of

absent residues at the N-terminus as amino acid differences. The computation of only the "common" areas of the molecule (Figure 38) ignores at least two real differences between Macrobrachium and the Insecta with regard to N-terminal "tails" and blocking groups. However, the sequences in this study represented too small a sample to form a stable phylogeny from both the "common" and "total" treatments. It is probable that the addition of more sequences of cytochrome c representative of each of the major invertebrate phyla, including the Chelicerata and Crustacea, would yield a phylogeny which would remain stable regardless of the treatments of the N-terminal region. For this study, the most realistic approach appears to have been the computation of the common region of the molecule together with an acknowledgement of the presence or absence of N-terminal "tails" and/or N-terminal blocking groups.

The true phylogenetic significance of the Arthropod sequences is, therefore, difficult to assess. Only the determination and computation of further arthropod sequences (as in other areas of the phylogeny) can reduce the relative importance of the choice of "tail" strategy together with the degree of phylogenetic error.

A number of theories with regard to the invertebrate origins of the vertebrates have been reviewed (Garstang, 1928; Fell, 1948; Berril, 1955; Carter, 1957; Bone, 1957; Whitear, 1957; Tarlo, 1960). The ascidians are often thought

of as degenerate descendants of free-swimming ancestors close to the vertebrate line of descent (Gregory, 1951; Eaton, 1953), whilst another school, commencing with Garstang (1928) and supported by Medawar (1951) and Berril (1955), considered that the vertebrate line arose through neoteny from the tadpole larvae of ascidian-like adult forms. Eaton (1970) and Jefferies (1967; 1971; 1973) agreed that the stems of calci-chordates are homologous with those of chordates, hemichordates and crinoids. Jefferies (1967) removed the carpoids, a class of Lower Paleozoic bilateral or asymmetrical animals considered to be primitive echinoderms, and placed them in the Chordata as the sub-phylum Calcichordata. He constructed (Jefferies, 1968) from the skeletal structure of numerous specimens the soft anatomies of species from the orders Cornuta and Mitrata, and concluded that the combined chordate and echinoderm stock originated from a detached, creeping pterobranch of the Hemichordata, rather similar to Cephalodiscus (see Appendices XI and XII). Jefferies (1973) defined the Calcichordata as a primitive sub-phylum of fossil marine chordates having a calcite skeleton of an echinoderm type.They all have a stem and a flattish theca adapted to lie on one side upon the sea floor. Various members of the group show chordate features of dorsal nerve cord and natochord in a post-anal tail, brachial slits in the thecal wall and a fish-like arrangement of cranial nerves. The Calcichordata were very asymmetrical animals and certain of these

asymmetries resemble those of ecent, primitive chordates, suggesting that the chordates are more closely related to the echinoderms than to the hemichordates, which are thought to form a separate, though related phylum (see Barrington, 1965). The chordates and echinoderms are thought to have evolved from a common ancestor resembling the hemichordate Cephalodiscus which had taken to crawling, right-side down on the sea floor and acquired a calcite skeleton (Jefferies, 1969). Of the three living subphyla of chordates, only the vertebrates possess a phosphatic skeleton, but Jefferies (1969) proposed that a calcite skeleton was a primitive feature of the Chordata and was lost in the two lines leading towards the Cephalochordata and the Vertebrata and Urochordata, possibly due to the adoption of a swimming mode of life. Later, it was proposed that the vertebrates only evolved a phosphatic skeleton.

Only the relatively recent estimated time of divergence for the Echinodermata (Table 38) from the vertebrate line of descent obtained in this study, directly supports the ideas of Jefferies (1973) and Eaton (1970). There is a great need for further representative sequences of cytochrome c to be determined from echinoderm and lower chordate sources, in order to help resolve the speculation with regard to the origin of the chordates from the invertebrates.

The findings of the molecular phylogenies can thus be summed up. The estimated times of divergence of the major

invertebrate phyla, and the phylogenies constructed by the ancestral sequence method, did not conflict with the broadly accepted patterns of invertebrate phylogeny based upon the classical studies of morphology, embryology and palaeontology. However, it was clear that particular evolutionary theories (e.g. Vagvolgyi, 1967; Manton, 1973; Jefferies, 1973; Cisne, 1974) could not be rigorously tested using such a small set of data as that applied during this study. The significance of the phylogenies constructed by the ancestral sequence method was difficult to assess, because of the possible errors associated with the treatment of a small data set of such high variance. The observed disagreements with the classical view were well within the estimated limits of error. The application of the numerical method was less successful in coping with the large branch-lengths than was expected from the evidence of model experiments (Peacock & Boulter, 1975). The numerical method yielded phylogenies which disagreed with classical views and were outside the estimated limits of error. The small number of invertebrate sequences available, and the large degree of variance between them, placed undue emphasis on the computation chosen, particularly with regard to the treatment of the N-terminal region of the standard alignment of cytochrome c. It was concluded that a realistic treatment of a small data set was to compute only the common regions of the molecule shared by the sequences, and to note the presence or absence of N-terminal "tails" and/or blocking

groups on the resulting phylogenies. Only the determination of further sequences of cytochrome c from all the major invertebrat phyla will remove the need for such a restrictive approach.

However, it is clearly unrealistic to expect the purification of quantities of cytochrome c sufficient for sequence determination, even on a micro-scale, from the more phylogenetically interesting genera. As noted above, a number of conflicting theories have been put forward to account for the invertebrate origins of the Chordata and the status of any of these may be enhanced by the construction of molecular phylogenies. Such studies would have to include the Echinodermata, the Cephalochordata (e.g. Amphioxos) and the Urochordata (e.g. Ciona, Oikopleura). The yields of cytochrome c and the negative result obtained with the echinoderm Holuthuria (Richardson, M., personal communication) indicate that very large quantities of starting material would be required, and to obtain, for example, more than six species per phylum in these quantities would be very difficult. It is possible that the investigation of a protein other than cytochrome c would be better suited to particular areas of study. However, this would mean the neglect of good quality data in the respect that cytochrome c sequences have been determined from all the major kingdoms of organisms and these have shown a relatively slow rate of acceptance of amino acid substitutions and a low level of convergent changes.

The major exception to the low yields of cytochrome c expected from invertebrates, lies in the phylum Arthropoda whose members may be considered as the most muscularly active, and thus the best cytochrome c sources, within this group of animals. As noted earlier, the Arthropoda have excited interest with regard to their origins and, the evidence of the success of the insect molecular phylogeny constructed in this study, it would appear that the accumulation and analysis of further, more representative arthropod data would be both a realistic and phylogenetically rewarding course of study.

BIBLIOGRAPHY

AMBLER, R. P. (1963). Biochem. J. 89, 341-378.

AMBLER, R. P. (1971). FEBS Letters 18, 351-353.

AMBLER, R. P. & WYNN, M. (1973). Biochem. J. 131, 485-535.

AMBLER, R. P., BRUSCHI, M. & LE GALL, J. (1971). FEBS Letters 18, 347-350.

ANDERSON, D. T. (1973). Embryology and Phylogeny in Annelids and Arthropods. Pergamon, Oxford.

ARMENTROUT, R. W. & DOOLITTLE, R. F. (1969). Arch. Biochem. Biophys. 132, 80-90.

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

AUGUSTEYN, R., McDOWALL, M., WEBB, E. C. & ZERNER, B. (1973). Biochim. Biophys. Acta 303, 1-7.

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

BARLOW, G. H. & MARGOLIASH, E. (1966). J. Biol. Chem. 241, 1473-1477.

BARRINGTON, E. J. W. (1965). The Biology of the Hemichordata and Protochordata. Oliver and Boyd, Edinburgh.

BARTNICKI-GARCIA, S. (1970). In: Phytochemical Phylogeny (Ed. J. B. Harborne), pp. 81-103. Academic Press.

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

BERRIL, N. J. (1955). The Origin of the Vertebrates. Oxford University Press.

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

BOARDMAN, N. K. & PARTRIDGE, S. M. (1953). Nature 171, 208-210.

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

BOARDMAN, N. K. and PARTRIDGE, S. M. (1955). Biochem. J. 59, 543-552.

BOCK, R. M. & LING, N-S. (1954). Analyt. Chem. 26, 1543-1546.

BONE, Q. (1957). Nature 180, 1462-1464.

BOULTER, D. (1970). Ann. Rev. Plt. Physiol. 21, 91-114.

BOULTER, D. (1972). Sci. Prog. Oxf. 60, 217-229.

BOULTER, D. (1973). Pure and Applied Chem. 34, 539-552.

BOULTER, D. (1974). Curr. Adv. Plant Sci. 8, 1-15.

BOULTER, D. & RAMSHAW, J. A. M. (1972). Phytochem. 11, 553-561.

BOULTER, D., THURMAN, D. A. & TURNER, B. L. (1966). Taxon 15, 135-143.

BOULTER, D., THOMPSON, E. W., RAMSHAW, J. A. M. & RICHARDSON, M. (1970). Nature 228, 552.

BOULTER, D., RAMSHAW, J. A. M., THOMPSON, E. W., RICHARDSON, M. & BROWN, R. H. (1972). Proc. Roy. Soc. London B. 181, 441-445.

BØVRE, K. & SZYBALSKI, W. (1971). In: Methods in Enzymology, vol. 21 (Eds. S. P. Colowick and N. O. Kaplan), pp. 350-383. Academic Press, N.Y.

