The Amino Acid Sequence of Sesame (*Sesamum indicum* L.) and Castor (*Ricinus communis* L.) Cytochrome *c*

By E. W. THOMPSON, M. RICHARDSON AND D. BOULTER

Department of Botany, University of Durham, Durham City, U.K.

(Received 6 October 1970)

The amino acid sequences of sesame (*Sesamum indicum* L.) and castor (*Ricinus communis* L.) cytochrome *c* were determined by using 1.5 μ mol of protein from each species. Both molecules consist of a single chain of 111 amino acid residues and are homologous with other mitochondrial cytochrome *c* molecules. Both have an *N*-acetylated 'tail' of eight amino acids and two ϵ -*N*-trimethyl-lysine residues, as also reported for wheat germ (Delange, Glazer & Smith, 1969) and mung-bean cytochrome *c* (Thompson, Laycock, Ramshaw & Boulter, 1970). Two different preparations of castor cytochrome *c* differed by one residue. This was glutamic acid for glutamine in position 100. The results for sesame and castor cytochrome *c* led to a re-examination and subsequent correction to the *N*-terminal region of the mung-bean cytochrome *c* sequence, as given by Thompson *et al.* (1970).

The determination of the amino acid sequences of sesame and castor cytochrome *c* forms part of a comparative investigation, the eventual aim of which is to construct a phylogenetic tree for plants by using the established methods of Fitch & Margoliash (1967) and Dayhoff & Eck (1968).

MATERIALS

Sesame seeds were obtained from Frank Fehr and Co. Ltd., London E.C. 3, U.K., and the incorporated Oil Seed Corporation, 24 St Mary Axe, London E.C.3, U.K. Castor oil seeds were a generous gift from British Oil and Cake Mills, Hull, Yorks., U.K. Other materials were as described by Thompson, Laycock, Ramshaw & Boulter (1970).

METHODS

The isolation and purification of cytochrome *c* from sesame and castor is described by Richardson, Laycock, Ramshaw, Thompson & Boulter (1970). The E_{550} (reduced)/ E_{280} (oxidized) ratio was 1.22 and 1.00 for castor and sesame cytochrome *c* respectively.

The methods used for the determination of the amino acid sequences of sesame and castor cytochrome *c* were as described by Thompson *et al.* (1970), with the following exceptions. (1) After digestion of denatured cytochrome *c* with either trypsin or chymotrypsin, the digest was applied directly to paper for electrophoresis. (2) For removal of the haem group the haem peptides of castor were taken up in 200 μ l of 90% (v/v) formic acid, 10 μ l of 30% (v/v) H_2O_2 was added and the mixture shaken to oxidize the thioether link (Nolan & Margoliash, 1966). After 10 min a further 10 μ l was added. The mixture was freeze-dried after 20 min, by which time most of the haem colour had disappeared. The haem group was removed from haem

peptides of sesame, as described previously (Thompson *et al.* 1970). (3) The haem group was not removed before determination of the amino acid composition of the total protein. (4) The conditions of Stevens, Glazer & Smith (1967) were used for papain digestions. (5) The third-dimension solvent, butyl acetate-methanol-acetic acid (30:20:1, by vol.) (Ramshaw, Thompson & Boulter, 1970), was used in the polyamide-layer chromatography of DNS-amino acids. (6) *N*-Terminal blocking groups were determined by the hydrazinolysis method of Schmer & Kreil (1969), modified as described by Ramshaw *et al.* (1970).

Nomenclature of peptides. Sesame and castor cytochrome *c* peptides are distinguished by the prefixes S and C respectively. Arabic numerals prefixed by the letter C refer to purified chymotryptic peptides and those prefixed 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 of long peptides are denoted by a subscript to the larger peptide.

RESULTS

Sequences of sesame and castor cytochrome *c*. Amino acid compositions of sesame and castor cytochrome *c* determined by analysis and calculated from the complete sequence are shown in Table 1. Two residues of ϵ -*N*-trimethyl-lysine were found in both species, as in wheat germ and mung-bean cytochrome *c*. The analysis for sesame cytochrome *c* is in close agreement with the composition calculated from the sequence, with the exception of the value for glutamic acid. The castor cytochrome is high in serine and low in glycine and alanine in comparison with the sequence values. Castor cytochrome also

Table 1. *Amino acid composition of sesame and castor cytochrome c*

Abbreviations: nd, not determined; TML, trimethyl-lysine.

