Phytochemistry, 1973, Vol 12, pp. 2617 to 2622. Pergamon Press. Printed in England.

THE AMINO ACID SEQUENCE OF *CANNABIS SATIVA* CYTOCHROME-*c*

DONALD G. WALLACE*, RICHARD H. BROWN† and DONALD BOULTER

Department of Botany, University of Durham, Durham City DHI 3LE

(*Received* 15 *January* 1973. *Accepted* 5 *June* 1973)

Key Word Index—*Cannabis sativa*; Cannabinaceae; hemp; cytochrome-*c*; amino acid sequence; dansyl-Edman method.

Abstract—The amino acid sequence of hemp (*Cannabis sativa* L.) cytochrome-*c* was determined using 0·7 μmol of protein. Analysis of chymotryptic and tryptic peptides by the dansyl Edman method showed that the molecule consisted of 111 residues and was homologous with other mitochondrial plant cytochromes-*c*.

INTRODUCTION

THE AMINO ACID sequence of hemp (*Cannabis sativa* L.) cytochrome-*c* has been determined as part of an investigation of the sequences of a wide range of plant cytochromes-*c*, from which it is hoped to establish a computer-generated phylogenetic tree for higher plants.[1]

RESULTS AND DISCUSSION

Because of difficulties involved in obtaining sufficient quantities of viable hemp seed, purified cytochrome was available only for chymotryptic and tryptic digestions; the amino acid composition of the intact protein was not determined. The electrophoretic mobilities and the amino acid sequences of hemp tryptic and chymotryptic peptides are given in Tables 1 and 2. The complete amino acid sequence was deduced from these peptides and is given in Fig. 1. The evidence for the sequence of the residues is indicated.

Peptides obtained by chymotryptic digestion of the protein gave the sequence of about three-fourths of the molecule; the blocked *N*-terminal peptide (C1) and the haem peptide (C2) were the least satisfactorily analysed. The tryptic peptides, however, permitted the sequence determination for the missing sections, as well as providing overlapping results for much of the chymotryptic sequence. Some of the long tryptic peptides were sequenced part way through, and then for a section in the middle of the peptide, Edman degradation was carried out without removing aliquots for dansylation.[2] After several steps in this manner, normal dansyl-Edman analysis was resumed, and sequencing was continued to the end of the peptide. In this way, it was possible to determine residues near the *C*-terminal end, which ordinarily might have been difficult to identify because of diminished amounts of peptide. Peptides T6, T7 and T9 were sequenced by such a procedure. The middle residues which were 'skipped', already had been determined from analysis of chymotryptic peptides.

* Carnegie Institution of Washington, Department of Terrestrial Magnetism, 5241 Broad Branch Road, N.W., Washington, DC 20015, U S.A.

† Postgraduate School of Biological Sciences, University of Bradford, Bradford, Yorkshire DB7 1DP.

[1] BOULTER, D., RAMSHAW, J. A. M., THOMPSON, E. W., RICHARDSON, M. and BROWN, R. H. (1972) *Proc. Roy. Soc.* (*London*) **181B**, 441.

[2] THOMPSON, E. W., LAYCOCK, M. V., RAMSHAW, J. A. M. and BOULTER, D. (1970) *Biochem. J.* **117**, 183.

Consideration of the electrophoretic mobilities of the non-haem peptides indicated that there were ten amide groups, seven of which were placed by the mobilities of the peptides in which they occurred. Two of the remaining three amides were tentatively placed by homology. Comparison with other plant cytochromes-*c* suggests that residue 98 is acidic, and therefore 97 is probably glutamine. By similar reasoning, position 5 should be glutamic acid, and therefore, either position 4 or 10 is an amide. Peptide mobilities also require that either position 68 or 69 is amidated.

