The Amino Acid Sequence of *Phaseolus aureus* L. (Mung-Bean) Cytochrome *c*

By E. W. THOMPSON, M. V. LAYCOCK,* J. A. M. RAMSHAW AND D. BOULTER
Department of Botany, University of Durham, Durham City, U.K.

(Received 27 October 1969)

The amino acid sequence of *Phaseolus aureus* L. (mung-bean) cytochrome *c* has been determined. The molecule consists of a single polypeptide chain of 111 amino acid residues and is homologous with other mitochondrial cytochromes *c*. Comparison with the amino acid sequence of wheat-germ cytochrome *c* (Stevens, Glazer & Smith, 1967) shows 14 differences. On alignment with mammalian cytochromes *c*, mung-bean cytochrome *c* has an *N*-acetylated 'tail' of eight amino acid residues similar to that found in wheat-germ cytochrome *c*. Of the 22 positions in wheat-germ cytochrome *c* that contain amino acid residues unique to these positions, 20 were found to contain the same ones in mung-bean cytochrome *c*. The ϵ -*N*-trimethyl-lysine residues reported for wheat-germ cytochrome *c* (Delange, Glazer & Smith, 1969) in positions 72 and 86 were also found in these positions in mung-bean cytochrome *c*. The sequence was determined from 3 μ mol, by using chymotryptic and tryptic peptides which were analysed by the 'dansyl'-Edman method (Gray & Hartley, 1963a), with confirmation by amino acid analysis.

The amino acid sequences of a large number of cytochromes *c*, mainly from vertebrate sources, have now been determined (Dayhoff & Eck, 1968), but so far wheat-germ cytochrome *c* is the only higher-plant one that has been studied (Stevens, Glazer & Smith, 1967). Using these sequences, Fitch & Margoliash (1967) constructed a phylogenetic tree of the species concerned, which is in close agreement with that derived from fossil, morphological and cytological evidence.

The aim of the present investigation was to determine the amino acid sequence of cytochrome *c* from a second higher-plant source, to find whether many differences exist between the cytochromes *c* of two plant species and to see if their sequences have any characteristics in common that can be considered typical of plant cytochromes *c*. Since the concentration of cytochrome *c* in plant material is low compared with that in many animals, it was also of interest to see if the investigation could be completed with relatively small quantities of protein so that the potential range of possible future sources might be increased.

MATERIALS

The mung-bean cytochrome *c* was isolated from mung-bean seeds obtained locally. Trypsin (salt-free, twice recrystallized) and chymotrypsin (three times recrystallized)

were obtained from Worthington Biochemical Corp., Freehold, N.J., U.S.A. Carboxypeptidase A (di-isopropyl phosphorofluoridate-treated; crystalline suspension in water) was obtained from the Sigma Chemical Co., St Louis, Mo., U.S.A. Other chemicals used were of the highest purity readily available. Polyamide sheets were obtained from the Cheng Chin Trading Co. Ltd., Taipei, Taiwan.

METHODS

Preparation of cytochrome *c*. The isolation and purification of the cytochrome *c* from mung beans were carried out as described by Richardson, Laycock, Ramshaw, Thompson & Boulter (1970). The preparations were examined by polyacrylamide-gel electrophoresis, by isoelectric focussing and by spectral ratios, and were found to be pure. The ratio E_{550} (reduced)/ E_{280} (oxidized) was 0.91 for the material used to determine the sequence.

Denaturation of cytochrome *c*. Cytochrome *c* was denatured with ethanol by using the method of Margoliash, Kimmel, Hill & Schmidt (1962). Addition of 5M-(NH₄)₂-SO₄ was not required, and the precipitate was not dissolved in aq. ammonia or dialysed.

Digestion with chymotrypsin. Denatured cytochrome *c* (1.25 μ mol) was suspended in 3 ml of water and adjusted to pH 8.0 with 25mM-NaOH on a Radiometer TTT1C autotitrator. The digestion was performed at 40°C by the addition of 2% (w/w, enzyme/substrate) α -chymotrypsin made up in water at zero time and a further 1% after 90 min. The digestion was followed on the autotitrator and was terminated after 120 min by adjusting the clarified solution to pH 4.5 with acetic acid. The digest was allowed to stand overnight at this pH at 4°C. The dark-brown precipitate that formed was centrifuged from the almost

* Present address: Atlantic Regional Laboratory, Halifax, N.S., Canada.

colourless solution containing the soluble peptides, which was then freeze-dried.

Digestion with trypsin. Denatured cytochrome *c* (1.25 μ mol) was suspended in 3 ml of water and adjusted to pH 8.0 with 25 mM-NaOH on a recording Radiometer TTT1C autotitrator. The digestion was performed at 35°C by the addition of 1% (w/w, enzyme/substrate) trypsin made up in 1 mM-HCl at zero time and a further 1% after 60 min. The digestion was followed on the autotitrator and was terminated after 100 min by adjusting the clarified solution to pH 4.5 with acetic acid. The digest was then treated as described above for the chymotryptic digest.

Purification of soluble peptides. The soluble chymotryptic and tryptic peptides were separated by high-voltage paper electrophoresis at pH 6.5 (pyridine-acetic acid-water, 25:1:225, by vol.) on Whatman 3 MM paper, in a flat-plate apparatus (107 cm $\times$ 15 cm; The Locarte Co., London S.W.3, U.K.), at 90 V/cm for 120–150 min. The peptides were located on guide strips by the cadmium-ninhydrin reagent of Heilman, Barollier & Watzke (1957) and the starch-iodide method of Pan & Dutcher (1956). The Ehrlich, Sakaguchi and platinic iodide reagents recommended by Easley (1965) were also used. Peptides were eluted with water and their purity was determined by *N*-terminal analysis by the 'dansyl' method (Gray & Hartley, 1963b). Peptides requiring further separation and those obtained from further digestions of long tryptic peptides were purified by electrophoresis at pH 1.9 (acetic acid-formic acid-water, 4:1:45, by vol.) in the flat-plate apparatus.