BROWN, R. H. & BOULTER, D. (1973a). Biochem. J. 131, 247-251.

BROWN, R. H. & BOULTER, D. (1973b). Biochem. J. 133, 251-254.

BROWN, R. H. & BOULTER, D. (1974). Biochem. J. 137, 93-100.

BROWN, R. H., RICHARDSON, M., BOULTER, D., RAMSHAW, J. A. M. & JEFFERIES, R. P. S. (1972). Biochem. J. 128, 971-974.

BROWN, R. H., RICHARDSON, M., SCOGIN, R. & BOULTER, D. (1973). Biochem. J. 131, 253-256.

BRYSON, V. & VOGEL, H. J. (1965). In: Evolving Genes and Proteins. Academic Press, N.Y.

CANTOR, C. R. & JUKES, T. H. (1966). Proc. Natl. Acad. Sci. U.S.A. 56, 177-184.

CARBON, J., BERG, P. & YANOFSKY, C. (1966). Cold Spring Harb. Symp. Quant. Biol. 31, 487-498.

CARPENTER, F. H. & BAUM, W. E. (1962). J. Biol. Chem. 237, 409.

CARTER, G. S. (1954). Syst. Zool. 4, 163-167.

CARTER, G. S. (1957). Syst. Zool. 6, 187-192.

CHAN, S. K. (1970). Biochim. Biophys. Acta 221, 497-501.

CHAN, S. K. & MARGOLIASH, E. (1966). J. Biol. Chem. 241, 335-338.

CHAN, S. K., TULLOSS, L. & MARGOLIASH, E. (1966). Biochemistry 5, 2586-2597.

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

CISNE, J. L. (1974). Science 186, 13-18.

CLARK, R. B. (1964). Dynamics in Metazoan Evolution. Clarendon Press, Oxford.

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

COWIE, J. W. (1967). In: The Fossil Record (Ed. W. B. Harland). Geological Society of London.

CRICK, F. H. C. (1966). Sci. American 215, 217-223.

CRICK, F. H. C., BARNETT, L., BRENNER, S. & WATTS-TOBIN, R. J. (1961). Nature 192, 1227-2132.

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

DAHL, E. (1969). Opusc. Ent. 34, 1-9.

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

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

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

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

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

DELANGE, R. J., GLAZER, A. N. & SMITH, E. L. (1969). J. Biol. Chem. 244, 1385-1388.

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

DICKERSON, R. E. (1971). J. Molec. Biol. 57, 1-15.

DICKERSON, R. E., TAKANO, T., EISENBERG, D., KALLAI, O. B., SAMSON, L., COOPER, A. & MARGOLIASH, E. (1971). J. Biol. Chem. 246, 1511-1535.

DIXON, M. (1953). Biochem. J. 54, 457-458.

DOOLITTLE, R. F. & ARMENTROUT, R. W. (1968). Biochem. 7, 516-521.

DOUGHERTY, E. C., BROWN, Z. N., HANSON, E. D. & HARTMAN, W. D. (1963). The Lower Metazoa - Comparative Biology and Phylogeny. Univ. of California Press, Los Angeles.

DUS, K., SLETTEN, K. & KAMEN, M. D. (1968). J. Biol. Chem. 243, 5507-5518.

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

EATON, T. H. (1953). Syst. Zool. 2, 1-6.

EATON, T. H. (1970). J. Paleont. 44, 969-979.

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

FELL, H. B. (1948). Biol. Rev. 23, 81-107.

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

FERNANDEZ-SOUSA, J. M., GAVILANES, J. G., MUNICIO, A. M., PERERA, J., PEREZ-ARANDA, A. & RODRIGUEZ, R. (1974). FEBS Letters 41, 315-316.

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

FITCH, W. M. (1966). J. Molec. Biol. 16, 17-27.

FITCH, W. M. (1967). J. Molec. Biol. 26, 499-507.

FITCH, W. M. (1969). Biochem. Genet. 3, 99-108.

FITCH, W. M. (1970). J. Molec. Biol. 49, 1-14.

FITCH, W. M. & MARGOLIASH, E. (1967a). Science 155, 279-284.

FITCH, W. M. & MARGOLIASH, E. (1967b). Biochem. Genet. 1, 65-71.

FITCH, W. M. & MARGOLIASH, E. (1969). Brookhaven Symp. Biol. 21, 217—241.

FITCH, W. M. & MARGOLIASH, E. (1970). Evolutionary Biology 4, 67-109.

FITCH, W. M. & MARKOWITZ, E. (1970). Biochem. Genet. 4, 579-593.

FLATMARK, T. (1964). Acta Chem. Scand. 18, 1517-1527.

FLATMARK, T. (1966). Acta Chem. Scand. 20, 1487-1496.

FLATMARK, T. & VESTERBERG, O. (1966). Acta Chem. Scand. 20, 1497-1503.

FRIDMAN, C., LIS, H., SHARON, N. & KATCHALSKI, E. (1968). Arch. Biochem. Biophys. 126, 299-304.

GARSTANG, W. (1928). Quart. J. Micr. Sci. 72, 51-187.

GHIRETTI-MAGALDI, A., GIUDITTA, A. & GHIRETTI, F. (1956). Biochem. J. 66, 303.

GHIRETTI, A., GHIRETTI-MAGALDI, A. & TOSI, L. (1959). J. Gen. Physiol. 42, 1185.

GIBBS, A. J & McINTYRE, G. A. (1970). Eur. J. Biochem. 16, 1-11.

GILLESPIE, D. (1968). In: Methods in Enzymology, vol. 12 (Eds. S. P. Colowick & N. P. Kaplan), pp. 641-668. Academic Press, N.Y. and London.

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

GRANTHAM, R. (1974). Nature 248, 791-793.

GRAY, W. R. (1967). In: Methods in Enzymology, vol. 11 (Eds. S. P. Colowick & N. P. Kaplan), pp. 139-151. Academic Press, N.Y. and London.

GRAY, W. R. (1972). In: Methods in Enzymology, vol. 25 (Eds. S. P. Colowick & N P. Kaplan), pp. 121-138. Academic Press, N.Y. and London.

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

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

GRAY, W. R. & SMITH, J. F. (1970). Anal. Biochem. 33, 36-42.

GREGORY, S. (1951). Evolution Emerging. New York.

GROBBEW, K. (1908). Verh. Zool. Bot. Ges. Wien. 58, 491.

GROS, C. & LABOUESSE, B. (1969). Eur. J. Biochem. 7, 463-470.

GURTLER, L. & HORSTMANN, J. (1970). Eur. J. Biochem. 12, 48-57.

HADZI, J. (1953). Syst. Zool. 2, 145-153.

HADZI, J. (1963). The Evolution of the Metazoa. Pergamon, Oxford.

HAGIHARA, B., TAGAWA, K., MORIKAWA, I., SHIN, M. & OKUNUKI, K. (1958). Nature 181, 1590-1591.

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

HANSON, E. D. (1961). Animal Diversity. Prentice-Hall, New Jersey.

HARDY, A. C. (1953). Quart. J. Micr. Sci. 94, 441-443.

HARLAND, W. B., HOLLAND, C. H., HOUSE, M. R., HUGHES, W. F., REYNOLDS, A. B., RUDWICK, M. J. S., SLATTERTWAITE, G. E., TARLO, W. B. H. & WILLEY, E. C. (1967). The Fossil Record. Geological Society of London.

HAWKES, J. G. (1968). In: Chemotaxonomy and Serotaxonomy. Academic Press, N.Y. and London.

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

HELLER, J. & SMITH, E. L. (1966). J. Biol. Chem. 241, 3158-3180.

HILDEMANN, W. H. (1974). Life Sciences 14, 605-614.

HILL, R. L., DELANEY, R., FELLOWS, R. E. Jnr. and LEVOVITZ, H. C. (1960). Proc. Natl. Acad. Sci. U.S.A. 56, 1762-1769.

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

HIRS, C. H. W. (1956). J. Biol. Chem. 219, 611-621.

HOWARD, N. L., JOUBERT, F. J. & STRYDOM, P. J. (1974). Comp. Biochem. Physiol. 48, 75-85.

HYMAN, L. H. (1940-1959). The Invertebrates, vol. 1-5. McGraw Hill, N.Y.

ILSE, D. & EDMAN, D. E. (1963). Aust. J. Chem. 16, 411-416.

INAGAMI, T. (1972). Proteins - Structure and Function, vol. 1, p. 1 (Ed. M. Funatsu). John Wiley and Sons.

INAGAMI, T. & STURTEVANT, J. M. (1960). J. Biol. Chem. 235, 1019-1023.

INGLISH, A. S., NICHOLLS, P. W. & ROXBORGH, C. M. (1971). Aust. J. Biol. Sci. 24, 1247.

JACOBS, E. E. & SAWADI, D. R. (1960). J. Biol. Chem. 235, 531-534.

JAGERSTEN, G. (1955). Zool. Bidrag, Uppsala 30, 321-354.

JAGERSTEN, G. (1959). Zool.Bidrag, Uppsala 33, 79-107.

JAGERSTEN, G. (1972). Evolution of the Metazoan Life Cycle. Academic Press.

JARDINE, N., VAN RUSBERGEN, C. J. & JARDINE, C. J. (1969). Nature 224, 185.

JEFFERIES, R. P. S. (1967). Symp. Zool. Soc. (London) 20, 163-208.