	Sesame		Castor	
	Amino acid analysis (integral)	Sequence	Amino acid analysis (integral)	Sequence
Asp	10	10	11	11
Thr	7	7	7	7
Ser	6	6	5	4
Glu	11	9	10	10
Pro	8	8	7	7
Gly	12	12	11	12
Ala	11	11	12	13
Val	4	4	4	4
Met	2	2	2	2
Cystine	nd	2	nd	2
Ile	4	4	3	3
Leu	8	8	8	8
Tyr	5	6	4	6
Phe	4	4	4	4
TML	2	2	2	2
Lys	11	11	11	11
His	2	2	2	2
Arg	2	2	2	2
Trp	nd	1	nd	1
		111		111

had an E_{440} peak in the position of hydroxyproline. (Hydroxyproline was not found during the dansyl-Edman analysis and does not wholly occupy any position in the sequence. Total hydrolysis followed by dansylation indicated traces of this amino acid in certain peptide zones of the electrophoretogram. The possibility of limited hydroxylation of certain proline residues is being investigated.) The results of the dansyl-Edman analysis on the chymotryptic and tryptic peptides of sesame and castor cytochrome *c* are shown in Tables 2 and 3. The amino acid compositions of peptides are shown in Tables 4 and 5 and the results from peptides subjected to further digestion are shown in Table 6. The complete castor and sesame cytochrome *c* sequences are shown in Fig. 1, except that in both sequences two aspartyl residues and one glutamyl residue are left as Asx or Glx, as discussed below.

Application of the dansyl-Edman method to the majority of peptides gave unambiguous sequences. Exceptions were the peptides containing residue -5 in sesame and residue 104 in both sesame and castor cytochrome *c*. The *N*-terminus of peptide SC2, (residues -5 to 10) by the dansyl-Edman method gave serine and aspartic acid in the ratio 1:1. The amino acid analysis of peptide ST1 (residues -8 to 4), the tryptic equivalent, gave lower serine and higher aspartic acid values than would be expected, if the peptide had serine as *N*-terminus. In a further tryptic digest of sesame cytochrome *c*, the blocked

peptide ST1 was isolated by electrophoresis at pH 6.5, and was purified as usual by electrophoresis at pH 2 and had a mobility of 35% relative to DNS-arginine. The peptide gave no *N*-terminus by the dansyl method, but ϵ -DNS-lysine was present in high yield, which confirmed that the peptide had a blocked *N*-terminus. The peptide was digested with chymotrypsin to cleave phenylalanine, position -6, and electrophoresis was performed on the digest at pH 2. This gave a single band of mobility of 65% consistent with the removal of three residues and an increase of one positive charge due to the new unblocked *N*-terminus. The *N*-terminus of this peptide was again aspartic acid and serine in a ratio of 1:1, indicating the presence of two peptides identical except for residue -5. However, in view of the amino acid composition of peptide ST1 (Table 4), the *N*-terminus has been taken to be an aspartyl residue.

The other ambiguity concerned a probable deletion in position 104 in both proteins. The relevant peptides were SC15, ST13 and CC14, CT14. Dansyl-Edman analysis was inconclusive because of incomplete Edman reaction of the free amino acid remaining after removal of the rest of the peptide. Amino acid analysis of peptides SC15 and CC14 supported the sequence Leu-Lys-Glx-Ala-Thr-Ala for both peptides, whereas analysis of the tryptic peptides supported the sequence Glx-Ala-Thr-Ala-Ala. The mobilities at pH 6.5 of these peptides,

Table 2. Dansyl-Edman analysis of chymotryptic and tryptic peptides of sesame cytochrome *c*

Amides and their parent acid residues established from peptide mobilities are shown, but those placed later by homology left as Asx or Glx. Mobilities at pH 6.5 are relative to DNS-Arg-Arg- and at pH 1.9 are relative to DNS-Arg. Arrows show the positions confirmed by dansyl-Edman analysis. Abbreviations: Ac, acetyl; CySO₃, cysteine acid; TML, trimethyl-lysine.

