Phytochemistry. 1975. Vol. 14. pp. 1945–1949. Pergamon Presss. Printed in England.

THE AMINO ACID SEQUENCE OF CYTOCHROME *c* FROM NIGER-SEED, *GUIZOTIA ABYSSINICA*

JOHN A. M. RAMSHAW* and DONALD BOULTER†

Department of Botany, University of Durham, Durham City DH1 3LE, England

(*Received* 10 *September* 1974)

Key Word Index—*Guizotia abyssinica*; Compositae; niger-seed; cytochrome *c*; amino acid sequence; dansyl-Edman; phylogeny.

Abstract—The amino acid sequence of cytochrome *c* from niger-seed has been determined by sequence analysis of chymotryptic and tryptic peptides using the dansyl–phenylisothiocyanate method and by qualitative analysis of peptide composition by the dansyl method. Although the spectral ratios indicated the protein was not completely pure, no indication of impurity was found during the sequence analysis and no peptides in addition to those given here were obtained. In certain cases the alignment of peptides was by homology with other cytochromes *c*. Four residues in the proposed sequence, alanine-1, cysteine-25, histidine-26 and lysine-61 were identified only from peptide compositions. The amino-terminus of the protein is acetylated. The sequence contains two residues of ϵ-*N*-trimethyllysine.

INTRODUCTION

The amino acid sequence of cytochrome *c* from *Guizotia abyssinica* has been determined as part of an investigation of the sequences of cytochromes *c* from a wide range of plant species, from which a computer generated affinity tree for higher plant cytochrome *c* has been established [1, 2].

RESULTS AND DISCUSSION

The proposed amino acid sequence of cytochrome *c* from *Guizotia* is given in Fig. 1. Every residue in the proposed sequence was identified during dansyl–phenylisothiocyanate analysis of either the chymotryptic or tryptic peptides, except for alanine-1, cysteine-25, histidine-26 and lysine-61, which were identified from peptide compositions. The amino acid compositions of the protein determined by analysis and from the sequence are in fair agreement (Table 1). However, the values obtained by analysis for serine, glycine and valine are higher than expected from the sequence, while the values of threonine and alanine are lower than expected. The values obtained for those amino acids which are susceptible to oxidation losses were also all low. None of the values given in Table 1 has been corrected for losses due to destruction during hydrolysis. The values obtained for the basic amino acids were not completely satisfactory from the 20 hr HCl hydrolyses, mainly due to bad base-line variations. In order to obtain better values an additional sample was hydrolysed with 3 M toluene–sulphonic acid [3]; this also allowed a value for tryptophan to be obtained. In neither of two analytical systems which were used was the ϵ-*N*-trimethyllysine resolved. The presence of two residues of this amino acid in positions 80 and 94 was clearly shown during the dansyl–phenylisothiocyanate analysis.

The low value of the E_{550} (reduced)/E_{280} (oxidized) ratio for the pooled material used for this investigation showed that the preparation was not entirely pure. A small quantity of material

* Present address: Department of Inorganic Chemistry, University of Sydney, N.S.W. 2006.

† To whom requests for reprints should be sent.