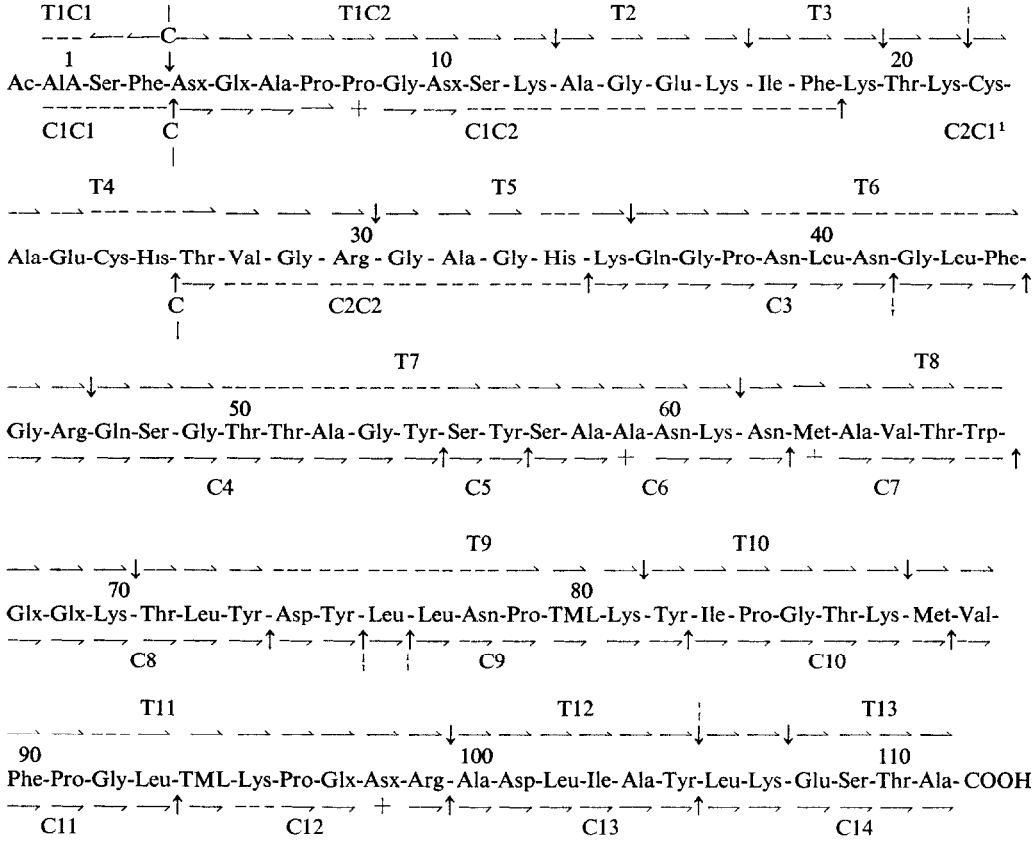

FIG. 1. AMINO ACID SEQUENCE OF *Cannibis sativa* CYTOCHROME-*c*.

--- Composition determined qualitatively.
—→ Sequence determined by dansyl-Edman degradation
←— Sequence determined using carboxypeptidase-A
\+ No assignment could be made during dansyl-Edman analysis.
↓ Major enzymic cleavage.
⋮
↓ Partial enzymic cleavage.
C
↓ Cleavage after second digestion of peptide with chymotrypsin.

Abbreviations: TML—E-*N*-trimethyllysine, T—tryptic peptide; C—chymotryptic peptide.