		Mobility				Mobility	
		pH 6.5	pH 1.9			pH 6.5	pH 1.9
	Chymotryptic peptides						
SC1	Blocked N-terminus (Ac, Ala, Ser, Phe) (see the text)			ST1	Tryptic peptides Blocked N-terminus. Amino acid analysis: (Ac, Ala, Ser, Phe, Asx, Glx, Ala, Pro, Pro, Gly, Asx, Val, Lys)		
SC2(a)	Asx-Glx-Ala-Pro-Pro-Gly-Asx-Val(Lys)	-0.8		ST2	Ser-Gly-Glu-Lys	0	1.06
SC2(b)	Asx(Glx, Ala, Pro, Pro, Gly, Asx, Val, Lys, Ser, Gly, Glu, Lys, Ile, Phe). Papain digest (Table 6)	-0.5	0.59	ST3	Ile-Phe-Lys	1.35	1.04
SC2(c)	Ser-Gly-Glu(Lys, Ile, Phe)	0		ST4	Thr-Lys	1.8	
SC3	Haem peptide. Lys(Thr, Lys, Cys, Ala, Glx, Cys, His, Thr, Val, Asx, Gly, Ala, Gly, His)	0.5	0.72	ST5	Haem peptide (CySO ₃)Ala-Glx(CySO ₃)(His)Thr-Val-Asx(Lys)	0	0.25
SC4	Lys-Gln-Gly-Pro-Asn-Leu-Asn	0.9		ST6	Gly-Ala-Gly-His-Lys	2.3	
SC5	Gly-Leu-Phe	0	0.54	ST7(a)	Gln-Gly(Pro, Asn, Leu, Asn, Gly, Leu, Phe, Gly, Arg)	0.7	
SC6	Gly-Arg-Gln-Ser-Gly-Thr-Thr-Pro-Gly-Tyr	0.6		ST7(b)	Gln-Gly-Pro-Asn-Leu-Asn	0	0.35
SC7	Ser-Tyr	0	0.59	ST7(c)	Gly-Leu-Phe-Gly-Arg	1.2	0.83
SC8(a)	Ser-Ala-Ala-Asn(Lys, Asn, Met)	1.1		ST8	Gln-Ser(Gly, Thr, Thr, Pro, Gly, Tyr, Ser, Tyr, Ser, Ala, Ala, Asn, Lys). Digested with chymo- trypsin (Table 6)	0.5	0.27
SC8(b)	Lys-Asn-Met	1.4		ST9(a)	Asn-Met-Ala-Val-Ile(Trp, Gly, Glx, Asx, Thr, Leu, Tyr, Asp, Tyr, Leu, Leu, Asn, Pro, TML, Lys)	0	
SC9	Ala-Val-Ile(Trp)	0	0.34	ST9(b)	(Asp, Tyr, Leu, Leu, Asn, Pro, TML, Lys)	1.0	
SC10	Gly-Glx-Asx-Thr-Leu-Thr	-1.0	0.28	ST10	Tyr-Ile-Pro-Gly-Thr-Lys	1.2	0.86
SC11(a)	Asp-Tyr-Leu-Leu-Asn(Pro, TML, Lys, Tyr)	0.6		ST11	Met(Val, Phe, Pro, Gly, Leu, TML, Lys, Pro, Glx, Glx, Arg)		
SC11(b)	Asp-Tyr-Leu	-1.3		ST12(a)	Ala-Asp-Leu-Ile-Ala-Tyr(Lys)	0	0.56
SC11(c)	Asp-Tyr	-1.6	0.52	ST12(b)	Ala-Asp-Leu-Ile-Ala-Tyr	-1.0	
SC11(d)	Leu-Leu-Asn(Pro, TML, Lys, Tyr)	1.7	0.92	ST12(c)	Leu-Lys		
SC11(e)	Leu-Asn-Pro-TML-Lys-Tyr	1.8	1.02	ST13	Glu-Ala-Thr-Ala	-1.5	
SC12	Ile-Pro-Gly-Thr-Lys-Met	1.1	0.82				
SC13	Val-Phe-Pro-Gly-Leu	0	0.39				
SC14(a)	(TML, Lys, Pro, Glx, Glx, Arg, Ala, Asp, Leu)	-1.7	1.30				
SC14(b)	TML(Lys, Pro, Glx, Glx, Arg)	-1.0	0.28				
SC14(c)	Ala-Asp-Leu-Ile-Ala-Tyr	0	0.47				
SC14(d)	Ile-Ala-Tyr	0	0.78				
SC15	Leu-Lys-Glu-Ala-Thr-Ala	0					

Table 3. *Dansyl-Edman analysis of chymotryptic and tryptic peptides of castor cytochrome c*

Amides and their parent acid residues established from peptide mobilities are shown, but those placed later by homology left as Asx or Glx. Mobilities at pH 6.5 are relative to DNS-Arg-Arg, and at pH 1.9 are relative to DNS-Arg. Arrows show those positions confirmed by dansyl-Edman analysis. Abbreviations: Ac, acetyl; CySO₃, cysteic acid; TML, trimethyl-lysine.