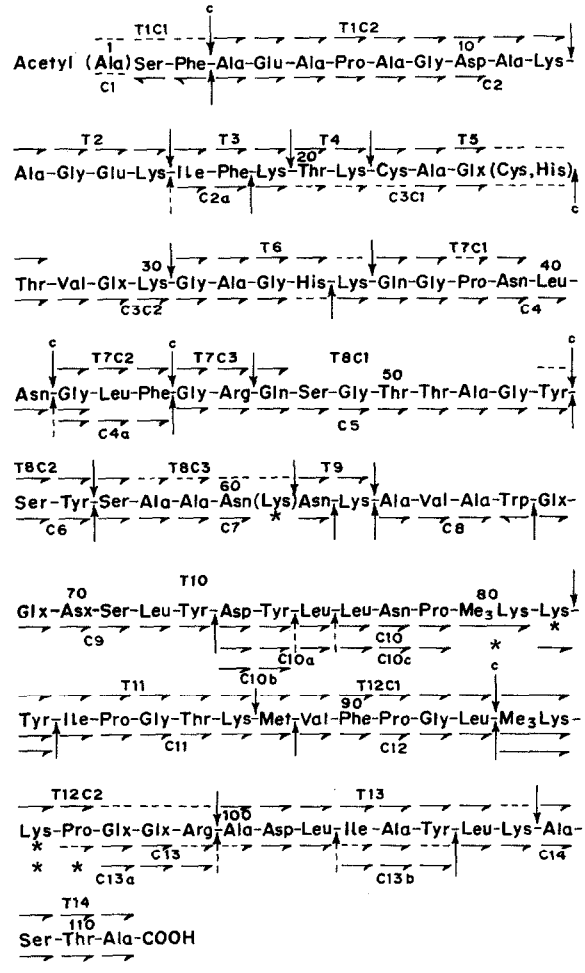

Fig. 1. The amino acid sequence of cytochrome *c* from *Guizotia abyssinica*. → Sequence determined by dansyl–phenylisothiocyanate method. ⇢ Sequence determined by dansyl–phenylisothiocyanate method, but results not considered entirely satisfactory [17]. * No assignment was made during the dansyl–phenylisothiocyanate degradation. ← Sequence determined by Carboxypeptidase A digestion. --- Composition determined qualitatively [5]. ( ) Residue inferred from peptide compositions only. ↑↓ Major enzyme cleavage. ↓↑ Partial enzyme cleavage. $\overset{\uparrow}{_{c}}$ Further chymotryptic cleavage of purified peptide. *Abbreviations*: Me_3Lys—ε-*N*-trimethyllysine; T—tryptic peptide; C—chymotryptic peptide.

with higher ratio was obtained during the purification but was insufficient for sequence analysis. However, during the sequence analysis, there was no indication of the impurity present and no peptides additional to those given in Fig. 1 and Table 1 were identified.

The peptides obtained from the chymotryptic digest (Table 1 and Fig. 1) enabled the majority of the residues in the sequence to be determined. The remaining residues, except for alanine-1, cysteine-25, histidine-26 and lysine-61, were all identified from the tryptic peptides (Table 1 and Fig. 1). These four other residues were identified only from peptide compositions. The tryptic peptides enabled a considerable number of the overlaps to be established; the remainder were established by homology with other cytochromes *c* [4].

Certain of the peptides needed special analysis. The sequence of the *N*-terminal peptide C1 was determined by digestion with carboxypeptidase A. After 3 hr phenylalanine only was released and after 24 hr both phenylalanine and serine were released. Since the compositions of this peptide and of peptide T1C1 indicated that

Table 1. The amino acid composition of cytochrome *c* from *Guizotia abyssinica*

	20 hr HCl hydrolysis*	70 hr Toluene–sulphonic acid hydrolysis†	Sequence value
Asp	9·8		9
Thr	4·4		6
Ser	7·4		6
Glu	9·9		10
Pro	5·8		6
Gly	12·6		11
Ala	15·5		17
Val	3·8		3
Cys‡	0·8		2
Met	0·5		1
Ile	2·8		3
Leu	8·0		8
Tyr	5·1		6
Phe	3·7		4
Trp		0·8	1
Lys	12·2	12·0	12
Me_3Lys§		4·2	2
His	2·4		2
Arg	3·2	2·2	2
Total			111

* Run on a Locarte amino acid analyser. Values have not been corrected for losses caused by destruction during hydrolysis. † Run on an EEL amino acid analyser basic amino acids column. ‡ As cysteic acid. § Me_3Lys was not resolved by either analyser system. In the HCl hydrolysate column, the value is given for Lys + Me_3Lys; in the toluene–sulphonic acid hydrolysate column, as Me_3Lys + His.

alanine was also present in the peptide, this amino acid was assumed to occupy position-1 in the sequence. Peptide T1C1 was pooled with the remainder of peptide C1 for the identification of the acetyl blocking group, which was identified as the 1-acetyl-2-dansyl–hydrazine derivative. The haem peptides were eluted after electrophoresis by 20% (v/v) pyridine. After removal of the haem group they were purified by electrophoresis at pH 1·9. In addition the chymotryptic haem peptide was further digested with chymotrypsin and after electrophoresis at pH 1·9, two fragments were recovered in low yield resulting from cleavage at histidine-26. The presence of this residue and of cysteine-25 were inferred from the compositions of the haem peptides and fragments, T5, C3C1 and C3C2 and their order based on the specificity observed for the chymotryptic digestion of peptide C3. The amino acid composition of the protein (Table 1), suggested that one tryptophan residue was present. From the chymotryptic digest only one Ehrlich positive peptide was located, peptide C8, and the C-terminal tryptophan residue of this peptide was identified after carboxypeptidase A digestion for 90 min. From the tryptic digest a single Ehrlich positive peptide was observed after electrophoresis at pH 6·5, but the peptide was not isolated after further electrophoresis at pH 1·9. Peptides C14 and T14 were both found to have alanine as their C-terminal residue. For peptide T14, a sample taken after three steps of degradation which was dansylated but not hydrolysed co-chromatographed with dansyl–alanine; this confirmed that alanine was the C-terminal of this peptide. A similar experiment confirmed the C-terminal of peptide C14. Since alanine is not normally susceptible to both chymotryptic and tryptic cleavage these peptides were assumed to be the C-terminal of the protein.