Separation of haem peptides. Part of the haem peptides was obtained from the dark-brown precipitates described above. However, some haem peptide was not precipitated on standing at pH 4.5. This was not eluted by water, but could be eluted by 20% (v/v) pyridine (C. J. Bailey, personal communication) and was combined with the insoluble material. The haem group was removed by the mercuric chloride method of Ambler (1963) and the resulting peptides were separated by electrophoresis at pH 1.9.

Electrophoretic mobilities of peptides. These were measured from DNS-arginine relative to DNS-Arg-Arg at pH 6.5 and from 1-dimethylaminonaphthalene-5-sulphonic acid relative to DNS-arginine at pH 1.9.

Haem-free cytochrome *c* for amino acid analysis. The haem was removed from the total protein by the silver sulphate method of Margoliash *et al.* (1962).

Amino acid-sequence methods. The 'dansyl'-Edman method of Gray & Hartley (1963a) was used. The *N*-terminal amino acids of the peptides were determined by the 'dansyl' procedure (Gray & Hartley, 1963b). A sample of peptide containing 10–50 nmol was taken and dried *in vacuo* over NaOH and conc. H₂SO₄. The peptide was dissolved in 10 μ l of 0.1 M-NaHCO₃ and again dried *in vacuo* over NaOH and conc. H₂SO₄. Then 5 μ l of water and 5 μ l of a solution containing 2.5 mg of DNS chloride/ml in acetone were added, and the tubes were sealed with Parafilm. After 1 h at 37°C the reaction was stopped by drying *in vacuo* over NaOH and conc. H₂SO₄. Then 40 μ l of 6 M-HCl was added and the 'dansyl' peptide hydrolysed in a sealed tube for 6–8 h at 110°C. The hydrolysate was dried *in vacuo* over NaOH and the 'dansyl' amino acids were separated by chromatography on polyamide sheets (Woods & Wang, 1967). DNS-amino acids

were spotted by using aq. 95% (v/v) ethanol followed by 1 M-ammonia. Development in the first dimension was by aq. 90% (v/v) formic acid-water (3:200, v/v) for 35 min and in the second dimension by benzene-acetic acid (9:1, v/v) for 40 min. This was followed by butyl acetate-glacial acetic acid-methanol (20:1:1, by vol.) for 40 min in the second dimension to resolve DNS-serine from DNS-threonine and DNS-aspartic acid from DNS-glutamic acid. Differentiation between asparaginyl and aspartyl residues and glutaminyl and glutamyl residues was achieved by examining peptide electrophoretic mobilities by the method of Offord (1966).

Peptides were degraded by the Edman procedure (Edman, 1956). The peptide (0.1–0.5 μ mol) was taken up in 150 μ l of distilled water, and 150 μ l of 5% (v/v) re-distilled phenyl isothiocyanate in pyridine was added. The tube was flushed with N₂, sealed with Parafilm and heated at 45°C for 1 h. The excess of reagent was then removed by drying *in vacuo* over NaOH and P₂O₅ at 60°C. When it was completely dry, 200 μ l of anhydrous trifluoroacetic acid was added, and the tube was resealed with Parafilm and heated at 45°C for 30 min. Excess of reagent was removed by drying *in vacuo* at 60°C over NaOH. The degraded peptide was dissolved in 200 μ l of water and extracted three times with 1.5 ml of butyl acetate, and finally dried *in vacuo* over conc. H₂SO₄ and NaOH. A suitable sample was removed for *N*-terminal analysis, and the rest was subjected to further degradation.

Amino acid composition. Quantitative analyses were made on haem-free protein and chymotryptic and tryptic peptides by using a Technicon automatic amino acid analyser with a 133 cm column of Chromobead type A resin. The nine-chamber Autograd 18 h elution system was used and the recorder had a scale expander fitted. Total protein (1–2 mg) or peptide (0.05–0.5 μ mol) was hydrolysed at 110°C in evacuated Pyrex tubes with constant-boiling 5.7 M-HCl (Moore & Stein, 1963). For peptides containing or suspected to contain tryptophan, 5% (v/v) thioglycolic acid was added with the HCl (Matsubara & Sasaki, 1969). A semi-quantitative method was also used to show the composition of peptides. A sample containing 10–50 nmol of peptide was dried *in vacuo* over NaOH and conc. H₂SO₄. It was then hydrolysed in a sealed tube by 25 μ l of 5.7 M-HCl for 24 h at 110°C. The amino acids liberated were dried *in vacuo* over NaOH for 16 h and then 'dansylated' by the procedure described above but without further hydrolysis. After the chromatographic separation of the DNS-amino acids the number of residues of each amino acid originally present in the peptide was estimated. An equal fluorescent yield for each amino acid was assumed (Gros & Labouesse, 1969).

Chymotryptic digestion of tryptic peptides. The peptides were digested with 5% (w/w) α -chymotrypsin in 0.2 M-NH₄HCO₃ buffer, pH 8.4, at 37°C for 60 min.

Carboxypeptidase A digestion of peptides. The peptides were digested with 0.2–0.5 mg of carboxypeptidase A/ μ mol in 0.2 M-NH₄HCO₃ buffer, pH 8.4, at 37°C. The amino acids liberated were detected by the 'dansyl' method.

Preparation of ϵ -N-trimethyl-L-lysine. ϵ -N-Trimethyl-L-lysine was prepared by the method of Delange *et al.* (1969) but was separated by paper chromatography with the solvent system of Stewart (1963). The *R_F* values and

electrophoretic mobility of the material agreed with those quoted by Delange *et al.* (1969), and it gave a single ninhydrin-positive peak on the analyser, eluted 43 min before lysine.

Nomenclature of peptides. Arabic numerals preceded by the letter C refer to purified chymotryptic peptides, and those preceded by the letter T refer to purified tryptic peptides, both numbered on the basis of their occurrence in the sequence starting at the *N*-terminus. Minor peptides derived by further cleavage are given by a suffix to the larger peptide.