			Mobility			Mobility	
			pH 6.5	pH 1.9		pH 6.5	pH 1.9
CC1	Chymotryptic	Blocked N-terminus (Ac, Ala, Ser, Phe) (see the text)			Tryptic		
CC2(a)		Asx-Glx-Ala(Pro, Pro, Gly, Asx, Val, Lys, Ser, Gly, Glu, Lys, Ile, Phe). Papain digest (Table 6)	-0.5		Blocked N-terminus. Analysis only: (Ac, Ala, Ser, Phe, Asx, Glx, Ala, Pro, Gly, Asx, Val, Lys)	0	1.05
CC2(b)		Asx-Glx-Ala-Pro(Pro, Gly, Asx, Val, Lys)	-0.8		Ala-Gly-Glu-Lys	1.4	
CC2(c)		Ala-Gly-Glu(Lys, Ile, Phe)	0	0.75	Ile-Phe-Lys	1.8	
CC3		Haem peptide. Lys(Thr, Lys, Cys, Ala, Glx, Cys, His, Thr, Val, Glx, Gly, Ala, Gly, His)	0.5		Thr-Lys	0	0.25
CC4(a)		Lys-Gly-Gln-Pro, Asn-Leu-Asn-Gly(Lys, Phe)	0.6	0.54	(CySO ₃)-Ala-Glx-(CySO ₃)-(His)-Thr-Val-Glx(Lys)	2.4	
CC4(b)		(Lys, Gly, Gln, Pro, Asn, Leu, Asn)	0	0.56	Gly-Ala-Gly(His, Lys)	0.6	
CC4(c)		Gly-Leu-Phe	0	0.58	Gln-Gly(Pro, Asn, Leu, Asn, Gly, Leu, Phe, Gly, Arg)	0	0.3
CC5		Gly-Arg-Gln-Ser-Gly-Thr-Thr-Ala-Gly-Tyr	0.6	0.59	Gln(Ser, Gly, Thr, Thr, Ala, Gly, Tyr, Ser, Tyr)	1.1	0.93
CC6		Ser-Tyr	0	0.58	Ser-Ala-Ala-Asn(Lys)	0	0.27
CC7(a)		Ser(Ala, Ala, Asn, Lys, Asn, Met)	1.1	0.86	Asn-Met-Ala(Val, Gln, Trp, Gly, Glx, Asn, Thr, Leu, Tyr, Asp, Tyr, Leu, Leu, Asn, Pro, TML, Lys)	0.5	0.19
CC7(b)		Lys-Asn-Met	1.4		Asn-Met-Ala(Val, Gln, Trp, Gly, Glx, Asn, Thr, Leu, Tyr)	0.7	0.81
CC8(a)		Met-Ala-Val-Gln(Trp)	0	0.30	Asp(Tyr, Leu, Leu, Asn, Pro, TML, Lys)	0.9	
CC8(b)		Ala-Val-Gln(Trp)	0	0.35	Tyr-Ile-Pro-Gly-Thr-Lys	1.1	0.83
CC9		Gly-Glx-Asn-Thr-Leu-Tyr	-0.9		Met-Val-Phe(Pro, Gly, Leu, TML, Lys, Pro, Glx, Asx, Arg)	0	0.59
CC10(a)		Asp(Tyr, Leu, Leu, Asn, Pro, TML, Lys, Tyr)	0.6		Ala-Asp(Leu, Ile, Ala, Tyr, Leu, Lys)	1.1	
CC10(b)		Asp-Tyr-Leu	-1.3	0.45	Ala-Asp-Leu-Ile-Ala-Tyr	1.83	
CC10(c)		Asp-Tyr	-1.6	0.53	Leu-Lys	0	0.50
CC10(d)		Leu-Leu-Asn-Pro-TML-Lys-Tyr	1.5		Gln-Ala-Thr-Ala		
CC10(e)		Leu-Asn-Pro-TML-Lys-Tyr	1.7				
CC11		Ile-Pro-Gly-Thr-Lys-Met	1.1	0.82			
CC12		Val-Phe-Pro-Gly-Leu	0	0.40			
CC13(a)		TML-Lys-Pro-Glx-Asx-Arg(Ala, Asp, Leu, Ile, Ala, Tyr)	0.6	0.81			
CC13(b)		TML-Lys-Pro-Glx-Asx-Arg(Ala, Asp, Leu)	0.6	1.05			
CC13(c)		Ile-Ala-Tyr	0	0.50			
CC14		Leu-Lys-Gln-Ala-Thr-Ala	1.0	0.80			

