The Amino Acid Sequence of Cytochrome *c* from *Abutilon theophrasti* Medic. and *Gossypium barbadense* L. (Cotton)

By E. W. THOMPSON, B. A. NOTTON,* M. RICHARDSON AND D. BOULTER
Department of Botany, University of Durham, Durham City, U.K.

(Received 17 March 1971)

The amino acid sequences of *Abutilon* and *Gossypium* cytochromes *c* were determined on 1 μ mol of protein. The molecules consist of 111 residues and are homologous with other mitochondrial plant cytochromes *c*. Experimental details are given in a supplementary paper that has been deposited as Supplementary Publication SUP 50005 at the National Lending Library for Science and Technology, Boston Spa, Yorks. LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1971), 121, 7.

The determination of the amino acid sequences of *Abutilon* and cotton cytochromes *c* is a continuation of the investigation of the sequences of a wide range of plant cytochromes *c* (Boulter, Thompson, Ramshaw & Richardson, 1970).

MATERIALS

Abutilon and cotton seeds were obtained from the Incorporated Oil Seed Association, London E.C.2, U.K. Other materials were as previously described (Thompson, Richardson & Boulter, 1971a).

METHODS

The isolation and purification of cytochrome *c* from *Abutilon* and cotton is described by Richardson, Laycock, Ramshaw, Thompson & Boulter (1970) and Richardson,

Richardson, Ramshaw, Thompson & Boulter (1971). The E_{550} (reduced)/ E_{280} (oxidized) ratios were 1.10 and 1.0 for *Abutilon* and cotton respectively.

The methods used for the determination of the amino acid sequence were as previously described (Thompson *et al.* 1971a). Haem groups were removed from haem peptides as described by Ramshaw, Thompson & Boulter (1970).

RESULTS

The amino acid compositions of *Abutilon* and cotton cytochromes *c* are given in Table 1. The complete sequences of these cytochromes have been determined (see Figs. 1 and 2 respectively). In common with castor cytochrome *c*, hydrolysates of both *Abutilon* and cotton cytochromes *c* contained 4-hydroxyproline (Thompson *et al.* 1971a). This amino acid was found in hydrolysates of a number of the proline-containing peptides of *Abutilon* cytochrome *c*, which were analysed on the automatic

Fig. 1. Amino acid sequence of *Abutilon* cytochrome *c*. ---, Composition determined qualitatively; —, sequence determined by the dansyl-Edman method; ←, sequence determined by using carboxypeptidase; ††, major enzymic cleavages; †, partial enzymic cleavages. Abbreviations: Me₃Lys, ε-trimethyllysine; T, tryptic; C, chymotryptic. For full details of individual peptides, see the supplementary paper SUP 50005.

Table 1. *Amino acid composition of Abutilon and cotton cytochromes c*

The values for the amino acid analysis are the averages for two 24 h and one 72 h hydrolysates. Abbreviations: n.d., not determined; Me₃Lys, ε-trimethyl-lysine.

	Amino acid composition (residues/molecule)			
	<i>Abutilon</i> cytochrome <i>c</i>		Cotton cytochrome <i>c</i>	
	Amino acid analysis	Calc. from sequence	Amino acid analysis	Calc. from sequence
Hyp	3.73	—		
Asp	11.20	11	11.15	11
Thr	7.06*	7	8.61*	7
Ser	4.73*	5	6.33*	5
Glu	10.00	10	8.31	10
Pro	8.30	8	6.64	7
Gly	11.73	12	10.48	12
Ala	11.42	12	12.91	13
Val	3.83†	3	4.72†	3
Met	1.90	2	1.77	2
Cys	n.d.	2	n.d.	2
Ile	3.47†	3	3.57†	3
Leu	8.20	8	7.98	8
Tyr	4.76	6	4.74	6
Phe	3.53	4	3.48	4
Me ₃ Lys	1.60	2		2
Lys	11.91	11	11.72	11
His	2.00	2	1.23	2
Arg	2.33	2	3.11	2
Trp	n.d.	1	n.d.	1

* Calculated assuming first-order kinetics for destruction.

† Maximal yield.

Fig. 2. Amino acid sequence of cotton cytochrome *c*. The abbreviations and symbols used are described in the legend of Fig. 1.

amino acid analyser, but proline was always present as well, and hydroxyproline was never in sufficient amounts to fully account for a whole residue. This was also true for some peptides of cotton cytochrome *c*, which were analysed by

hydrolysis followed by dansylation. Hydroxyproline was never found in any position in the sequence of either cytochrome by using the dansyl-Edman method. DNS-hydroxyproline has been shown to be stable to acid hydrolysis (J. A. M. Ramshaw,

Fig. 3. Positions of tryptic and chymotryptic cleavage of plant cytochromes *c*. The single-letter code used for the amino acids is that given in the Tentative Rules of the IUPAC-IUB Commission on Biochemical Nomenclature [*Biochem. J.* (1969), 113, 1]: A, alanine; B, aspartic acid or asparagine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; X, ϵ -trimethyl-lysine; Y, tyrosine; Z, glutamic acid or glutamine. The positions of enzymic cleavage are shown by: ↓, major tryptic; ↓, minor tryptic; ↑, major chymotryptic; ↑, minor chymotryptic.

unpublished work). The discrepancy between the results obtained with total amino acid analysis, where hydroxyproline was detected in some peptides, and the sequential dansyl analysis, where it was not, was most likely due to incomplete release of the DNS-hydroxyproline from peptides during dansyl-Edman analysis, because suspected proline residues were only hydrolysed for 6 h. If this was the case, and evidence for incomplete release under these conditions has now been obtained with leek peptides (R. Brown, unpublished work), the most likely explanation of our finding is that partial hydroxylation of certain proline residues occurs during the purification of the cytochrome *c*.