RESULTS

Amino acid analysis of cytochrome c

Samples of haem-free cytochrome *c* were hydrolysed for 24 h, and the average results for each amino acid are given in Table 1. The values were in reasonable agreement with those calculated from the complete sequence. A basic amino acid that was completely separated from the other basic amino acids present co-chromatographed with a synthetic preparation of ϵ -*N*-trimethyl-lysine and was therefore accepted as that amino acid.

Peptides from chymotryptic digestion

Mung-bean cytochrome *c*, unlike horse heart cytochrome *c*, precipitated immediately upon the addition of ethanol in the denaturation method. The denatured cytochrome *c* was readily digested by α -chymotrypsin, the reaction being essentially

complete after 2 h, leaving no insoluble core material. After overnight standing at 4°C and pH 4.5, some haem material precipitated, and this was removed by centrifugation. The soluble peptides were split into two equal parts and the peptides in each half were separated by high-voltage paper electrophoresis at pH 6.5. The results of the two electrophoreses were similar. The peptides were purified where necessary by further electrophoresis at pH 1.9. The sequences of peptides from the first half of the digest were obtained by the 'dansyl'-Edman method, and the peptides from the second half that were recovered in sufficient yield were hydrolysed and their amino acid compositions determined on the analyser to support the results of the 'dansyl'-Edman method. These analyses were corrected for small amounts of background amino acids always eluted from the paper with the peptides. The following chymotryptic peptides were examined and the results from the 'dansyl'-Edman method together with their electrophoretic mobilities are summarized in Table 2. The amino acid compositions of the peptides from total analysis are given in Table 3. The relative positions of each peptide after electrophoresis at pH 6.5 are shown schematically in Fig. 1(a). The numbers in parentheses after the peptide name are the residue numbers in the whole protein (see Fig. 2).

Peptide C1 (−8 → −6). This was detected only by the starch-iodide method between peptides C12 and C12A, and purified as the only peptide neutral at pH 1.9, suggesting that the *N*-terminal amino group was blocked. Carboxypeptidase A digestion and 'dansylation' showed phenylalanine and small amounts of serine after 3 h and equal amounts of phenylalanine and serine after 24 h. The amino acid composition confirmed the peptide sequence acetyl-Ala-Ser-Phe.

Peptide C2 (−5 → 7). This was an acidic peptide found in very low yield, which was platonic iodide-positive (methionine), with alanine as the *N*-terminus. There was insufficient material for an accurate amino acid analysis, and information on this part of the sequence was obtained from tryptic peptide T1.

Peptide C3 (8 → 10). This was pure after pH 6.5 electrophoresis but was obtained in poor yield. The 'dansyl'-Edman method gave the sequence Lys-Ile-Phe and this was supported by the amino acid composition obtained by 'dansylation' after hydrolysis.

Peptide C4 (11 → 26) (*a haem peptide*). Part of the sequence was determined by the 'dansyl'-Edman method. The amino acid analysis gave a very low threonine value, possibly due to oxidation by performic acid during removal of the haem. Owing to the presence of histidine and cysteic acid residues, the method of Offord (1966) could not be used to obtain information about amide content of the

Table 1. *Amino acid composition of haem-free mung-bean cytochrome c*

	Amino acid composition (residues/mol)		
	Analysis values	Integral values	Sequence values
Asp	9.8	10	11
Thr	7.6	8	8
Ser	6.3	7	6
Glu	8.6	9	8
Pro	6.5	7	7
Gly	12.1	12	11
Ala	9.9	10	11
Val	3.2	3	3
Cys	1.2	1	2
Met	2.6	3	3
Ile	3.4	3	4
Leu	8.0	8	8
Tyr	5.9	6	6
Phe	4.3	4	4
Tml	1.9	2	2
Lys	11.5	12	12
His	2.2	2	2
Arg	2.3	2	2
Trp	*	*	1

* Tryptophan content was not determined.

Table 2. 'Dansyl'-Edman procedure applied to chymotryptic peptides

Peptide no.	Mobility at pH 6.5 relative to DNS-Arg-Arg	Mobility at pH 1.9 relative to DNS-arginine	Sequence results
C1	-1.50	0	Does not react
C2	-0.65		<u>Ala-</u>
C3	1.20		<u>Lys-Ile-Phe</u>
C4	0.70	0.84	<u>Lys-Thr-Lys-CySO₃-Ala-Glx-CySO₃-His-</u>
C5	0.81	0.74	<u>Lys-Glx-Gly-Pro-Asx-Leu-Asx</u>
C6	0	0.58	<u>Gly-Leu-Phe</u>
C7	0.70	0.61	<u>Gly-Arg-Glx-Ser-Gly-Thr-Thr-Ala-Gly-Tyr</u>
C8	0	0.66	<u>Ser-Tyr</u>
C9	0.87	0.70	<u>Ser-</u>
C10	0	0.51	<u>Ala-Val-Ile-</u>
C11	-0.91		<u>Glx-Glx-Lys-Thr-Leu-Tyr</u>
C12	-1.65	0.50	<u>Asx-Tyr</u>
C12A	-1.37	0.47	<u>Asx-Tyr-Leu</u>
C13	1.53		<u>Leu-Leu-Asx-Pro-(Tml)-Lys-Tyr</u>
C13A	1.68		<u>Leu-Asx-Pro-(Tml)-Lys-Tyr</u>
C14	1.02		<u>Ile-Pro-Gly-Thr-Lys-Met</u>
C15	0	0.47	<u>Val-Phe-Pro-Gly-Leu</u>
C16	0.70	1.01	<u>(Tml)-Lys-Pro-Glx-Asx-(Arg)-Ala-Asx-Leu</u>
C17	0	0.55	<u>Ile-Ala-Tyr</u>
C18	0	0.82	<u>Leu-Lys-Glx-Ser-Thr-Ala</u>

peptide. Tryptic peptides T5 and T6 confirmed the sequence of performic acid-oxidized peptide C4 as Lys-Thr-Lys-CySO₃*-Ala-Glx-CySO₃-His (Thr, Val, Asx, Lys, Gly, Ala, Gly, His), and this sequence was supported by the amino acid composition with the exception of the low threonine value.