however, indicated the absence of the second terminal alanine residue. This was resolved by performing three Edman stages on peptide ST13, dansylation of a portion of the remaining peptide being followed directly by polyamide-layer chromatography. DNS-alanine and not DNS-alanyl-alanine was the only product found. The additional alanine residue given by the amino acid analysis of peptide ST13 and peptide CT14 cannot be explained, but it was noted that the threonine value was relatively low in both cases.

Two different preparations of castor cytochrome *c* were used in the present investigation. Peptide CT14 from the first preparation was neutral at pH 6.5 and was Gln-Ala-Thr-Ala, whereas the peptide from the second preparation, like that from sesame cytochrome, was acidic and was Glu-Ala-Thr-Ala. Whether this difference was due to polymorphism or to deamidation is uncertain. Both glutamine and glutamic acid have been put in position 100 in Fig. 1.

Apart from these exceptions, unequivocal dansyl-Edman analyses were closely supported by amino acid compositions except for occasional low values for the labile amino acids tyrosine, serine and threonine, and unexplained low values for aspartic

acid in peptides CT1 (residues -8-4) and CT10(c) (residues 66-73).

The blocked peptides SC1 and CC1 yielded phenylalanine on short digestion with carboxypeptidase A and phenylalanine+serine on longer digestion. The hydrazide, formed by hydrazinolysis of both peptides, co-chromatographed with standard acetyl hydrazide.

Sesame and castor cytochrome *c* contained 19 and 20 aspartyl+glutamyl residues respectively. Peptide mobilities decided the acid or amide form of 11 of these. Six of the others occurred in three pairs in peptides SC3, CC3; SC10, CC9; and SC14, CC13, each of which had one acid and one amide, and these positions were decided by homology with other cytochrome *c* species. The remaining aspartyl and glutamyl residues occurred in peptides SC2 and CC2, and their mobilities required that two out of the three residues were acid. These residues are given as Asx and Glx in Fig. 1.

Most of the expected specificities of chymotrypsin and trypsin were demonstrated in this study. Chymotrypsin cleaved at asparagine residues 33, 52 and 54 in both proteins, and also at arginine (residue 91) in sesame cytochrome. Trypsin cleaved at tyrosine (residue 97) in both proteins and at

Fig. 1. Complete sequence of sesame, castor and mung-bean cytochrome *c*.

Table 4. *Amino acid composition*

Values are given as mol/mol of peptide; values in parentheses are the compositions finally deduced

Peptide ...	SC1	SC3	SC5	SC6	SC8(b)	SC9*	SC10	SC11(b)	SC11(e)	SC13	SC15
Asp		1.5 (1)			1.0 (1)		1.0 (1)	1.0 (1)	1.0 (1)		
Thr		1.0 (2)		1.6 (2)			0.8 (1)				1.1 (1)
Ser	0.4 (1)			0.9 (1)							
Glu		1.2 (1)		1.2 (1)			1.2 (1)				1.3 (1)
Pro				1.5 (1)					1.0 (1)	1.1 (1)	
Gly		2.0 (2)	1.0 (1)	3.0 (3)		1.0 (1)				1.1 (1)	
Ala	0.9 (1)	1.9 (2)				1.0 (1)					2.1
Val		1.0 (1)				0.8 (1)				1.0 (1)	
Met					0.8 (1)						
Ile						0.8 (1)					
Leu			1.0 (1)				0.9 (1)	1.0 (1)	1.0 (1)	1.0 (1)	0.8 (1)
Tyr				0.6 (1)			0.7 (1)	0.3 (1)	1.1 (1)		
Phe	1.0 (1)		1.0 (1)							0.9 (1)	
TML									1.0 (1)		
Lys		2.5 (3)			1.0 (1)				1.1 (1)		1.0 (1)
His		1.3 (2)									
Trp						0.8 (1)					
Arg				1.0 (1)							
CySO ₃		nd									