The alternative possibility, that the hydroxyproline residues were present in peptides from a contaminating protein, is unlikely, since the cytochrome *c* preparations gave a single band on polyacrylamide gels after electrophoresis at various pH values.

Consideration of the electrophoretic mobilities of the peptides from both cytochromes indicated that each contained 13 amide groups, eight of which could be placed in position in the sequence from the mobilities of the peptides in which they occur. The remaining five amide positions were established from considerations of homology. Glutamine, rather than glutamic acid, was placed in positions 24 and 97, on the basis of homology with all other cytochrome *c* sequences so far determined (Dayhoff, 1969). Glutamine at position 4 and asparagine at positions 10 and 70 were placed by homology with other plant cytochrome *c* sequences determined (Stevens, Glazer & Smith, 1967; Ramshaw *et al.* 1970; Thompson, Laycock, Ramshaw & Boulter, 1970; Thompson, Richardson & Boulter, 1971*a,b,c*).

DISCUSSION

The complete sequence of *Abutilon* cytochrome was deduced unambiguously from the sequences of chymotryptic peptides and by homology with other plant cytochromes *c*. The sequence of cotton cytochrome *c* was deduced from both tryptic and chymotryptic peptides. When the latter were compared with chymotryptic peptides obtained from *Abutilon* cytochrome *c*, an equivalent number of peptides of identical length and sequence were found, except for three amino acid differences. This result clearly indicates the close homology between these two proteins.

In summarizing the specificity of action of chymotrypsin on the cytochromes *c* of mung bean, sunflower, castor, sesame, *Abutilon*, cotton, buckwheat, cauliflower and pumpkin, the following points may be made (see Fig. 3). Chymotrypsin had the expected specificity against phenylalanine and tyrosine, except that in no instance was there cleavage

Table 2. Matrix of amino acid differences in sequences of plant cytochromes *c*

	Wheat ¹	Mung bean ²	Castor ³	Sesame ³	Sunflower ⁴	Cauliflower ⁵	Buckwheat ⁵	Pumpkin ⁶	<i>Abutilon</i> ⁷	Cotton ⁷
Wheat										
Mung bean	15									
Castor	13	8								
Sesame	15	7	5							
Sunflower	15	12	12	12						
Cauliflower	10	5	6	8	14					
Buckwheat	18	12	11	10	15	13				
Pumpkin	13	5	7	6	14	3	13			
<i>Abutilon</i>	15	9	5	7	12	9	11	8		
Cotton	13	7	4	7	12	7	11	8	3	

References: ¹Stevens *et al.* (1967); ²Thompson *et al.* (1970); ³Thompson *et al.* (1971*a*); ⁴Ramshaw *et al.* (1970); ⁵Thompson *et al.* (1971*c*); ⁶Thompson *et al.* (1971*b*); ⁷present work.

at phenylalanine-90, which occurs beside proline. Cleavage at histidine-26 also never occurred, presumably owing to its association with the haem moiety. Cleavage always occurred at methionine residues except for methionine-63 in *Abutilon* cytochrome *c*, where cleavage was at the adjacent asparagine residue instead. Cleavage occurred at only some of the asparagine and leucine residues; for leucine, these were always at the same positions for all cytochromes. This was also normally so for asparagine residues except for asparagine-60 and -62; usually a cleavage occurred at one or other of these, although in cotton cytochrome *c* neither was hydrolysed. There were a few examples of cleavage at what are normally considered tryptic cleavage sites, i.e. castor cytochrome *c* at lysine-12, sesame cytochrome *c* at lysine-12 and arginine-99, and sunflower cytochrome *c* at lysine-61 and arginine-99.

The results in Fig. 3 show that trypsin hydrolysed at all lysine and arginine residues in all the cytochromes *c* except lysine-95, which occurs beside a proline residue. In addition, trypsin cleaved at tyrosine-105 in every cytochrome *c* sequence except that of cotton, and it also cleaved at asparagine-41 and tyrosine-56 and -73 in that of sesame, and at tyrosine-56 in that of pumpkin.

Table 2 is a matrix of amino acid differences between the cytochromes *c* of higher plants. These results show that the sequences of *Abutilon* and cotton cytochromes *c* are more similar to each other than either is to any other sequence. The fact that the two species are members of the same family, the Malvaceae, gives further strong support to the idea originally established from studies on mainly vertebrate animals (Fitch & Margoliash, 1967), that this type of data can be used to construct phylogenies objectively.

When viewed against the vast number of different plant genera that exist, the present data are indeed few, and the phylogenetic implication of these results must await the determination of sequences of more plant cytochromes *c*.

We thank Mrs L. Mayo for valuable assistance with the dansyl-Edman analyses, the Incorporated Oil Seed Association, London E.C.2, U.K., for materials, and the Agricultural Research Council, Nuffield Foundation and Science Research Council for financial support.

REFERENCES

- Boulter, D., Thompson, E. W., Ramshaw, J. A. M. & Richardson, M. (1970). *Nature, Lond.*, **228**, 552.
Dayhoff, M. O. (1969). *Atlas of Protein Sequence and Structure*, vol. 4. Silver Spring, Md.: National Biomedical Research Foundation.
Fitch, W. M. & Margoliash, E. (1967). *Science, N.Y.*, **155**, 279.
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.
Richardson, M., Richardson, D. L., Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1971). *J. Biochem., Tokyo*, **69**, 811.
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.
Thompson, E. W., Richardson, M. & Boulter, D. (1971a). *Biochem. J.* **121**, 439.
Thompson, E. W., Richardson, M. & Boulter, D. (1971b). *Biochem. J.* **124**, 779.
Thompson, E. W., Richardson, M. & Boulter, D. (1971c). *Biochem. J.* **124**, 783.