[1]The chymotryptic fragment from the haem peptide, C2C1, was not isolated. Aspartyl and glutamyl residues, whose state of amidation was uncertain, are represented as Asx and Glx, respectively. For full details of the individual peptides see Tables 1 and 2

TABLE 1. DANSYL-EDMAN ANALYSIS OF TRYPTIC PEPTIDES OF *Cannabis sativa* CYTOCHROME-*c*

Peptide	Dansyl-Edman results	Mobility at pH 6·5*
T1	Blocked *N*-terminus; digested with chymotrypsin for 2 hr	−1·24
T1C1	(Acetyl-Ala)-Ser-Phe	
T1C2	Asx-Glx-Ala-Pro-Pro-Gly-Asx-Ser-Lys	
T2	Ala-Gly-Glu-Lys	0
T3	Ile-Phe-Lys	+1·10
T4(a)	Thr-Lys	+1·60
T4(b)	Thr-Lys-Cys-Ala-Glu(Cys)(His)Thr-	
	Val-Gly-Arg	+0·60
	(peroxide oxidized to remove haem)	
T5	Gly-Ala-Gly(His)Lys	+2·14
T6	Gln-Gly-Pro(Asn,Leu,Asn,Gly,Leu)-	
	Phe-Gly-Arg	+0·60
T7	Gln-Ser-Gly(Thr,Thr,Ala,Gly,Tyr)	
	Ser-Tyr-Ser-Ala-Ala-Asn-Lys	+0·50
T8	Asn-Met-Ala-Val-Thr(Trp)Glx-Glx-Lys	−0·10
T9	Thr-Leu-Tyr(Asp,Tyr,Leu,Leu,Asn)Pro-	
	TML-Lys	+0·50
T10	Tyr-Ile-Pro-Gly-Thr-Lys	+0·74
T11	Met-Val-Phe-Pro(Gly)Leu-TML-Lys-	
	Pro-Glx-Asx-Arg	+1·05
T12(a)	Ala-Asp-Leu-Ile-Ala-Tyr(Leu,Lys)	−0·10
T12(b)	Leu-Lys	+1·64
T13	Glu-Ser-Thr-Ala	−1·50

* Peptides resulting from tryptic digestion were purified by paper electrophoresis at pH 6·5 only. They were not rerun at pH 1·9, as were the peptides of the chymotryptic digest. In cases where more than one peptide was present in the same fraction, such as peptides T8 and T12a, sequences of individual peptides were resolved by reference to the chymotryptic digest.

Amides and their parent acid residues established from peptide mobilities are shown, but those placed later by homology are 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 (—→) indicate positions confirmed by dansyl-Edman analysis. Arrows (←—) indicate positions confirmed by carboxy-peptidase-A digestion followed by dansylation. The sign (+) indicated that no assignment could be made for the position during the dansyl-Edman analysis. Residues in brackets were determined from peptide composition and the order determined by homology of other evidence.

TABLE 2. DANSYL-EDMAN ANALYSIS OF CHYMOTRYPTIC PEPTIDES OF *Cannabis sativa* CYTOCHROME-*c*

Peptide	Dansyl-Edman results	Mobilities at pH 6·5	Mobilities at pH 1·9
C1	Blocked *N*-terminus; further digestion with chymotrypsin 2 hr	−0·80	+0·38
C1C1	Not isolated (Acetyl,Ala,Ser,Phe)		
C1C2	Asx-Glx-Ala-Pro- + -Gly-Asx(Ser, Lys,Ala,Gly,Glu,Lys,Ile,Phe)		
C2	Haem peptide; peroxide oxidized and digested further with chymotrypsin	+0·50	
C2C1	Not isolated. (Lys,Thr,Lys,Cys,etc.)		
C2C2	Thr(Val,Gly,Arg,Gly,Ala,Gly,His)		
C3(a)	Lys-Gln-Gly-Pro-Asn-Leu-Asn-Gly-		
	Leu-Phe	+0·60	0·56
C3(b)	Gly-Leu-Phe	0	0·58
C4	Gly-Arg-Gln-Ser-Gly-Thr-Thr-Ala-		
	Gly-Tyr	0·60	0·56
C5	Ser-Tyr	0	0·58
C6	Ser-Ala- + -Asn-Lys-Asn	+1·10	0·73
C7	+ -Ala-Val-Thr(Trp)	0	0·35
C8	Glx-Glx-Lys-Thr-Leu-Tyr	0	0·65
C9(a)	Asp-Tyr	−1·68	—
C9(b)	Asp-Tyr-Leu	−1·30	0·47
C9(c)	Leu-Leu-Asn-Pro-TML(Lys,Tyr)	1·10	0·90
C9(d)	Leu-Asn-Pro-TML-Lys-Tyr	1·30	1·00
C10	Ile-Pro-Gly-Thr-Lys-Met	0·90	0·78
C11	Val-Phe-Pro-Gly-Leu	0	0·35
C12	TML(Lys)Pro-Glx- + -Arg	1·80	1·00
C13(a)	Ala-Asp-Leu-Ile-Ala-Tyr	−1·00	+0·38
C13(b)	Ile-Ala-Tyr	0	0·48
C14	Leu-Lys-Glu-Ser-Thr-Ala	0	0 78