JEFFERIES, R. P. S. (1968). Bull. Bi. Museum (Nat. Hist.) Geology 16, 243-339.

JEFFERIES, R. P. S. (1969). Paleontology 12, 494-535.

JEFFERIES, R. P. S. (1973). Phil. Trans. Roy. Soc. Lond. 265B, 409-469.

JEPPSSON, J. O. and SJÖQUIST, J. (1967). Analyt. Biochem. 18, 264-269.

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

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

KASPER, C. B. (1970). In: Molec. Biol. Biochem. and Biophys. 8, 139-184. (Ed. S. B. Needleman).

KEIL-DLOUHA, V., ZYLBER, N., IMHOFF, J. M., TONG, N. T. & KEIL, B. (1971). FEBS Letters 16, 287-290.

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

KERKUT, G. A. (1960). The Implications of Evolution. Pergamon Press, London.

KOHNE, D. E. (1968). In: Chemotaxonomy and Serotaxonomy (Ed. J. G. Hawkes). Academic Press.

KOSTKA, V. & CARPENTER, F. H. (1964). J. Biol. Chem. 239, 1799-1803.

LANCE, G. W. & WILLIAMS, W. T. (1966). Nature 212, 218.

LANCE, G. W. & WILLIAMS, W. T. (1967). Computer J. 9, 373-380.

LAYCOCK, M. V. (1972). Canad. J. Biochem. 50, 1311-1325.

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

LEMECHE, H. (1959). Proc. Int. Cong. Zool., vol. XV, 381-389.

MANTON, S. M. (1967). J. Nat. Hist. 1, 1-22.

MANTON, S. M. (1973). J. Zool. 171, 111-130.

MANSKI, I. W., HALBERT, S. P., JAVIER, P. and AVERBACH-PASCAL, T. (1967). Int. Archs. Allergy Appl. Immun. 31, 529-545.

MARCUS, E. (1958). Quart. Rev. Biol. 33, 24-58.

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

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

MARGOLIASH, E. (1962). Brookhaven Symp. Biol. 15, 266-281.

MARGOLIASH, E. (1963). Proc. Natl. Acad. Sci. U.S.A. 50, 672-679.

MARGOLIASH, E. (1964). Can. J. Biochem. 42, 745-753.

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

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

MARGOLIASH, E. & FITCH, W. M. (1968). Ann. N.Y. Acad. Sci. 151, 359-381.

MARGOLIASH, E. & FROWHIRT, W. (1959). Biochem. J. 71, 570-572.

MARGOLIASH, E. & LUSTGARTEN, J. (1962). J. Biol. Chem. 237, 3397-3405.

MARGOLIASH, E. & SCHEJTER, A. (1966). Adv. Prot. Chem. 21, 113-280.

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

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

MARGOLIASH, E., KIMMEL, J. R., HILL, R. L. & SCHMIDT, W. T. (1962). J. Biol. Chem. 237, 2148-2150.

MARGOLIASH, E., NEEDLEMAN, S. B. & STEWART, J. W. (1963). Acta Chem. Scand. 17, 250-256.

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

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

MATSUBARA, H. & SASAKI, R. M. (1969). Biochem. Biophys. Res. Commun. 35, 171-181.

MATSUBARA, H. & SMITH, E. L. (1962). J. Biol. Chem. 238, 2732-2753.

MATSUBARA, H., SASAK, R. M. & CHAIN, R. K. (1967). Proc.Natl. Acad. Sci. U.S.A. 57, 441-445.

McDOWALL, M. A. & SMITH, E. L. (1965). J. Biol. Chem. 240, 4635-4647.

McLACHLAN, A. D. (1971). J. Molec. Biol. 61, 409-424.

McLAUGHLIN, P. J. & DAYHOFF, M. O. (1973). J. Molec. Evol. 2, 99-116.

MEATYARD, B. T. M. (1974). Ph.D. Thesis, University of Durham.

MEDAWAR, P. B. (1951). Nature 167, 852-853.

MIWAKAMI, S., TITAWI, K., ISHIHORA, H. & TAKAHASHI, K. (1958). J. Biochem.45, 547-549.

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

MOORE, G. W. , GOODMAN, M. & BARNABAS, J. (1973). J. Theoret. Biol. 38, 423-445.

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

MORITA, Y. & IDA, S. (1972). J. Mol. Biol. 71, 807-808.

MORITA, Y., YAGI, F., IDA, S., ASADA, K. & TAKAHASHI, M. (1973). FEBS Letters 31, 186-188.

MORGAN, W. T., HENSLEY, C. P. & RIEHM, J. P. (1972). J. Biol. Chem. 247, 6555-6565.

MORTON, J. E. & YONGE, C. M. (1964). In: Physiology of Mollusca, vol. I, pp. 1-58 (Eds. K. M. Wilbur & C. M. Yonge). Academic Press, New York.

NEEDLEMAN, S. B. (Ed.). Molecular Biol. Biochem. and Biophys. 8, pp. 345.

NEEDLEMAN, S. B. & BLAIR, T. T. (1969). Proc. Natl. Acad. Sci. U.S.A. 63, 1227-1233.

NEEDLEMAN, S. B. & MARGOLIASH, E. (1966). J. Biol. Chem. 241, 853-863.

NIALL, H. D. (1973). In: Methods in Enzymology, vol. 17 (Eds. S. P. Colowick and N. O. Kaplan), p. 942. Academic Press, N.Y.

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

NOLAN, C. & MARGOLIASH, E. (1966). J. Biol. Chem. 241, 1049-1059.

NOLAN, C. & MARGOLIASH, E. (1968). Ann. Rev. Biochem. 37, 727-779.

NOVOTNY, J. (1971). FEBS Letters 14, 7-10.

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

PEACOCK, D. & BOULTER, D. (1975). In Press.

PERCY, M. E. & BUCHWALD, B. M. (1972). Anal. Biochem. 45, 60-67.

PETTIGREW, G. W. (1972). FEBS Letters 22, 64-66.

PETTIGREW, G. W. (1973). Nature 241, 531-533.

PISANO, J. J., BRONLERT, T. J. & BREWER, H. B. (1972). Anal. Biochem. 45, 43-60.

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

RAMSHAW, J. A. M. & BOULTER, D. (1975). Phytochemistry, in press.

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.

RICHARDSON, M., LAYCOCK, M. V., RAMSHAW, J. A. M., THOMPSON, E. W. & BOULTER, D. (1970). Phytochem. 9, 2271-2280.

RICHARDSON, M., RICHARDSON, D., RAMSHAW, J. A. M., THOMPSON, E. W. & BOULTER, D. (1971a). J. Biochem. 69, 811-813.

RICHARDSON, M., RAMSHAW, J. A. M. & BOULTER, D. (1971b). Biochim. Biophys. Acta 251, 331-333.

RIFKIN, D. B., HIRSH, D. I., RITKIN, M. R. & KONIGSBERG, W. (1966). Cold Spring Harb. Symp. 31, 715-718.

ROMER, A. S. (1966). Vertebrate Palaeontology (3rd edn.). University of Chicago Press.

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

SCHMER, C. & KREIL, G. (1969). Analyt. Biochem. 29, 186-192.

SCOTT, W. A. & MITCHELL, H. K. (1969). Biochemistry 8, 4282-4289.

SHAROV, A. G. (1966). Basic Arthropodan Stock. Oxford.

SHROCK, R. R. & TWENHOFEL, W. H. (1953). Principles of Invertebrate Palaeontology. McGraw-Hill, New York.

SIEWING, R. (1960). Z. Wiss. Zool. 164, 238-270.

SLONIMSKI, P. P., SCHER, R., PERE, G., SELS, A. & SOMLO, M. (1965). In: Mechanismes de Regulation des Activites Cellulaires chez les Microorganisms, p. 435. Du Centre National dela Recherche Scientifique, Paris.

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

SMITH, E. L. (1968). In: Structure and Function of Cytochromes. (Eds. K. Okunuki, M. D. Kamen and I. Sekuzu), pp. 282-288. University of Tokyo Press.

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

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

STANIER, R. Y. (1974). In: Evolution in the Microbial World (Eds. M. J. Carlile and J. J. Skehel). Cambridge University Press.

STEERS, E. Jnr., CRAVEN, G. R., ANFINSEN, C. B. & BETHUNE, J. L. (1965). J. Biol. Chem. 240, 2478-2484.

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

STEVENS, F. C., GLAZER, A. N. & SMITH, E. L. (1967). J. Biol. Chem. 242, 2764-2779.

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

SWAIN, T. (1963). In: Chemical Plant Taxonomy (Ed. Swain, T.). Academic Press, London and New York.

TARLO, L. B. (1960). Proc. Int. Pal. Um. (XXI Int. Geol. Congr.), part 22, 113-123.

THOMPSON, E. W., RICHARDSON, M. & BOULTER, D. (1970). Biochem. J. 121, 439-446.

THOMPSON, E. W., RICHARDSON, M. & BOULTER, D. (1971). Biochem. J. 124, 779-791.

TIEGS, O. W. & MANTON, S. M. (1958). Biol. Rev. 33, 255-337.

TUPPY, H. (1957). Z. Naturforsch 12, 784.

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

VAGVOLGYI, J. (1967). Syst. Zool. 16, 153-167.