Peptide C5 (27 → 33). This gave a weak positive reaction with the Ehrlich reagent, but no tryptophan could be found in the peptide. The 'dansyl'-Edman method gave the sequence Lys-Glx-Gly-Pro-Asx-Leu-Asx, which was supported by the amino acid composition. The mobility of 0.81 at pH 6.5 showed a charge of 1 for a peptide of this molecular weight, which indicates that residues 28, 31 and 33 are all amides.

Peptide C6 (34 → 36). This was neutral at pH 6.5 and was separated by electrophoresis at pH 1.9. A yellow colour with cadmium-ninhydrin reagent was observed and the sequence Gly-Leu-Phe was given by the 'dansyl'-Edman method. This was supported by the amino acid composition.

Peptide C7 (37 → 46). This was yellow with cadmium-ninhydrin reagent and was Sakaguchi positive (arginine). As the amino acid composition gave no lysine in the peptide, DNS-arginine was observed with no masking by ϵ -DNS-lysine, which

* Abbreviations: CySO₃, cysteic acid; Tml, ϵ -N-trimethyl-lysine.

co-chromatographed with DNS-arginine in the solvents used. The 'dansyl'-Edman method gave the sequence Gly-Arg-Glx-Ser-Gly-Thr-Thr-Ala-Gly-Tyr, which was supported by the amino acid composition. A charge of 1 was indicated by the mobility of 0.70 at pH 6.5 for a peptide of this molecular weight, so residue 39 was glutamine.

Peptide C8 (47 → 48). This was neutral at pH 6.5 and was purified at pH 1.9. The 'dansyl'-Edman method gave the sequence Ser-Tyr, which was supported by the amino acid composition.

Peptide C9 (49 → 55). This was platinum iodide-positive (methionine) and was only partially separated from peptide C5, which had almost identical mobilities at both pH 6.5 and pH 1.9 and occurred in much greater yield. N-Terminal serine was found. Tryptic peptides T8C3 (Table 6) and T9 (Table 5) suggested that this peptide was Ser(Thr, Asn, Ala, Lys, Asn, Met).

Peptide C10. This was neutral at pH 6.5. It was Ehrlich-positive (tryptophan) and was not obtained in sufficient yield to allow an accurate amino acid analysis. Carboxypeptidase A digestion followed by 'dansylation' revealed tryptophan as the C-terminus and the 'dansyl'-Edman method gave the sequence Ala-Val-Ile.

Peptide C11 (60 → 65). The 'dansyl'-Edman method gave the sequence Glx-Glx-Lys-Thr-Leu-

Table 3. *Amino acid analysis of chymotryptic peptides*

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

Peptide no.	Composition (molar ratios)					
C1	Ser	1.0 (1),	Ala	1.0 (1),	Phe	0.9 (1)
C2	Not analysed					
C3*	Ile	(1),	Phe	(1),	Lys	(1)
C4	CySO ₃	2.5 (2),	Asp	1.3 (1),	Thr	0.3 (2),
	Ala	1.6 (2),	Val	0.8 (1),	Lys	2.9 (3),
					His	1.8 (2)
C5	Asp	2.2 (2),	Glu	1.0 (1),	Pro	0.8 (1),
	Lys	1.1 (1)			Gly	1.2 (1),
					Leu	0.9 (1),
C6	Gly	1.1 (1),	Leu	1.0 (1),	Phe	0.8 (1)
C7	Thr	1.6 (2),	Ser	0.8 (1),	Glu	0.9 (1),
	Tyr	0.9 (1),	Arg	1.0 (1)	Gly	3.3 (3),
					Ala	0.8 (1),
C8	Ser	1.0 (1),	Tyr	0.8 (1)		
C9	Not analysed					
C10	Not analysed					
C11	Thr	0.8 (1),	Glu	1.8 (2),	Leu	1.0 (1),
					Tyr	0.7 (1),
					Lys	1.0 (1)
C12	Asp	1.0 (1),	Tyr	0.7 (1)		
C12A	Asp	1.3 (1),	Leu	0.8 (1),	Tyr	0.7 (1)
C13	Not analysed (see C13A)					
C13A	Asp	1.0 (1),	Pro	0.6 (1),	Leu	1.1 (1),
	Lys	1.3 (1)			Tyr	0.5 (1),
					Tml	0.9 (1),
C14	Thr	0.8 (1),	Pro	0.7 (1),	Gly	1.1 (1),
	Lys	1.0 (1)			Met	0.5 (1),
					Ile	0.6 (1),
C15	Pro	0.8 (1),	Gly	1.0 (1),	Val	1.0 (1),
					Leu	1.3 (1),
					Phe	(1)
C16	Asp	2.0 (2),	Glu	1.2 (1),	Pro	0.8 (1),
	Tml	0.8 (1),	Lys	1.1 (1),	Ala	0.7 (1),
					Leu	1.1 (1),
					Arg	1.1 (1)
C17	Ala	0.9 (1),	Ile	0.6 (1),	Tyr	0.3 (1)
C18	Thr	0.9 (1),	Ser	0.8 (1),	Glu	1.0 (1),
	Lys	1.1 (1)			Ala	1.0 (1),
					Leu	1.0 (1),

* Composition shown by hydrolysis plus 'dansylation'.

Tyr, which was supported by the amino acid composition. The mobility at pH 6.5 was -0.91 which for a peptide of this molecular weight indicates a charge of -1 . This means that both residues 60 and 61 were glutamic acid.

Peptides C12 and C12A (66 $\rightarrow$ 67 and 66 $\rightarrow$ 68). The 'dansyl'-Edman method gave Asx-Tyr and Asx-Tyr-Leu for the sequences of these peptides, which were confirmed in both cases by the amino acid composition. Position 66 was aspartic acid because both peptides were acidic at pH 6.5.

Peptides C13 and C13A (68 $\rightarrow$ 74 and 69 $\rightarrow$ 74). These both contained the basic amino acid accepted as ϵ -*N*-trimethyl-lysine. The 'dansyl'-Edman method gave the sequences Leu-Leu-Asx-Pro-(Tml)-Lys-Tyr and Leu-Asx-Pro-(Tml)-Lys-Tyr. ϵ -*N*-Trimethyl-lysine was placed in position 72 because no other amino acid was found in this position. DNS-Tml co-chromatographs with ϵ -DNS-lysine and was not observed, as this peptide contains lysine. The amino acid composition supported these sequences. The mobilities of 1.53 and 1.68 at pH 6.5 were equivalent to a charge of 2 for peptides of these molecular weights, so residue 70 was asparagine.