* Hydrolysed with 5%

Table 5. *Amino acid composition*

Values are given as mol/mol of peptide; values in parentheses are the compositions finally deduced

Peptide ...	CC1	CC2(a)	CC3	CC4(a)	CC5	CC9	CC10(b)	CC10(e)	CC11	CC12	CC14
Asp		1.3 (2)		2.3 (2)		1.2 (1)	1.5 (1)	1.0 (1)			
Thr			1.7 (2)		1.6 (2)	0.7 (1)			0.8 (1)		0.8 (1)
Ser	0.8 (1)			1.0 (1)	0.8 (1)						
Glu		2.0 (2)	2.0 (2)	1.0 (1)	1.0 (1)	1.1 (1)					1.0 (1)
Pro		2.0 (2)						1.0 (1)	1.1 (1)	1.0 (1)	
Gly		2.0 (2)	2.0 (2)	2.0 (2)	2.8 (3)	1.0 (1)			1.0 (1)	1.1 (1)	
Ala	1.0 (1)	2.0 (2)	2.0 (2)	1.0 (1)	1.0 (1)						2.1
Val		1.1 (1)	0.9 (1)							0.9 (1)	
Met									1.0 (1)		
Ile		0.9 (1)							1.0 (1)		
Leu				1.7 (2)		1.0 (1)	1.0 (1)	1.0 (1)		1.1 (1)	1.0 (1)
Tyr					0.7 (1)	0.3 (1)	0.7 (1)	0.2 (1)			
Phe	1.0 (1)	0.8 (1)		0.8 (1)						1.0 (1)	
TML								1.1 (1)			
Lys		1.7 (2)	2.9 (3)	0.8 (1)				1.1 (1)	1.2 (1)		1.2 (1)
His			2.0 (2)								
Trp											
Arg					1.0 (1)						
CySO ₃			1.9 (2)								

asparagine (residue 33) in sesame cytochrome, and at tyrosine (residues 48 and 68) in castor cytochrome. The break at residue 48 was found in wheat cytochrome *c* (Stevens *et al.* 1967), and that at residue 68 was found to a lesser extent in sesame cytochrome.

Corrections to the sequence of mung-bean cytochrome c. The successful use of papain digestion in determining the sequence of the castor and sesame peptides CC2 and SC2 led us to reinvestigate the equivalent region of mung-bean cytochrome *c* (Thompson *et al.* 1970).

Incomplete digestion by chymotrypsin at residue

—6 had resulted in low yields of peptides C2 (residues —5 to 10) and TIC2 (residues —5 to 4), so that the sequence could be determined only on small amounts of the latter. The lack of homology between the mung-bean sequence and those of castor and sesame in this region made us suspect the former.

We therefore repeated the chymotryptic digest of mung-bean cytochrome *c* and isolated peptide MC2 (residues —5 to 10) in good yield. Papain digests of this peptide gave fragments exactly analogous to those obtained from sesame and castor, with the

of sesame cytochrome c peptides

from the sequence. Abbreviations: nd, not determined; CySO₃, cysteic acid; TML, trimethyl-lysine.

Peptide ...	ST1	ST2	ST3	ST6	ST7(b)	ST7(c)	ST8	ST10	ST11	ST12(b)	ST13
Asp	1.5 (2)				1.7 (2)		0.9 (1)			1.1 (1)	
Thr							1.9 (2)	0.8 (1)			0.7 (1)
Ser	1.3 (1)	1.0 (1)					2.3 (3)				
Glu	1.2 (1)	1.0 (1)			0.7 (1)		1.1 (1)		1.8 (2)	0.6 (0)	1.0 (1)
Pro	1.9 (2)				1.2 (1)		1.4 (1)	1.1 (1)	1.6 (2)		
Gly	1.0 (1)	1.0 (1)		2.0 (2)	1.1 (1)	2.2 (2)	2.5 (2)	1.2 (1)	1.1 (1)	0.5 (0)	
Ala	2.0 (2)			1.0 (1)			2.0 (2)			1.7 (2)	3.0
Val	1.0 (1)								0.8 (1)		
Met									0.4 (1)		
Ile			1.1 (1)					1.0 (1)		0.9 (1)	
Leu					1.1 (1)	1.1 (1)			1.1 (1)	1.0 (1)	
Tyr							1.5 (2)	0.8 (1)		0.8 (1)	
Phe	1.0 (1)		1.0 (1)			1.2 (1)			1.2 (1)		
TML									0.8 (1)		
Lys	1.0 (1)	1.0 (1)	1.2 (1)	0.9 (1)			0.9 (1)		1.0 (1)		
His				1.0 (1)							
Trp											
Arg						1.0 (1)			1.0 (1)		
CySO ₃											