VAUGHAN, J. G. (1968a). Science Prog. Oxford 56, 205-222.

VAUGHAN, J. G. (1968b). In: Chemotaxonomy and Serotaxonomy (Ed. J. G. Hawkes), p.93. Systematics Association Special Volume No. 2.

WALASEK, O. F. & MARGOLIASH, E. (1969). Unpublished by Margoliash et al. (1969) (q.v.).

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

WASSERMAN, A. R., GAVER, J. C. & BURRIS, R. H. (1963). Phytochem. 2, 7.

WEINER, A. M., PLATT, T. & WEBER, K. (1972). J. Biol. Chem. 247, 3242-3251.

WHITEAR, M. (1957). Ann. Mag. Nat. Hist. Ser. 12. 10, 338.

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

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

WU, R., PADMANABHAW, P. & BAMBARA, R. (1974). In: Methods in Enzymology, vol. 12 (Eds. S. P. Colowick and N. O. Kaplan), pp. 231-253. Academic Press, New York.

YAMADA, S. & ITANO, H. A. (1966). Biochim. Biophys. Acta 130, 538-540.

YAMANAKA, T. (1966). Ann. Report Biol. Works Fac. Sci. Osaka Univ. 14, 47.

YAMANAKA, T. (1967). Nature 213, 1183-1186.

YAMANAKA, T. & KAMEN, M. D. (1965). Biochim. Biophys. Acta 96, 328-330.

YAMANAKA, T. & KAMEN, M. D. (1967). Biochim. Biophys. Acta 143, 425-426.

YAMANAKA, T. & OKUNUKI, K. (1963). Biochim. Biophys. Acta 67, 379.

YAMANAKA, T., MIZUSHIMA, H., NOZAKI, M., HORIO, T. & OKUNUKI, K. (1959). J. Biochem. (Tokyo) 46, 121-132.

YAMANAKA, T., TOKUYAMA, S. & OKUNUKI, K. (1963). Biochim. Biophys. Acta 77, 592-601.

YAMANAKA, T., MIZUSHIMA, H., KATANO, H. & OKUNUKI, K. (1964a). Biochim. Biophys. Acta 81, 223-228.

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

YONGE, C. M. (1957). Nature 179, 672-673.

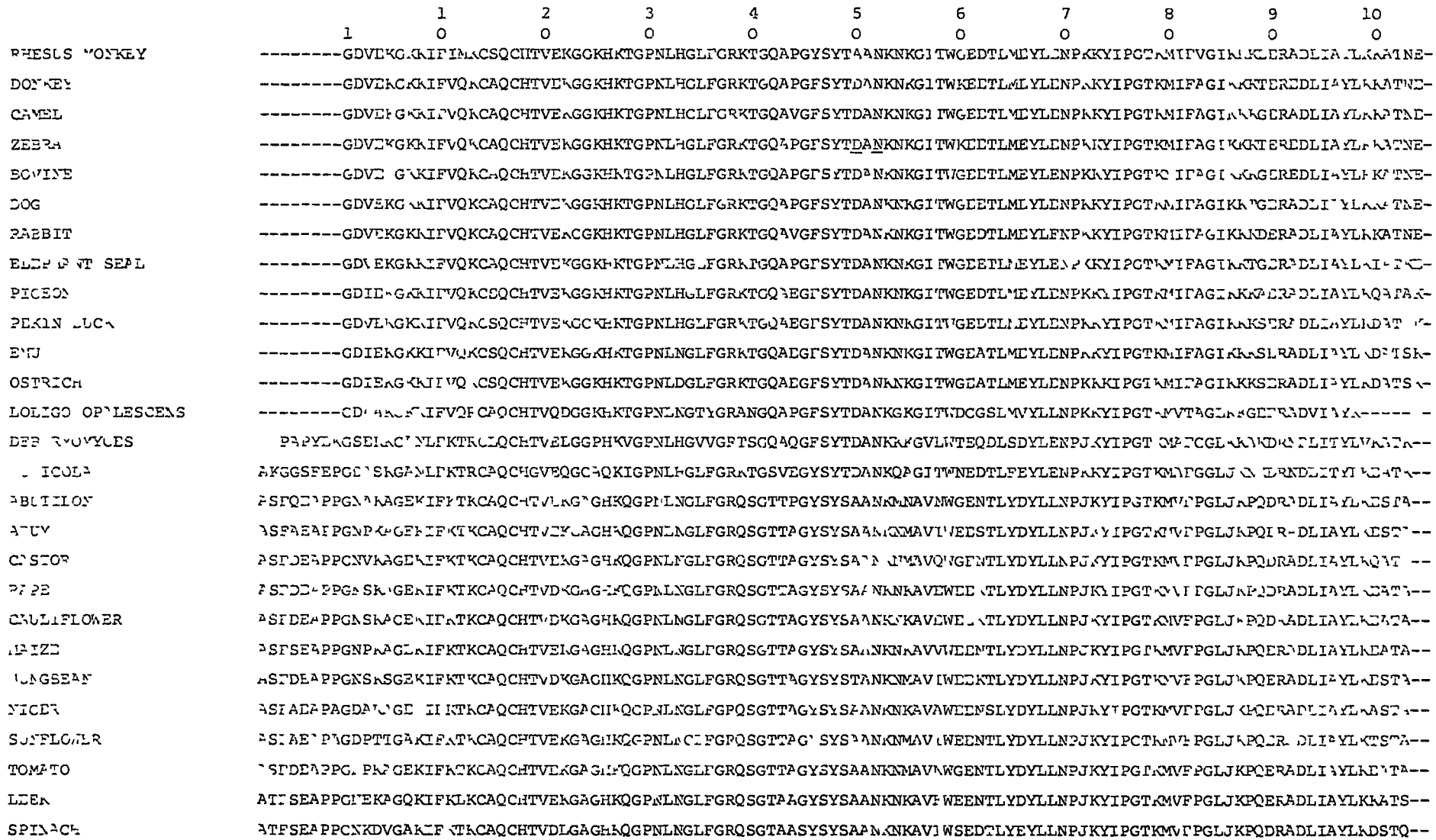
YONGE, C. M. (1960). In: Treatise on Invertebrate Palaeontology (Ed. R. C. Moore), pp. 13-136. University of Kansas Press.

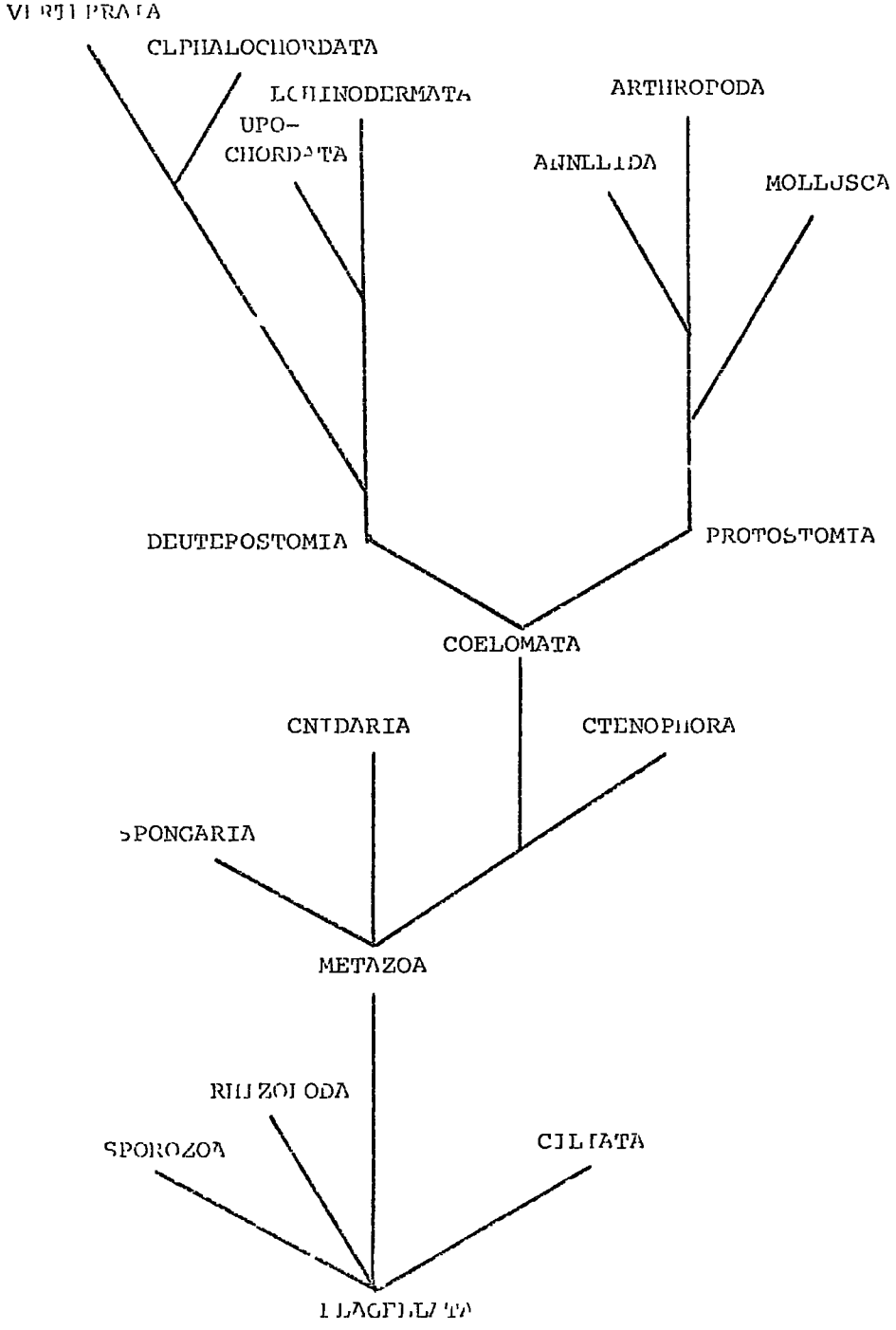
YOSHIDA, T. (1966). Ann. Report Biol. Works Fac. Sci. Osaka Univ. 14, 1-37.