Peptide C14 (75 $\rightarrow$ 80). This was platinum iodide-

positive (methionine). The sequence Ile-Pro-Gly-Thr-Lys-Met was given by the 'dansyl'-Edman method and was supported by the amino acid composition.

Peptide C15 (81 $\rightarrow$ 85). This was neutral at pH 6.5 and was purified at pH 1.9. The sequence Val-Phe-Pro-Gly-Leu was given by the 'dansyl'-Edman method and was supported by the amino acid composition.

Peptide C16 (86 $\rightarrow$ 94). This was Sakaguchi positive (arginine). The *N*-terminus gave a strong ϵ -DNS-lysine, DNS-Tml or DNS-arginine, but no other amino acid, and amino acid analysis showed that the peptide contained lysine, ϵ -*N*-trimethyl-lysine and arginine. Lysine was shown in position 87 by di-DNS-lysine, but arginine and ϵ -*N*-trimethyl-lysine could only be placed by homology with wheat-germ cytochrome *c*, which has ϵ -*N*-trimethyl-lysine in position 86 and invariant arginine in position 91. The 'dansyl'-Edman method gave the sequence Tml-Lys-Pro-Glx-Asx-Arg-Ala-Asx-Leu and this was supported by the amino acid composition and homology with wheat. The mobility at pH 6.5, 0.70, indicated that a peptide of this molecular weight had a charge of 1, so two of the positions 89, 90 and 93

Fig. 1. Peptides from (a) chymotryptic and (b) tryptic digests of mung-bean cytochrome c after electrophoresis at pH 6.5. DNS-OH, 1-dimethylaminonaphthalene-5-sulphonic acid.

were amides, but which two cannot be decided from these results.

Peptide C17 (95 → 97). This was a neutral peptide at pH 6.5 and was obtained after electrophoresis at pH 1.9. The 'dansyl'-Edman method gave the sequence Ile-Ala-Tyr, which was supported by the amino acid composition.

Peptide C18 (98 → 103). This was a neutral peptide at pH 6.5 and was purified at pH 1.9. The 'dansyl'-Edman method gave the sequence Leu-Lys-Glx-Ser-Thr-Ala, which was supported by the amino acid composition. As the peptide was neutral at pH 6.5, position 100 was glutamic acid. The observed *C*-terminal amino acid, alanine, is not normally susceptible to chymotryptic cleavage, indicating that this is the *C*-terminal peptide for the protein.

Peptides from tryptic digestion

Peptides obtained after tryptic digestion were separated by the methods described for the chymotryptic peptides. A summary of the results of the 'dansyl'-Edman method together with the electro-

phoresis mobilities are shown in Table 4 and amino acid compositions in Table 5. The positions of the major tryptic peptides after pH 6.5 electrophoresis are shown schematically in Fig. 1(b). The following tryptic peptides were examined.

Peptide T1 (−8 → 5). This was weakly ninhydrin-positive and gave no *N*-terminus by the 'dansyl' method. The peptide was digested with α -chymotrypsin, which gave two resulting peptides which it was considered unnecessary to separate since peptide T1C1 did not react in the 'dansyl'-Edman method. Results from peptide T1 are shown in Table 6. The mobility at pH 6.5, −1.24, showed a charge of −1 for a peptide of this molecular weight, indicating that one amide group was present, but this could not be placed as there were three acidic residues in the peptide. The 'dansyl'-Edman method gave the sequence Ala-Glx-Ala-Pro-Pro-Gly-Asx-Ser-Asx-Lys for peptide T1C2. This was supported by the amino acid composition of peptide T1.

Peptide T2 (6 → 8). This was platinum iodide-positive (methionine). The 'dansyl'-Edman method gave the sequence Gly-Met-Lys, which was supported by the amino acid composition.

Peptide T3 (9 → 11). The sequence Ile-Phe-Lys was given by the 'dansyl'-Edman method and supported by the amino acid composition.

Peptide T4 (12 → 13). This was Thr-Lys by the 'dansyl'-Edman method.

Peptide T5 (14 → 22) (a haem peptide). This was recovered in low yield after removal of the haem prosthetic group. The sequence of the whole peptide was determined by the 'dansyl'-Edman method, and hydrolysis followed by 'dansylation' indicated the presence of only one lysine and one histidine residue. The electrophoretic mobilities could not be used, because of the cysteic acid and histidine residues, for deduction of the amide content. The sequence was found to be $\text{CySO}_3\text{-Ala-Glx-CySO}_3\text{-His-Thr-Val-Asx-Lys}$ by the 'dansyl'-Edman method.

Peptide T6 (23 → 27). The 'dansyl'-Edman method gave the sequence Gly-Ala-Gly-His-Lys, which was supported by the amino acid composition.

Peptide T7 (28 → 38). Five Edman degradations gave the sequence Glx-Gly-Pro-Asx-Leu-Asx(Gly-Leu, Phe, Gly, Arg), which was supported by the amino acid composition. The last five residues were placed by peptides C6 and C7. The mobility at pH 6.5 (0.50–0.65) suggested a charge of 1 for a peptide of this molecular weight, so both positions 28 and 31 were amides.