Abbreviations as for tryptic peptides.

In order to decide whether the haem peptides (T4(b) and C2) contained an amide group at residue 24, comparison was made with the haem peptides of rape[3] and spinach[4] cytochromes-*c*. The electrophoretic mobilities of these peptides and their sequences, are shown in Table 3, along with those of hemp cytochrome-*c*. Placement of amides in rape and spinach was unequivocal and was determined by mobilities or pronase digestions. In comparing peptide mobilities of haem peptides from these three species of cytochrome-*c*, contributions to the charge due to *His*, *Cys*, or haem were ignored, since they should be the same for all three. Therefore, at pH 6·5, positive charges were assumed to be due to *Lys* and the *N*-terminal amino group; and negative charges, to *Asp*, *Glu*, and the *C*-terminal carboxyl group. From these considerations, if residue 24 is glutamic acid, the hemp chymotryptic and tryptic peptides should have a charge of +2 and +1, respectively. Similarly, rape peptides should have charges of +2 and +1; and spinach peptides, +1 and 0, in the same order. Since rape and hemp peptides have nearly the same mobilities, we have assumed that hemp residue 24 is not amidated. If it were amidated, the peptide charges would be +3 and +2 respectively, and should give a higher mobility. Therefore, the total number of amide groups remains 10.

TABLE 3. ELECTROPHORETIC MOBILITIES AND AMINO ACID SEQUENCES OF HAEM PEPTIDES FROM CYTOCHROMES-*c* OF THREE SPECIES OF HIGHER PLANTS

Species	Haem peptide mobility at pH 6·5 relative to dansyl-arg-arg Chymotryptic	Tryptic
Hemp (*Cannabis sativa* L.)	0·50	0·60
Rape (*Brassica napus* L.)[6]	0·50	0·46
Spinach (*Spinacia oleracea* L.)[7]	0·40	0·0

Species	Amino Acid Sequence of Haem Peptide
	19 ↓ ⇣ ↓ 34
Hemp	Lys-Thr-Lys-Cys-Ala-Glu-Cys-His-Thr-Val-Gly-Arg-Gly-Ala-Gly-His
Rape	Lys-Thr-Lys-Cys-Ala-Gln-Cys-His-Thr-Val-Asp-Lys-Gly-Ala-Gly-His
Spinach	Lys-Thr-Lys-Cys-Ala-Gln-Cys-His-Thr-Val-Asp-Leu-Gly-Ala-Gly-His
	↑ ↑

As in Fig. 1, vertical solid arrows indicate positions of major enzymic cleavage; vertical dotted arrows indicate positions of partial enzymic cleavage. Arrows from above indicate positions of tryptic cleavage; arrows from below, positions of chymotryptic cleavage.

The sequence of hemp cytochrome-*c* is found to possess those features characteristic of this protein in higher plants, viz. a blocked tail of eight residues compared with mammalian cytochrome-*c*, two residues of trimethyllysine (residues 80 and 94), and threonine in both positions 50 and 51. When compared to the plant sequences reported to date, the hemp sequence is most like that of mung bean[2] and elder.[1,5] A detailed discussion of the phylogenetic implications of this and other recent data will be the subject of a separate publication.