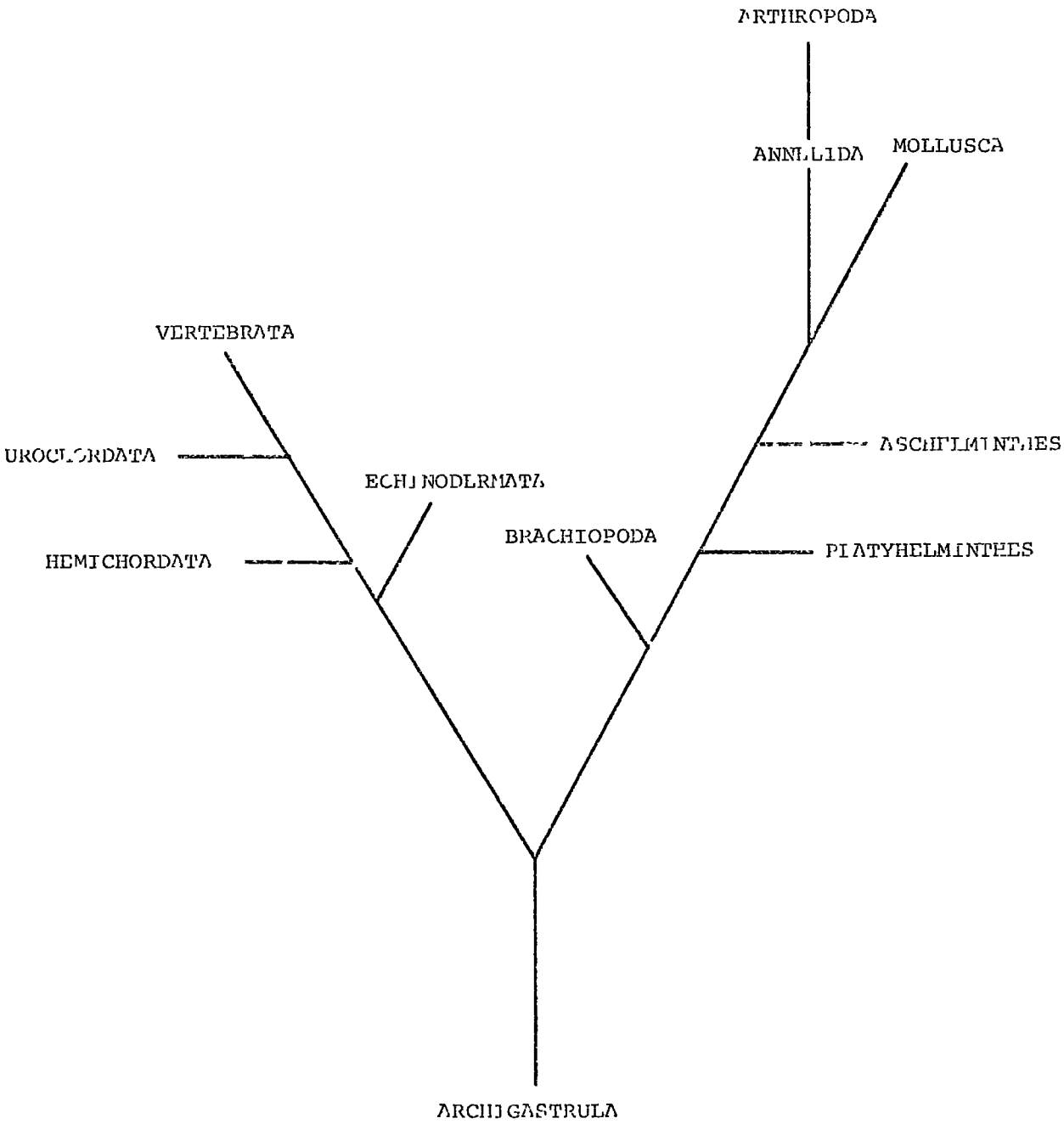
ZUCKERKANDL, E. & PAULING, L. (1965). J. Theoret. Biol. 8, 357-369.