Peptide T8 (39 → 53). This was a large peptide. It was digested with α -chymotrypsin. This gave three peptides, which were separated at pH 1.9, the results being shown in Table 6. These results, together with evidence from the sequences of chymo-

Table 4. '*Dansyl*'-Edman procedure applied to tryptic peptides

Peptide no.	Mobility at pH 6.5 relative to DNS-Arg-Arg	Mobility at pH 1.9 relative to DNS-arginine	Sequence results
T1	-1.24		Does not react; digested with chymotrypsin (see Table 6).
T2	1.61		<u>Gly-Met-Lys</u>
T3	1.38		<u>Ile-Phe-Lys</u>
T4	1.75	1.45	<u>Thr-Lys</u>
T5	0	0.28	<u>CySO₃-Ala-Glx-CySO₃-His-Thr-Val-Asx-Lys</u>
T6	2.20		<u>Gly-Ala-Gly-His-Lys</u>
T7	*	0.47	<u>Glx-Gly-Pro-Asx-Leu-Asx</u>
T8	*	0.39	Digested with chymotrypsin (see Table 6).
T9	-0.55		<u>Asx-Met-Ala-Val-Ile-</u>
T10	*	0.66	<u>Thr-Leu-Tyr-Asp-</u>
T11	1.00		<u>Tyr-Ile-Pro-Gly-Thr-Lys</u>
T12	1.15		<u>Met-</u>
T13	0	0.56	<u>Ala-Asp-Leu-Ile-Ala-Tyr-Leu-Lys</u>
T14	-1.45		<u>Glx-Ser-Thr-Ala</u>

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

Table 5. *Amino acid analysis of tryptic peptides*

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

Peptide no.	Composition (molar ratios)					
T1	Asp 2.2 (2), Ala 2.4 (3),	Ser 1.6 (2), Phe 1.1 (1),	Glu 1.2 (1), Lys 1.1 (1)	Pro 2.2 (2),	Gly 1.2 (1),	
T2	Gly 1.0 (1),	Met 0.4 (1),	Lys 1.2 (1)			
T3	Ile 0.7 (1),	Phe 1.0 (1),	Lys 1.0 (1)			
T4	Thr 1.0 (1),	Lys 1.0 (1)				
T5	CySO ₃ (2), Val (1),	Asp (1), Lys (1),	Thr (1), His (1)	Glu (1),	Ala (1),	
T6	Gly 1.9 (2),	Ala 0.9 (1),	His 1.0 (1),	Lys 1.0 (1)		
T7	Asp 1.9 (2), Phe 1.0 (1),	Glu 0.8 (1), Arg 1.0 (1)	Pro 0.8 (1),	Gly 3.1 (3),	Leu 2.0 (2),	
T8	See Table 6					
T9	Asp 1.0 (1), Ile 0.8 (1),	Glu 2.2 (2), Lys 1.3 (1)	Ala 0.8 (1),	Val 0.9 (1),	Met 1.0 (1),	
T10	Asp 2.1 (2), Tml 1.1 (1),	Thr 0.8 (1), Lys 1.1 (1)	Pro 1.0 (1),	Leu 3.3 (3),	Tyr 0.5 (2),	
T11	Thr 0.8 (1), Lys 1.1 (1)	Pro 0.7 (1),	Gly 1.2 (1),	Ile 0.7 (1),	Tyr 0.6 (1),	
T12	Asp 1.0 (1), Met 0.5 (1), Arg 1.3 (1)	Glu 1.0 (1), Leu 1.0 (1)	Pro 1.7 (2), Phe 1.0 (1)	Gly 1.5 (1), Tml 0.6 (1),	Val 0.8 (1), Lys 0.9 (1),	
T13	Asp 1.0 (1), Lys 1.1 (1)	Ala 1.8 (2),	Ile 0.7 (1),	Leu 1.9 (2),	Tyr 0.8 (1),	
T14	Thr 0.9 (1),	Ser 0.9 (1),	Glu 1.0 (1),	Ala 0.9 (1)		

* Composition shown by hydrolysis plus 'dansylation'.

tryptic peptides, gave the sequence Glx-Ser-Gly-Thr-Thr-Ala-Ala-Gly-Tyr-Ser-Tyr-Ser-Thr-Ala-Asx-Lys. The mobility at pH 6.5 (0.50–0.65) indicates a charge of 1 for a peptide of this molecular weight, so positions 39 and 52 were both amides.

Peptide T9 (54 → 62). This was strongly Ehrlich-positive (tryptophan) and was also pink in the trifluoroacetic acid stage of the first Edman degradation (Uphaus, Grossweiner, Katz & Kopple, 1959). The peptide was also platonic iodide-positive

Table 6. *Results from peptides T1 and T8 after chymotryptic digestion*

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

(a) 'Dansyl'-Edman results		
Peptide no.	Mobility at pH 1.9	Sequence results
T1C1		Does not react with DNS-Cl
T1C2		<u>Ala-Glx-Ala-Pro-Pro-Gly-Asx-Ser-Asx-Lys</u>
T8C1	0.30	<u>Glx-</u> ; not run further, but taken for amino acid analysis
T8C2	0.66	<u>Ser-Tyr</u>
T8C3	0.92	<u>Ser-Thr-Ala-Asx-Lys</u>
(b) Amino acid analysis results		
Peptide no.	Composition (molar ratios)	
T8C1	Thr 1.8 (2), Ser 0.8 (1), Glu 1.0 (1), Gly 2.1 (2), Ala 0.9 (1), Tyr 0.6 (1)	
T8C2*	Ser (1), Tyr (1)	
T8C3*	Ala (1), Asp (1), Lys (1), Ser (1), Thr (1)	

* Composition found by 'dansylation' of the hydrolysate

(methionine). The sequence Asx-Met-Ala-Val-Ile-(Trp,Glu,Glu,Lys) was given in part by the 'dansyl'-Edman method. The last four residues were placed by peptide C11. The mobility at pH 6.5 indicates a charge of -1 and, as positions 60 and 61 were both glutamic acid (see peptide C11), position 54 was asparagine.

Peptide T10 (63 → 73). Amino acid analysis and three Edman degradations, together with chymotryptic peptides C12 and C13, gave the sequence Thr-Leu-Tyr-Asx (Tyr,Leu,Leu,Asn,Pro,Tml,Lys) for this peptide. No tryptic break occurred between ϵ -*N*-trimethyl-lysine and lysine because of the proline residue adjacent to the former. The mobility at pH 6.5 (0.50-0.65) suggested a charge of 1 for a peptide of this molecular weight, indicating one amide and one acid residue. The former was placed in position 70 by peptide C13 and the latter in position 66 by peptide C12.