(v/v) thioglycolic acid.

of castor cytochrome c peptides

deduced from the sequence. Abbreviations: tr, trace; CySO₃, cysteic acid; TML, trimethyl-lysine.

Peptide ...	CT1	CT2	CT3	CT6	CT7	CT8	CT9	CT10(b)	CT10(c)	CT11	CT12	CT13(b)	CT14
Asp	1.3 (2)				2.0 (2)		1.0 (1)	1.8 (2)	1.0 (2)		1.1 (1)	1.4 (1)	
Thr						1.5 (2)		1.0 (1)		0.6 (1)			0.6 (1)
Ser	0.5 (1)					1.3 (2)	1.0 (1)			0.4 (0)			
Glu	1.2 (1)	1.0 (1)			1.0 (1)	1.0 (1)		1.8 (2)			1.0 (1)		1.1 (1)
Pro	2.4 (2)				1.0 (1)				1.0 (1)	1.0 (1)	2.0 (2)		
Gly	1.2 (1)	1.0 (1)		2.0 (2)	3.0 (3)	2.0 (2)		1.3 (1)		1.0 (1)	1.1 (1)	0.7 (0)	
Ala	2.0 (2)	1.0 (1)		1.0 (1)		1.0 (1)	1.4 (2)	0.8 (1)				2.0 (2)	2.7
Val	1.0 (1)							0.9 (1)			1.0 (1)		
Met								0.5 (1)			0.7 (1)		
Ile			1.0 (1)							0.9 (1)		0.9 (1)	
Leu					1.9 (2)			0.9 (1)	2.0 (2)		1.0 (1)	1.0 (1)	
Tyr						0.6 (2)		0.7 (1)	tr	0.7 (1)			
Phe	0.8 (1)		0.9 (1)		1.0 (1)						1.0 (1)		
TML									1.0 (1)		1.0 (1)		
Lys	1.0 (1)	1.0 (1)	1.2 (1)	1.0 (1)			1.0 (1)		1.2 (1)	1.2 (1)	0.9 (1)		
His				1.0 (1)									
Trp								tr					
Arg					1.0 (1)								
CySO ₃											0.8 (1)		

analyses and sequences shown in Table 6. Further, only traces of the peptide Gly-Met-Lys were found in a tryptic digest of the mung-bean cytochrome compared with high yields of Ser-Gly-Glu-Lys and Ile-Phe-Lys. We therefore consider the correct sequence of mung-bean cytochrome to be that shown in Fig. 1.

The errors of Thompson *et al.* (1970) can be ascribed to the low yields of peptide C2, insufficient for analysis. The evidence was derived entirely from peptide T1C2 (residues -5 to 4) where poor amino acid analysis, an incorrect *N*-terminal analysis and the presence of more DNS-aspartic acid than bis-DNS-lysine after eight steps of Edman degradation suggested the sequence Ala-Glx-Ala-Pro-Pro-Gly-

Asx-Ser-Asx-Lys. Tryptic fragments Gly-Met-Lys and Ser-Gly-Glu-Lys were found, and also a chymotryptic peptide, Lys-Ile-Phe, in low yield. The sequence of residues -5 to 9 was wrongly deduced, but we feel confident that the sequence shown in Fig. 1 is correct.

As with castor and sesame cytochrome *c*, amides cannot be placed unambiguously in three cases, but a possible explanation for the limited split at phenylalanine (-6) could be that positions -5 and -4 are both acids, leaving position 2 as asparagine. Thus no chymotryptic split was found at tryptophan 59 in the sequence Glu-Trp-Glu-Glu in wheat cytochrome *c* (Stevens *et al.* 1967). Sesame cytochrome *c* and castor cytochrome *c* may have only

Table 6. *Further digests of sesame and castor cytochrome c peptides*

Mobilities are relative to DNS-Arg. Arrows show those positions confirmed by dansyl-Edman analysis.