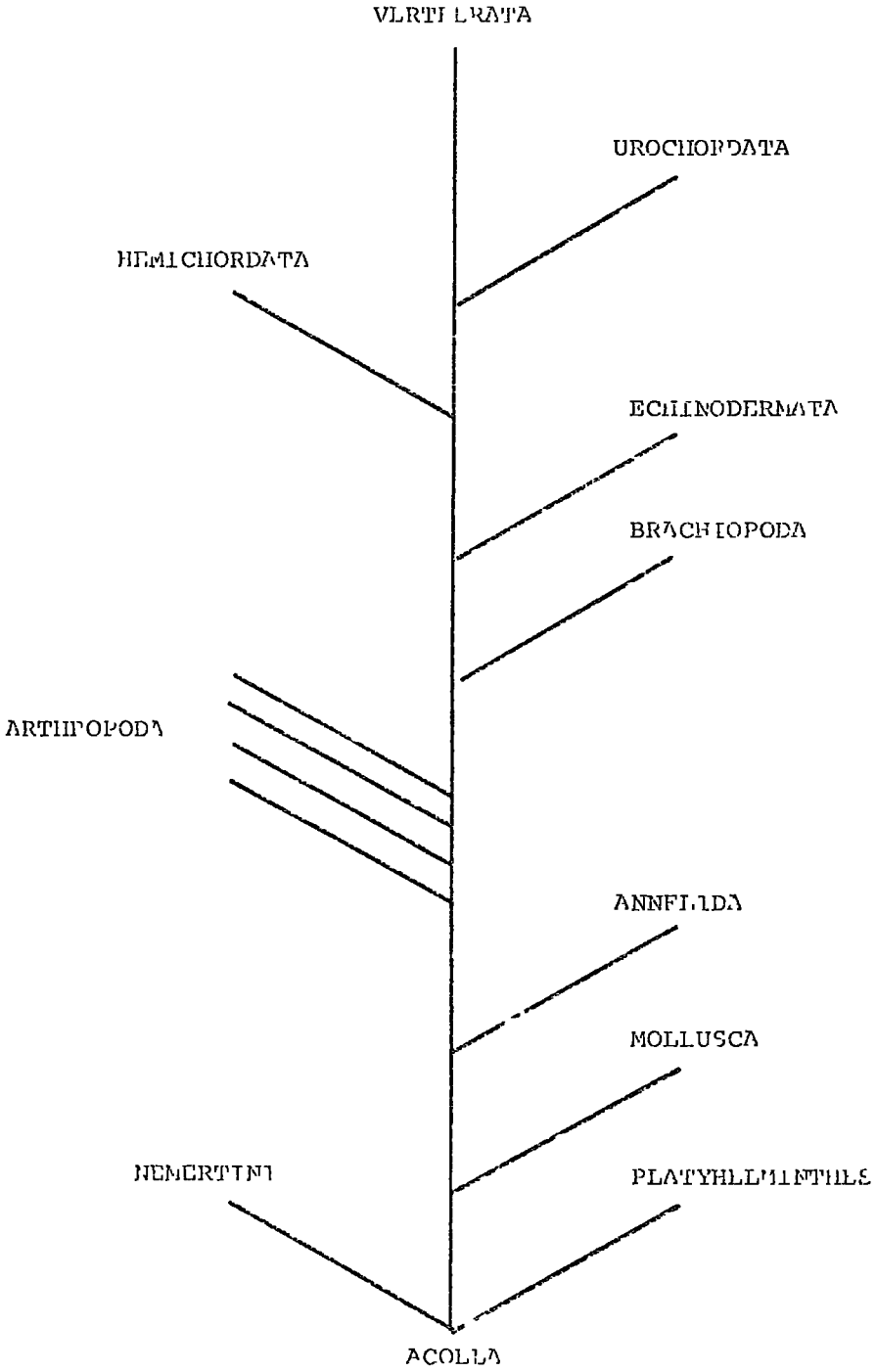
A P P E N D I C E S

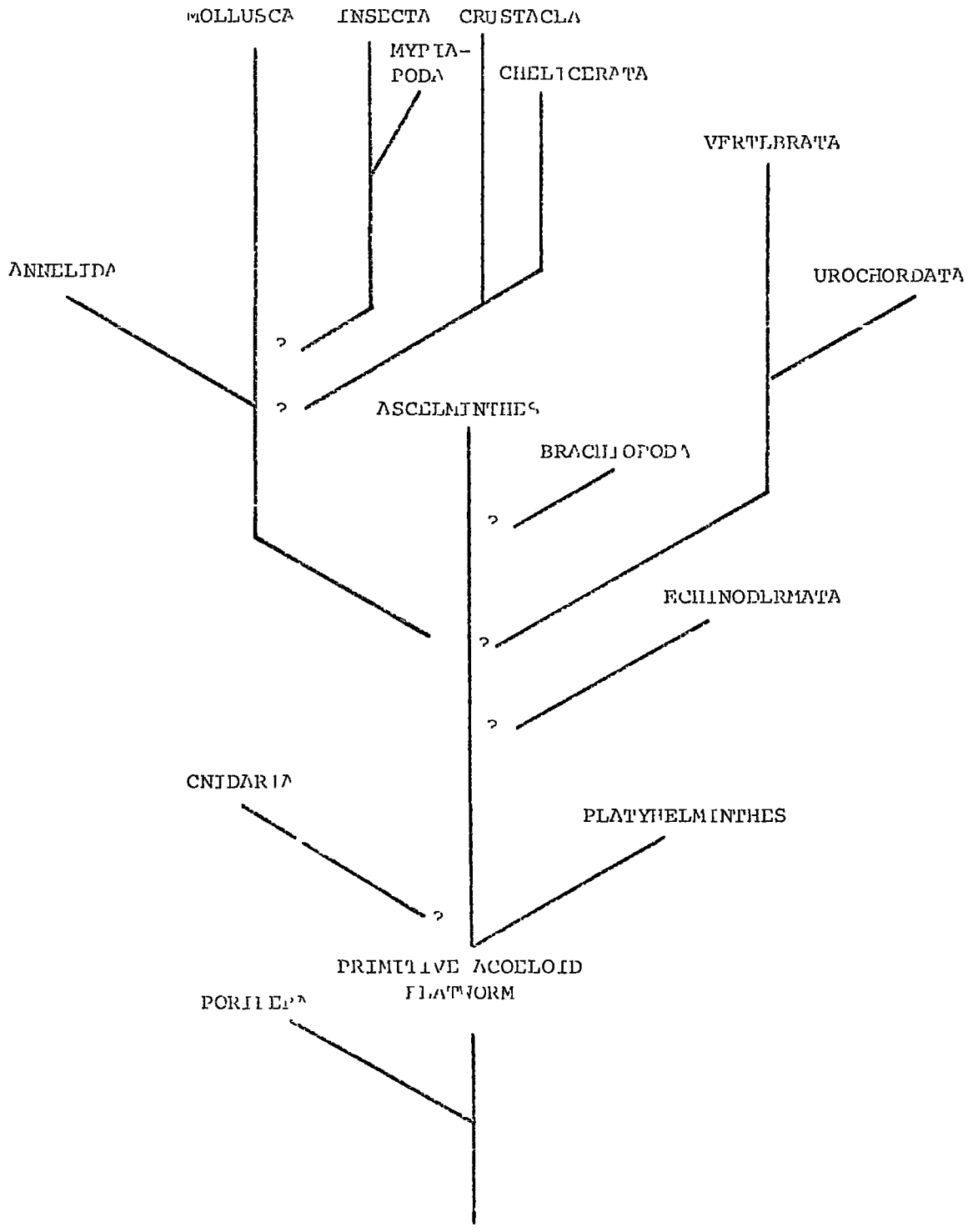
	HUMAN	HORSE	WHALE	KANGAROO	CHICKEN	PIGEON	TURTLE	RATTLESNAKE	BULL-FROG	TUNA	BONITO	CARP	LAMPREY	DOGFISH	FRUIT FLY	CERATITIS	SCREW-WORM FLY	TOBACCO H W MOTH	SAMIA	LOCUST	PRAWN	SNAIL	WORM	STARFISH	SACCHAROMYCES	CANDIDA	NEUROSPORA	USTILAGO	PHYSARUM	ENTEROMORPHA	GINKGO	WHEAT	BUCKWHEAT	PUMPKIN	COTTON	SESAME	ELDER	CRITHIDIA	EUGLENA	RHODOSPIRILLUM
HUMAN	0																																							
HORSE	12	0																																						
WHALE	10	5	0																																					
KANGAROO	10	7	6	0																																				
CHICKEN	13	11	9	12	0																																			
PIGEON	11	10	7	10	4	0																																		
TURTLE	15	11	8	11	8	5	0																																	
RATTLESNAKE	14	22	19	21	19	17	22	0																																
BULL-FROG	13	14	11	13	11	11	10	24	0																															
TUNA	21	19	17	18	17	17	18	26	15	0																														
BONITO	21	18	16	18	17	17	17	27	15	2	0																													
CARP	19	14	12	14	16	15	14	26	14	8	8	0																												
LAMPREY	20	16	15	17	18	18	19	27	21	19	19	12	0																											
DOGFISH	24	17	16	20	19	17	19	26	20	20	20	15	17	0																										
FRUIT FLY	29	24	24	26	25	[illegible]	24	31	22	25	26	23	29	26	0																									
CERATITIS	30	22	22	24	23	22	23	32	30	23	23	21	27	26	3	0																								
SCREW-WORM FLY	27	22	22	24	23	22	24	29	22	24	25	22	28	25	2	5	0																							
TOBACCO H W MOTH	[illegible]	28	27	28	28	26	29	33	30	30	31	27	33	31	14	17	12	0																						
SAMIA CYNTHIA	31	29	27	28	28	27	28	[illegible]	[illegible]	32	33	28	32	32	15	17	14	5	0																					
LOCUST	25	21	20	23	23	21	23	28	26	27	25	22	23	22	17	15	15	18	15	0																				
PRAWN	29	24	23	25	26	22	25	32	37	30	31	27	28	26	26	26	24	26	[illegible]	[illegible]	0																			
SNAIL	29	[illegible]	[illegible]	26	23	24	25	28	28	30	31	26	26	26	29	28	27	29	29	25	30	0																		
WORM	36	32	31	31	33	30	34	34	35	37	38	34	36	35	27	26	25	29	30	25	33	32	0																	
STARFISH	28	20	20	21	23	22	20	28	22	26	27	24	24	24	25	24	24	27	27	23	26	22	28	0																
SACCHAROMYCES	45	46	45	46	46	46	49	48	47	47	45	42	49	48	45	43	45	45	47	44	54	49	43	46	0															
CANDIDA	51	51	50	51	50	50	52	52	51	47	46	50	55	52	47	45	47	46	47	44	53	51	47	49	27	0														
NEUROSPORA	[illegible]	42	42	45	43	42	43	44	45	44	22	44	48	45	37	35	37	42	43	36	43	46	43	45	36	38	0													
USTILAGO	47	44	43	47	[illegible]	41	45	47	46	45	42	43	47	47	42	40	40	42	42	38	48	44	41	45	37	31	34	0												
PHYSARUM	47	43	[illegible]	47	49	44	49	50	48	51	52	50	50	50	50	50	50	52	51	50	52	50	51	48	57	54	52	54	0											
ENTEROMORPHA	46	46	44	44	45	44	43	48	47	47	[illegible]	45	50	43	45	43	3	43	45	40	47	50	45	47	50	51	49	50	57	0										
GINKGO	44	47	44	47	45	44	45	46	47	50	49	49	53	53	49	48	47	46	48	46	52	54	50	52	49	55	52	53	57	31	0									
WHEAT	45	48	45	49	48	48	43	43	50	51	49	50	53	51	49	48	47	44	47	44	51	53	50	53	47	50	51	50	58	23	17	0								
BUCKWHEAT	42	46	43	44	42	43	42	43	45	46	48	48	51	50	46	35	44	42	41	42	49	51	47	47	48	54	51	52	54	27	14	17	0							
PUMPKIN	45	49	48	47	50	48	48	46	52	52	51	50	54	53	52	51	50	46	47	45	52	53	51	52	49	51	50	50	54	26	20	13	15	0						
COTTON	47	51	48	49	50	48	48	46	51	53	51	52	54	53	52	51	50	47	45	44	53	54	50	52	48	51	51	50	54	25	19	13	12	8	0					
SESAME	42	47	45	44	48	45	45	43	49	50	46	48	52	51	50	49	48	45	45	42	50	52	50	50	49	52	51	52	51	26	19	15	12	7	7	0				
ELDER	48	52	50	51	51	48	50	48	53	54	53	53	54	54	53	52	51	48	49	47	53	56	50	53	50	53	54	52	54	25	18	13	14	8	7	11	0			
CRITHIDIA	59	56	55	55	57	56	54	63	56	56	56	64	67	69	65	65	64	65	65	65	68	70	53	55	59	57	63	54	59	52	56	57	53	54	53	53	55	0		
EUGLENA	49	47	48	48	50	50	48	52	49	50	49	48	[illegible]	51	54	54	55	55	53	54	51	54	57	51	62	61	53	54	51	[illegible]	[illegible]	61	61	58	59	60	60	57	0	
RHODOSPIRILLUM	64	62	63	64	62	62	62	63	63	63	62	64	66	64	65	64	64	64	65	66	66	63	69	69	73	78	70	73	68	72	72	72	71	72	71	72	72	71	70	0