Peptide T11 (74 → 79). The sequence Tyr-Ile-Pro-Gly-Thr-Lys was given by the 'dansyl'-Edman method and was supported by amino acid analysis.

Peptide T12 (80 → 91). This was both Sakaguchi (arginine) and platinic iodide (methionine) positive, and the *N*-terminus was methionine. Amino acid composition and chymotryptic overlaps (C15-C16) suggested that the sequence of this peptide was Met(Val,Phe,Pro,Gly,Leu,Tml,Lys,Pro,Glx,Asx,Arg). Again no break was found between ϵ -*N*-trimethyl-lysine and lysine because of the adjacent proline. As with peptide C16, the mobility at pH 6.5 indicated one acidic and two amide residues but these could not be placed.

Peptide T13 (T13A and T13B) (92 → 99). Peptide T13 was neutral at pH 6.5. Peptides T13A and T13B were the result of chymotryptic activity

causing a split at tyrosine (97) to give two peptides, the sequences of which were Ala-Asp-Leu-Ile-Ala-Tyr and Leu-Lys by the 'dansyl'-Edman method (supported by amino acid compositions). This was the only evidence of any chymotryptic activity in the tryptic digest. Position 93 is aspartic acid as peptide T13 was neutral at pH 6.5.

Peptide T14 (100 → 103) was acidic, the mobility proving glutamic acid in position 100. No ϵ -DNS-lysine was observed in the 'dansyl' method, so this is considered to be the *C*-terminal peptide in the protein as it was not Sakaguchi-positive (arginine). The 'dansyl'-Edman method gave the sequence Glu-Ser-Thr-Ala, which was confirmed by the amino acid composition.

Allocation of amide groups

Together both digests give the full sequence with many overlaps. Homology with other cytochromes *c* was used only to decide between Tml-86-arginine-91 and arginine-86-Tml-91. Ten amides were suggested from all peptide mobilities and of these 8 can be unambiguously placed by the peptide mobilities. These are glutamine-28, asparagine-31, glutamine-39, asparagine-52, asparagine-54 and asparagine-70. The analyses of peptides T1 and C2 show three acidic amino acids and the charge indicates only one of these to be an amide. By homology with wheat-germ cytochrome *c* the amide is placed in position 2, leaving positions -4 and 3 as the acidic residues. Similarly in peptides C16 and T12, homology with all cytochromes places residue 89 as glutamine and residue 90 as aspartic acid. The remaining uncertain positions are 16 and 21 in the haem peptides. As described above the haem peptides both contain cysteic acid and histidine so their mobilities cannot be used for determining amide

content. Electrophoresis of a chymotryptic digest of horse heart cytochrome *c* at pH 6.5 showed that the horse heart haem peptide had a similar mobility to the chymotryptic haem peptide of mung-bean cytochrome *c*. As the charge and cysteic acid and histidine contents are the same in both haem peptides (Margoliash, Smith, Kreil & Tuppy, 1961), it was concluded that in common with horse heart this peptide had one acidic and one amide residue. Homology with all cytochromes placed glutamine in position 16 and aspartic acid in position 21. The nature of the *N*-terminal block is assumed to be an acetyl group by homology with wheat-germ cytochrome *c* and the majority of the other cytochromes *c* so far examined. Both chymotrypsin and trypsin left no insoluble core after digestion, although some haem peptide precipitated on standing at pH 4.5 at 4°C overnight. Trypsin cleaved specifically at the *C*-terminal side of lysine and arginine except when these residues were adjacent to proline. No evidence for cleavage at ϵ -*N*-trimethyl-lysine was found because of adjacent proline residues. The cleavage of the tyrosine-leucine bond at 97–98 was the only evidence of chymotryptic activity in the tryptic digest. Chymotrypsin was found to cleave at the *C*-terminal side of three phenylalanine residues. The fourth (82) was adjacent to proline and was not cleaved. Cleavage occurred at the six tyrosine and three methionine residues, but at only three of the eight leucine residues. The tryptophan-glutamic acid bond at 59–60 was also broken. This did not occur in wheat-germ cytochrome *c*, probably because residues 58, 60 and 61 were all acidic, whereas mung-bean cytochrome *c* had isoleucine in position 58. In common with wheat-germ cytochrome *c*, cleavage occurred at the histidine in position 26 but not at that in position 18, possibly because of the close proximity of the latter to the haem group. The only other chymotryptic cleavage was at asparagine in position 33. This was also found in kangaroo cytochrome *c*, which also has asparagine in this position (Nolan & Margoliash, 1966).

DISCUSSION

Mung-bean cytochrome *c* has been shown to be a mitochondrial type cytochrome *c* consisting of a single chain of 111 amino acids, with a total molecular weight of 12 904 including the prosthetic group. Determination of the sequence of mung-bean cytochrome *c* has not changed the number of amino acids that are invariant in all cytochromes *c* studied to date, but methionine residues in positions 7 and 55 and lysine in position 62 have not previously been found in these positions.

The sequence of mung-bean cytochrome *c* resembles that of wheat-germ cytochrome *c* in the

Fig. 2. Amino acid sequence of mung-bean cytochrome *c*. Chymotryptic and tryptic peptides are indicated.

following points, all of which suggest a common higher plant ancestor.