EXPERIMENTAL

Hemp seeds in 20 kg quantities (as bird seed) were purchased from H. Garnham and Son, 85 Front Street, Sacriston, Co. Durham. Isolation and purification of cytochrome-*c* from etiolated hemp seedlings was by chromatography on Amberlite CG-50, CM-Sephadex and Bio-gel P-30 as described by Richardson *et al.*[6,7] The purity ratio (E_{550}red./E_{280}ox.) was 0·8 for the protein used in sequence determination.

[3] RICHARDSON, M., RAMSHAW, J. A. M. and BOULTER, D. (1971) *Biochim. Biophys. Acta* **251**, 331. Also B.B.A. Deposition Number: BBA/DD/001/36010/251 (1971) 331.

[4] BROWN, R. H., RICHARDSON, M., SCOGIN, R. and BOULTER, D. (1973) *Biochem. J.* in press.

[5] BROWN, R. H. and BOULTER, D. (1973) to be published.

[6] RICHARDSON, M., LAYCOCK, M. V., RAMSHAW, J. A. M., THOMPSON, E. W. and BOULTER, D. (1970) *Phytochemistry* **9**, 2271.

[7] RICHARDSON, M., RICHARDSON, D. L., RAMSHAW, J. A. M., THOMPSON, E. W. and BOULTER, D. (1971) *J. Biochem. Tokyo* **69**, 811.

The cytochrome-*c* was oxidized with $K_3Fe(CN)_6$ and then denatured with EtOH prior to digestion with either chymotrypsin or trypsin.[8] For chymotryptic digestion 0·4 μmol of cytochrome-*c* was used; for tryptic digestion, 0·3 μmol. The resulting peptides were separated by paper electrophoresis at 9 kV and at pH of 6·5 and 1·9.[2] The amino acid sequence of the peptides were determined using the method of Gray[9] as previously described.[2] The dansyl-amino acids were identified by TLC on polyamide sheets using solvent systems previously described.[10]

The haem moiety was removed from haem peptides by oxidation with performic acid.[10] The haem-containing chymotryptic peptide was further digested with chymotrypsin (5% w/w, enzyme/peptide) in 0·2 M NH_4HCO_3 at 37° for 4 hr Aliquots of the tryptic peptide T1Cl (Table 1) was further digested with carboxypeptidase A using 0·2–0 5 mg of enzyme per mol of peptide in 0·2 M NH_4HCO_3 buffer at pH 8·4 for 3 and 29 hr.

That the protein was blocked at the *N*-terminus was shown by an analysis of peptides T1 and C1 (Tables 1 and 2). Neither peptide reacted with ninhydrin on paper nor with dansyl chloride after elution from paper. Both were identified by the amino acid composition after total hydrolysis and dansylation of aliquots The sequences of the blocked peptides were obtained after cleavage with chymotrypsin, followed by dansyl-Edman or carboxypeptidase A analysis Because of limiting amounts of protein, the nature of the *N*-terminal blocking group was not ascertained. In all higher plant cytochromes-*c* which have been sequenced, the *N*-terminal block is an acetyl group This was assumed to be the case in hemp, as well

Acknowledgements—We thank Messrs. G. Bainbridge and J. Gilroy and Mrs E. Innes, for invaluable technical assistance. Thanks are due to the Nuffield Foundation and the Science Research Council for financial assistance. This work was also supported, in part, by the award of a Leverhulme Trust Fellowship to D.G W

[8] THOMPSON, E. W., RICHARDSON, M. and BOULTER, D. (1970) *Biochem J.* **121,** 439.
[9] GRAY, W. R. and HARTLEY, B S. (1963) *Biochem. J.* **89,** 379.
[10] RAMSHAW, J. A. M., THOMPSON, E. W. and BOULTER, D. (1970) *Biochem. J* **119,** 535.