For both digests, the observed enzyme specificities were consistent with those expected. Full chymotryptic cleavage also occurred at histidine-34, asparagine-62, lysine-63 and methionine-88; partial cleavage occurred at lysine-16, asparagine-41, tyrosine-75, leucine-76, arginine-99 and leucine-102. The cleavage at asparagine-62 and lysine-63 meant that there was no chymotryptic peptide involving lysine-63.

The electrophoretic mobilities of the peptides at pH 6·5 (Table 2) indicated the presence of 10 amide residues in the sequence. The presence of one amide residue in the haem peptide region was suggested by comparison of the haem peptide mobilities with those of haem peptides of known amide content [5]. Of the amide residues, 7 can be placed directly; these are asparagine in positions 39, 41, 60, 62 and 78 and glutamine in positions 36 and 47.

Guizotia abyssinica is the second member of the Compositae whose cytochrome *c* sequence has been determined. Although its sequence differs from that of the other, *Helianthus annuus*, in at least 8 positions, the two sequences are more closely related to each other than either sequence is to any other plant cytochrome *c* sequence since they are always found together on the same line of descent on affinity trees calculated using sequence data [1, 2]. In the other instances where the cytochromes *c* of closely related species have been examined, i.e.

Table 2. The electrophoretic mobilities of purified peptides from *G. Abyssinica* cytochrome *c*

Peptide	Mobility* at pH 6·5	Mobility* at pH 1·9	Peptide	Mobility* at pH 6·5	Mobility* at pH 1·9
C1	−1·30	0	T1		0·28
C2	−0·66		T1C1		0
C2(a)	0	0·67	T1C2		0·53
C3	0·65		T2	0	1·06
C3C1		0·61	T3	1·30	
C3C2		0·98	T4	1·69	
C4	0·90†	0·55	T5	0	0·31
C4(a)	0	0·59	T6	2·18	
C5	0·90	0·50	T7	0·64	0·53
C6	0	0·64	T7C1		0·30
C7	1·09		T7C2		0·58
C8	0	0·45	T7C3		1·46
C9	−1·67	0·34	T8	0·64	0·45
C10	(0·6)	0·71	T8C1		0·32
C10(a)	−1·67	0·51	T8C2		0·58
C10(b)	−1·53	0·45	T8C3		0·95
C10(c)	1·70	0·99	T9	1·69	
C11	1·09	0·83	T10	0	
C12	0	0·47	T11	0·91	
C13	(0·6)		T12	1·10	(0·9)
C13(a)	(1·7)		T12C1		0·44
C13(b)	0		T12C2		1·26
C14	(1·0)	0·85	T13	0	0·63
			T14	0	0·56

* Mobilities at pH 6·5 were measured from the position of the neutral amino acids relative to dansyl-Arg-Arg. Mobilities at pH 1·9 were measured from 1-dimethyl-aminonaphthalene-5-sulphonic acid relative to dansyl-arginine. † Leading edge of zone. Values given in brackets were not accurately determined.

cauliflower [6] and black rape [7] (both Cruciferae), cotton and *Abutilon* [8] (both Malvaceae), the number of sequence differences found between the pairs of sequences have been very small, i.e. 0 and 3 respectively. Thus, while the Compositae form a closely related natural group with characteristic morphology and cytology, the finding of eight differences in the two sequences from two members of the same tribe (Heliantheae), suggests that considerable evolutionary diversity may exist.

EXPERIMENTAL

Materials. G. abyssinica seeds were obtained from H. Garnham and Sons. Chymotrypsin, E.C. 3.4.21.1 (3× recrystallized) and trypsin, E.C. 3.4.21.4 (2× recrystallized, salt-free) were obtained from the Worthington Biochemical Corp. Carboxypeptidase A, E.C. 3.4.12.2 (diisopropylphosphorofluoridate-treated, crystalline suspn in H_2O) was obtained from the Sigma Chemical Co. Other chemicals used were of the highest purity readily available.