Both cytochromes *c* have an *N*-acetylated 'tail' of eight amino acids before glycine in position 1, starting acetyl-Ala-Ser-Phe- (see Fig. 2). Twenty-two positions in wheat-germ cytochrome *c* had an amino acid not previously found in that position in any other cytochrome *c* studied, and mung-bean cytochrome *c* has the same amino acid as wheat-germ cytochrome *c* in 20 of these positions (see Fig. 3). The high proline content observed in wheat-germ cytochrome *c* (eight residues) compared with other cytochromes *c* (e.g. four in horse heart), is also observed in mung bean (seven residues). Both proteins have two residues of ϵ -*N* trimethyl-lysine, in positions 72 and 86. The presence of this unusual

	-8	-7	-6	-5	-4	-3	-2	-1	
	Ala	Ser	Phe	<u>Ala</u>	—	Ala	—	Pro	
1	2	3	4	5					10
—	—	<u>Ser</u>	Asp	<u>Lys</u>	—	<u>Met</u>	—	—	—
—	—	—	—	15	—	—	—	—	20
—	—	—	—	25	—	—	—	—	30
Asp	<u>Lys</u>	—	Ala	Gly	—	—	Gln	—	—
—	—	—	—	35	—	—	—	Gln	40
—	—	<u>Asn</u>	—	—	—	—	—	—	—
—	Thr	Thr	—	45	—	—	—	—	50
—	—	—	—	55	—	—	—	—	60
—	—	—	—	<u>Met</u>	Ala	—	<u>Ile</u>	—	Glu
—	—	—	—	65	—	—	—	—	70
—	<u>Lys</u>	—	—	Tyr	—	—	—	Leu	—
—	—	—	—	75	—	—	—	—	80
—	—	—	—	85	—	—	—	—	90
—	—	Pro	—	—	Tml	—	—	Gln	—
—	—	—	—	95	—	—	—	—	100
—	—	—	—	—	—	—	—	—	<u>Glu</u>
101	102	103	104						
<u>Ser</u>	—	<u>Ala</u>	***						

Fig. 3. Comparison of wheat and mung-bean cytochromes *c*. The sequence partially shown is that of mung-bean cytochrome *c*. The underlined positions are those which have a different amino acid residue in wheat-germ cytochrome *c*. The others shown are those amino acids which have been found in these positions only in mung-bean and wheat-germ cytochromes *c*. ***, Residue (serine) found in wheat-germ but absent from mung-bean cytochrome *c*.

amino acid in position 72 is of particular interest, as in all 'mammalian' cytochromes *c* residues 70–80 have been invariant, and, as this is the longest invariant region in the molecule, it is probably functionally significant. It is possible that further unusual amino acids may be unequivocally identified in other plant cytochromes as a result of sequence studies.

Although these similarities between mung-bean and wheat-germ cytochromes *c* strongly suggest a higher-plant common ancestor, 14 differences have been found between the two primary structures. These 14 differences are equivalent to a total minimum mutation distance of 18, including a deletion in position 104, which has been considered as a change from serine to a chain terminator.

The results show that the sequence of a higher-plant cytochrome *c* can be obtained with 3 μ mol by using high-voltage paper electrophoresis to separate the peptides for analysis by the 'dansyl'–Edman

method. The yield of pure cytochrome *c* from higher plants is only 1–4 mg/kg, so by using the subtractive Edman method, the number of readily available sources from which sufficient material can be obtained to study the sequence of the protein is limited. The methods we have used should considerably widen the scope of possible sources.

The determination of this sequence is part of an investigation into the molecular evolution of plant cytochromes *c* (Boulter *et al.* 1970). The 14 differences found between the monocotyledon wheat and the dicotyledon mung bean strongly suggest that ordination of the major plant orders can be achieved by using cytochrome *c* sequence determination.

We thank Dr B. S. Hartley, Medical Research Council Laboratory of Molecular Biology, Cambridge, U.K., for much helpful advice, and Dr M. E. Boulter for help with the 'dansyl'–Edman analysis. We thank the Nuffield Foundation and the Science Research Council for financial support.

REFERENCES

- Ambler, R. P. (1963). *Biochem. J.* **89**, 341.
 Boulter, D., Laycock, M. V., Ramshaw, J. A. M. & Thompson, E. W. (1970). In *Phytochemical Phylogeny*, p. 80. Ed. by Harborne, J. B. London and New York: Academic Press Inc.
 Dayhoff, M. O. & Eck, R. V. (1968). *Atlas of Protein Sequence and Structure*. Silver Spring, Md.: National Biomedical Research Foundation.
 Delange, R. J., Glazer, A. N. & Smith, E. L. (1969). *J. biol. Chem.* **244**, 1385.
 Easley, C. W. (1965). *Biochim. biophys. Acta*, **107**, 386.
 Edman, P. (1956). *Acta chem. scand.* **10**, 761.
 Fitch, W. M. & Margoliash, E. (1967). *Science, N.Y.*, **155**, 279.
 Gray, W. R. & Hartley, B. S. (1963a). *Biochem. J.* **89**, 379.
 Gray, W. R. & Hartley, B. S. (1963b). *Biochem. J.* **89**, 59p.
 Gros, C. & Labouesse, B. (1969). *Eur. J. Biochem.* **7**, 463.
 Heilman, J., Barollier, J. & Watzke, E. (1957). *Hoppe-Seyler's Z. physiol. Chem.* **309**, 219.
 Margoliash, E., Kimmel, J. R. Hill, R. L. & Schmidt, W. R. (1962). *J. biol. Chem.* **237**, 2148.
 Margoliash, E., Smith, E. L., Kreil, G. & Tuppy, H. (1961). *Nature, Lond.*, **192**, 1121.
 Matsubara, H. & Sasaki, R. M. (1969). *Biochem. biophys. Res. Commun.* **35**, 175.
 Moore, S. & Stein, W. H. (1963). In *Methods in Enzymology*, vol. 6, p. 819. Ed. by Colowick, S. P. & Kaplan, N. O. New York: Academic Press Inc.
 Nolan, C. & Margoliash, E. (1966). *J. biol. Chem.* **241**, 1049.
 Offord, R. E. (1966). *Nature, Lond.*, **211**, 591.
 Pan, S. C. & Dutcher, J. D. (1956). *Analyt. Chem.* **28**, 836.
 Richardson, M., Laycock, M. V., Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1970). *Phytochemistry* (in the Press).
 Stevens, F. C., Glazer, A. N. & Smith, E. L. (1967). *J. biol. Chem.* **242**, 2764.
 Stewart, I. (1963). *J. Chromat.* **10**, 404.
 Uphaus, R. A., Grossweiner, L. I., Katz, J. J. & Kopple, K. D. (1959). *Science, N.Y.*, **129**, 641.
 Woods, K. R. & Wang, K. T. (1967). *Biochim. biophys. Acta*, **133**, 369.