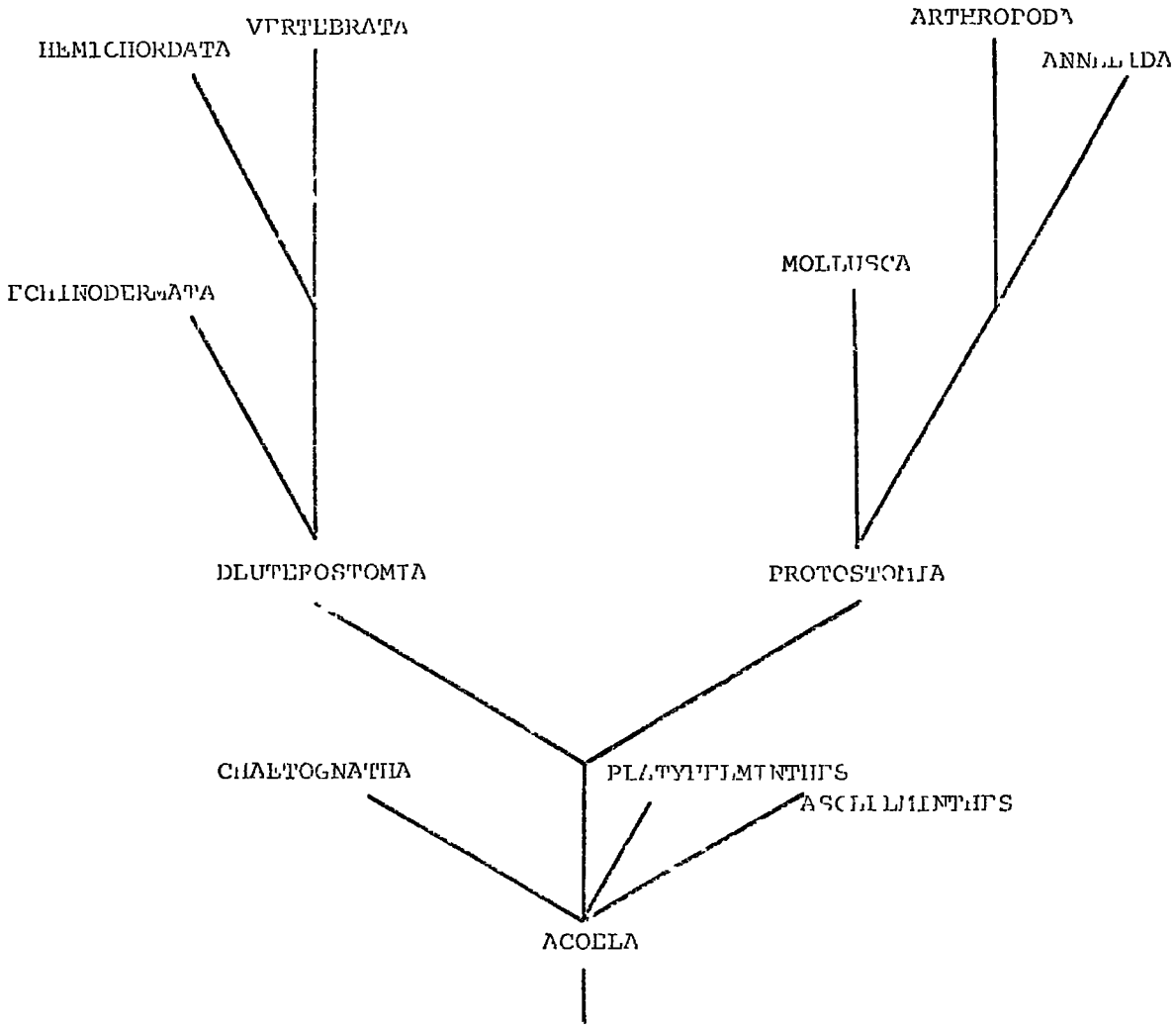


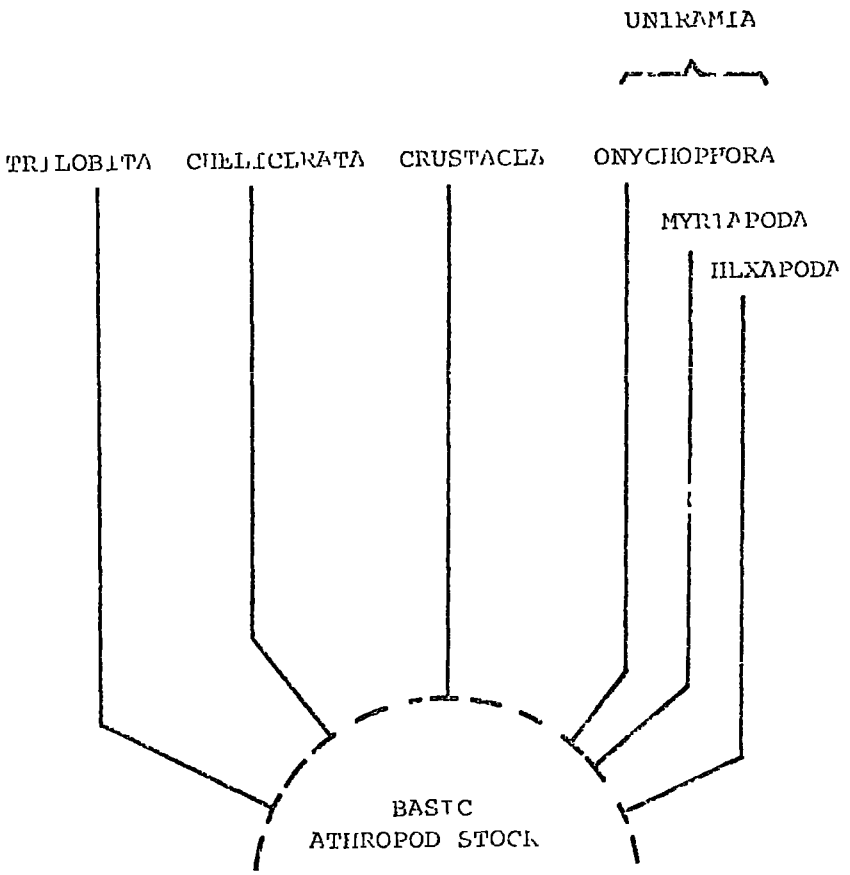

VERTEBRATA
CEPHALOCHORDATA
ECHINODERMATA
UPO-
CHORDATA
ARTHROPODA
ANNELIDA
MOLLUSCA
DEUTEROSTOMIA
PROTOSTOMIA
COELOMATA
CNIDARIA
CTENOPHORA
SPONGARIA
METAZOA
RHIZOPODA
SPOROZOA
CILIATA
FLAGELLATA

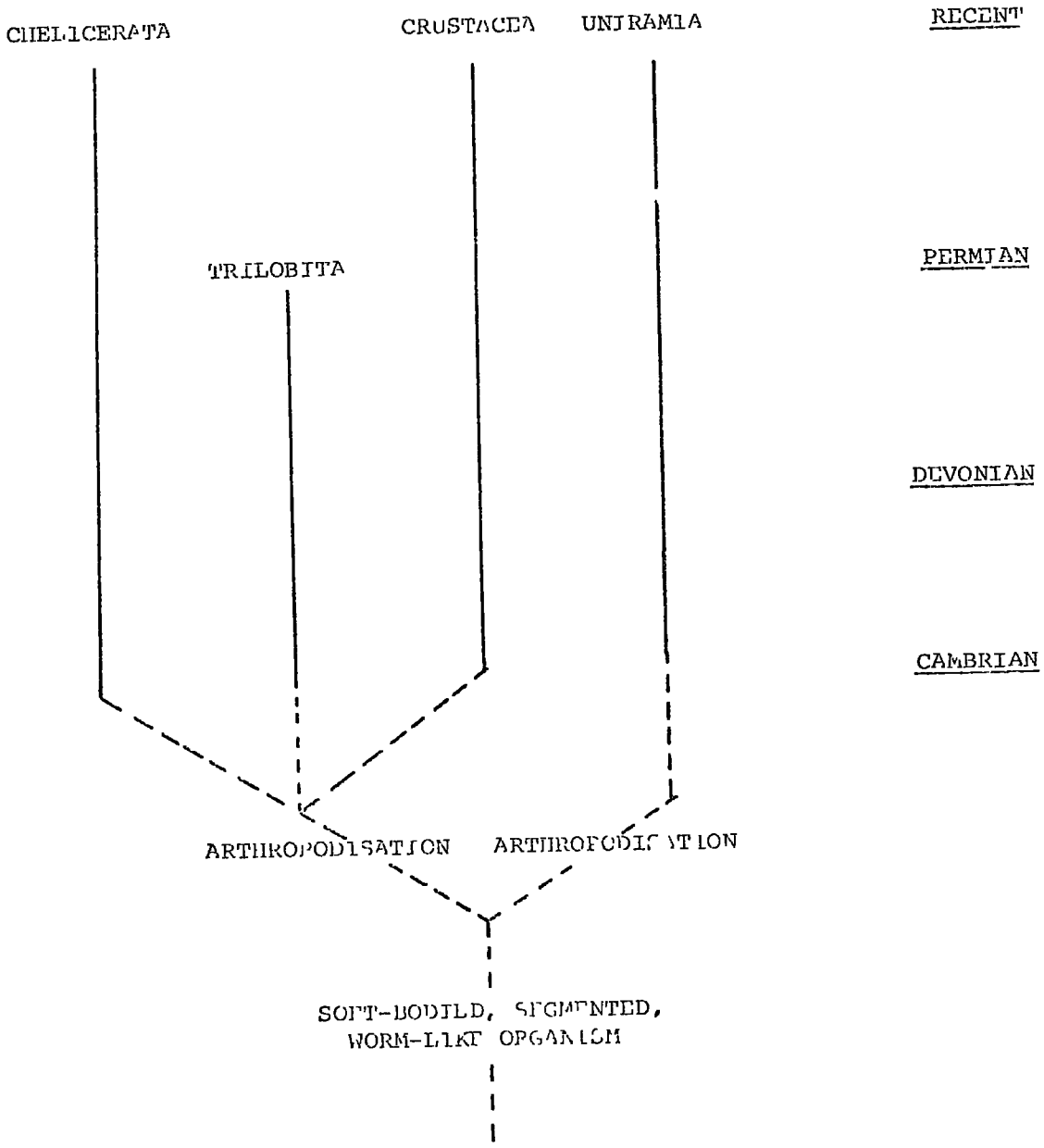

ARTHROPODA
ANNELIDA
MOLLUSCA
VERTEBRATA
UROCHORDATA
ECHINODERMATA
ASCHELMINTHES
BRACHIOPODA
HEMICHORDATA
PLATYHELMINTHES
ARCHIGASTRULA