Chymotrypsin digestion		Mobility at pH 1.9
Peptide		
ST8	<u>Gln-Ser</u> (Gly,Thr,Thr,Pro,Gly,Tyr,Ser,Tyr,Ser,Ala,Ala,Asn,Lys)	
ST8C1	} Not analysed	0.32
ST8C2		0.59
ST8C3		0.97
	<u>Ser-Ala-Ala-Asn-Lys</u>	
ST11	Met(<u>Val,Phe,Pro,Gly,Leu,TML,Lys,Pro,Glx,Glx,Arg</u>) digested with chymotrypsin. Only one peptide recovered	
ST11C1	<u>TML-Lys-Pro-Glx-Glx-Arg</u>	1.28
Papain digestion		
SC2(a)	<u>Asx</u> (Glx,Ala,Pro,Pro,Gly,Asx,Val,Lys,Ser,Gly,Glu,Lys,Ile,Phe)	
SC2(a)P1	<u>Ser-Glx-Ala</u>	0.64
SC2(a)P2	<u>Ser-Gly-Glu</u> (Lys,Ile,Phe)	0.74
SC2(a)P3	<u>Asx-Val-Lys</u>	1.03
CC2(a)	<u>Asx-Glx-Ala</u> (Pro,Pro,Gly,Asx,Val,Lys,Ala,Gly,Glu,Lys,Ile,Phe)	
CC2(a)P1	<u>Asx-Glx-Ala-Pro-Pro</u> (Gly)	0.35
CC2(a)P2	<u>Asx-Glx-Ala-Pro-Pro</u> (Gly,Asx,Val,Lys)	0.56
CC2(a)P3	<u>Ala-Gly-Glu</u> (Lys,Ile,Phe)	0.75
CC2(a)P4	<u>Asx-Val-Lys-Ala-Gly</u> (Glu,Lys,Ile,Phe)	1.86
CC2(a)P5	<u>Asx-Val-Lys</u>	

one rather than two acid residues in these positions, and therefore more cleavage occurred. Both the total amino acid analysis of peptide T1 and of the whole protein are now in even closer agreement with the sequence information. The correct sequence of mung-bean cytochrome *c* is shown in Fig. 1.

DISCUSSION

Sesame, castor and mung-bean cytochromes *c* have been shown to be a closely related set of mitochondrial-type cytochrome *c* species, each consisting of 111 residues. The phylogenetic implications of these results are discussed elsewhere (Boulter, Thompson, Richardson & Ramshaw, 1970). Determination of these sequences does not change the number of invariant amino acid residues found in cytochrome *c* generally, but glutamic acid has not been found previously in position 7. All plant cytochromes examined so far have an *N*-acetylated 'tail' of eight amino acids and two residues for ϵ -*N*-trimethyl-lysine in positions 72 and 86.

We thank Dr B. S. Hartley, Medical Research Council Laboratory of Molecular Biology, Cambridge, U.K., for

advice and discussion, and Mrs L. Mayo for invaluable assistance with the dansyl-Edman analysis. We thank the Nuffield Foundation and the Science Research Council for financial support.

REFERENCES

- Boulter, D., Thompson, E. W., Richardson, M. & Ramshaw, J. A. M. (1970). *Nature, Lond.*, **228**, 552.
 Dayhoff, M. O. & Eck, R. V. (1968). *Atlas of Protein Sequence and Structure*, p. 7. Silver Spring, Md.: National Biomedical Research Foundation.
 Delange, R. J., Glazer, A. N. & Smith, E. L. (1969). *J. biol. Chem.* **244**, 1385.
 Fitch, W. M. & Margoliash, E. (1967). *Science, N.Y.*, **155**, 279.
 Nolan, C. & Margoliash, E. (1966). *J. biol. Chem.* **241**, 1049.
 Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1970). *Biochem. J.* **119**, 535.
 Richardson, M., Laycock, M. V., Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1970). *Phytochemistry*, **9**, 2271.
 Schmer, G. & Kreil, G. (1969). *Analyt. Biochem.* **29**, 186.
 Stevens, F. C., Glazer, A. N. & Smith, E. L. (1967). *J. biol. Chem.* **242**, 2764.
 Thompson, E. W., Laycock, M. V., Ramshaw, J. A. M. & Boulter, D. (1970). *Biochem. J.* **117**, 183.