Methods. Isolation and purification of cytochrome *c* from *G. abyssinica* has been described previously [9]; the absorbance ratio at 550 nm (reduced) 280 nm (oxidized) = 1·08 for pooled material used in this investigation. Before digestion, cytochrome was oxidized by $K_3Fe(CN)_6$, denatured, and washed with EtOH [5]. It was resuspended for digestion and adjusted to pH 8·0 with 0·025 M NaOH on a Radiometer-TTTlc autotitrator at 37°. For chymotryptic digestion, 2% (w/w, enzyme–substrate) of α-chymotrypsin made up in H_2O (1 mg/ml) was added to 3·5 mg of denatured cytochrome. After 80 min a further 2% enzyme was added and digest terminated after 120 min. For tryptic digestion 2% (w/w, enzyme–substrate) trypsin made up in 1 mM HCl (1 mg/ml) was added to 2·5 mg of denatured cytochrome. After 60 min a further 2% enzyme was added, and digest terminated after 90 min. Both digests were terminated by adjusting to pH 4·5 with AcOH and then freezing and lyophilising. Resulting peptides were purified by high-voltage paper electrophoresis at pH 6·5 and 1·9, were detected, and had their mobilities calculated, as described in [5]. Peptide sequences of purified peptides were determined using the dansyl–phenylisothiocyanate method [10], and in some cases by a micromethod of this procedure [11]. Identification of dansyl amino acids was by chromatography on polyamide sheets [12] using solvent systems described in [13]. Amide residues were determined where possible from electrophoretic mobility of the peptide at pH 6·5 [14]. Certain purified peptides were further digested with chymotrypsin for 1 hr. Digestion was with 5% (w/w, enzyme–peptide) in pH 8·4, 0·2 M NH_4HCO_3 at 37°. Digestion of peptides with carboxypeptidase A was by using 0·2–0·5 mg enzyme/mol peptide in pH 8·4, 0·2 M NH_4HCO_3 at 37°. Liberated amino acids were detected as

their dansyl derivatives [5]. For removal of the haem moiety before sequence analysis, haem peptides were dissolved in 200 μl of 90% (v/v) formic acid and 2 10 μl portions of 30% (w/v) H_2O_2 were added at 10 min intervals. The reaction was allowed to proceed at room temp for 30 min and then stopped by freezing and drying [13]. The *N*-terminal blocking group was identified by the method of Schmer and Kreil [15] as previously described [13]. The amino acid composition of protein was determined after hydrolysis in 6 M HCl [16] and after hydrolysis in 3 M toluene–sulphonic acid [3] by automatic amino acid analysis [16]. Qualitative analysis of peptide compositions was by the dansyl method as previously described [5].

Acknowledgements—We thank Mr. G. Bainbridge, Mr. J. Gilroy, Mrs. L. Mayo and Mrs. E. Innes for technical assistance. We also thank the Nuffield Foundation and the Science Research Council for financial support.

REFERENCES

1. Boulter, D., Ramshaw, J. A. M., Thompson, E. W., Richardson, M. and Brown, R. H. (1972) *Proc. Roy. Soc. Lond. Ser. B.* **181,** 441.
2. Boulter, D. (1973) in *Nobel Symposium* 25, *Chemistry in Botanical Classification* (Bendz, G. and Santesson, J., eds.), pp. 211–216. Nobel Foundation, Stockholm, Sweden; Academic Press, New York.
3. Liu, T. Y. and Chang, Y. H. (1971) *J. Biol. Chem.* **246,** 2842.
4. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, Vol. 5. National Biomedical Research Foundation, Silver Spring, Md., U.S.A.
5. Thompson, E. W., Laycock, M. V., Ramshaw, J. A. M. and Boulter, D. (1970) *Biochem. J.* **117,** 183.
6. Thompson, E. W., Richardson, M. and Boulter, D. (1971) *Biochem. J.* **124,** 783.
7. Richardson, M., Ramshaw, J. A. M. and Boulter, D. (1971) *Biochim. Biophys. Acta* **251,** 331.
8. Thompson, E. W., Notton, B. A., Richardson, M. and Boulter, D. (1971) *Biochem. J.* **124,** 787.
9. Richardson, M., Richardson, D. L., Ramshaw, J. A. M., Thompson, E. W. and Boulter, D. (1971) *J. Biochem. (Tokyo)* **69,** 811.
10. Gray, W. R. and Hartley, B. S. (1963) *Biochem. J.* **89,** 379.
11. Bruton, C. J. and Hartley, B. S. (1970) *J. Mol. Biol.* **52,** 165.
12. Woods, W. R. and Wang, K. T. (1967) *Biochim. Biophys. Acta* **133,** 369.
13. Ramshaw, J. A. M., Thompson, E. W. and Boulter, D. (1970) *Biochem. J.* **119,** 535.
14. Offord, R. E. (1966) *Nature (Lond.)* **211,** 591.
15. Schmer, G. and Kreil, G. (1969) *Anal. Biochem.* **29,** 186.
16. Moore, S. and Stein, W. H. (1963) in *Methods in Enzymology*, (Colowick, S. P. and Kaplan, N. O., eds.), Vol. 6, pp. 817–831. Academic Press, New York.
17. Ambler, R. P. and Wynn, M. (1973) *Biochem. J.* **131,** 485.