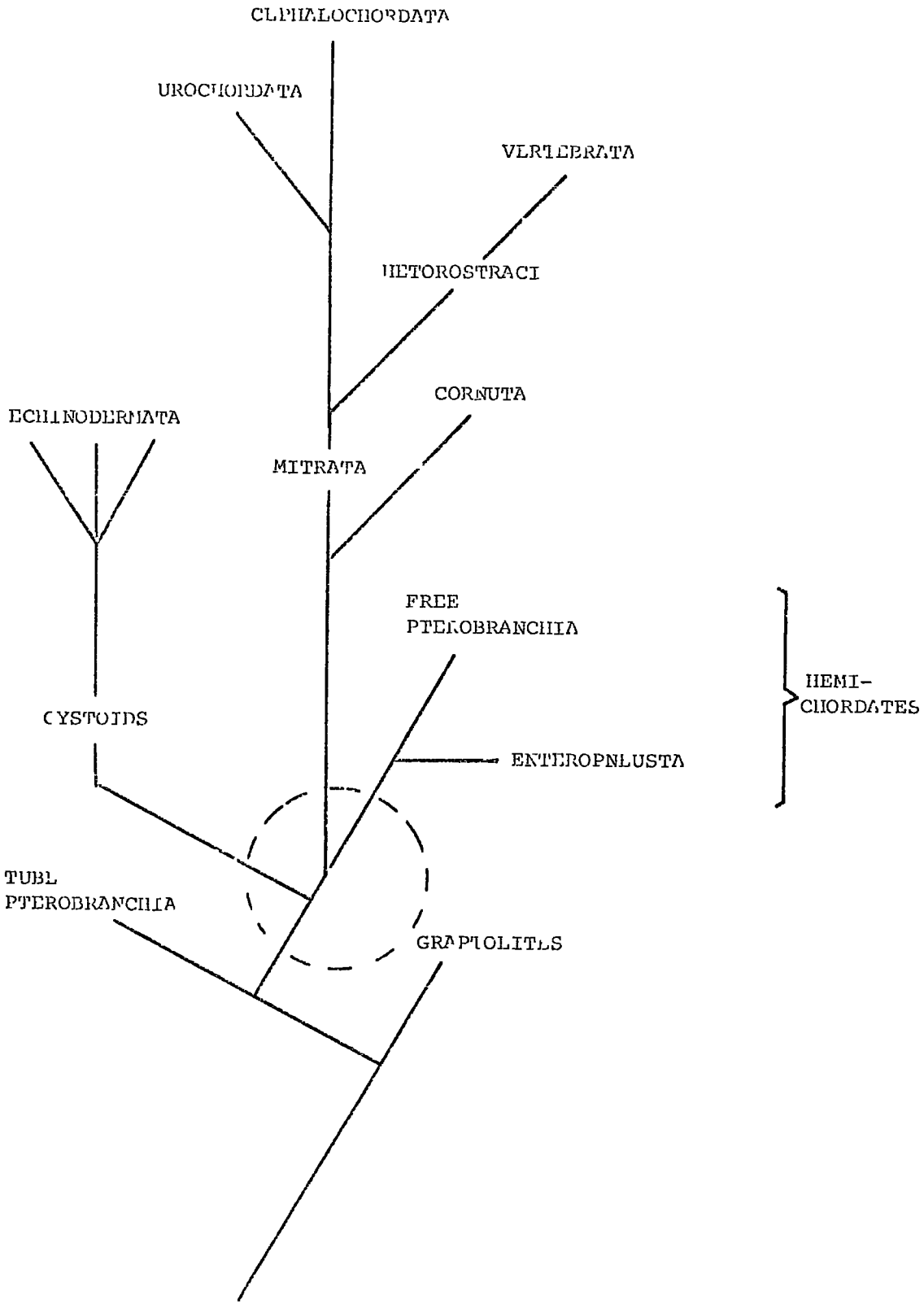

VERTEBRATA
UROCHORDATA
HEMICHORDATA
ECHINODERMATA
BRACHIOPODA
ARTHROPODA
ANNELIDA
MOLLUSCA
NEMERTINI
PLATYHELMINTHES
ACOELA

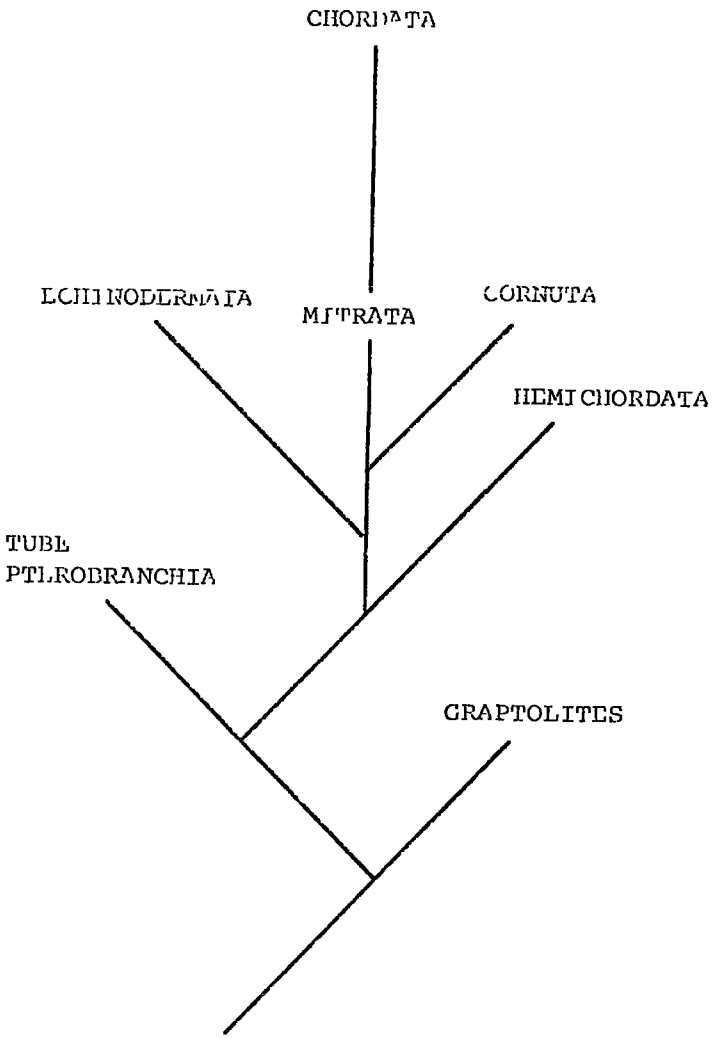

MOLLUSCA
INSECTA
CRUSTACLA
MYRIA-
PODA
CHELICERATA
VERTEBRATA
ANNELIDA
UROCHORDATA
?
?
ASCELMINTHES
BRACHIOPODA
?
ECHINODERMATA
?
?
CNIDARIA
PLATYHELMINTHES
?
PRIMITIVE ACOELOID
FLATWORM
PORIFERA


VERTEBRATA
HEMICHORDATA
ARTHROPODA
ANNELIDA
MOLLUSCA
ECHINODERMATA
DEUTEROSTOMIA
PROTOSTOMIA
CHAETOGNATHA
PLATYHELMINTHES
ASCHELMINTHES
ACOELA


UNIRAMIA
TRILOBITA
CHELICERATA
CRUSTACEA
ONYCHOPHORA
MYRIAPODA
HEXAPODA
BASIC
ARTHROPOD STOCK


CHELICERATA
CRUSTACEA
UNIRAMIA
RECENT
TRILOBITA
PERMIAN
DEVONIAN
CAMBRIAN
ARTHROPODISATION
ARTHROPODISATION
SOFT-BODIED, SEGMENTED,
WORM-LIKE ORGANISM


CEPHALOCHORDATA
UROCHORDATA
VERTEBRATA
HETOROSTRACI
CORNUTA
ECHINODERMATA
MITRATA
FREE
PTEROBRANCHIA
HEMI-
CHORDATES
CYSTOIDS
ENTEROPNEUSTA
TUBE
PTEROBRANCHIA
GRAPTOLITES


CHORDATA
ECHINODERMATA
MITRATA
CORNUTA
HEMICHORDATA
TUBE
PTEROBRANCHIA
GRAPTOLITES