The Amino Acid Sequence of Cytochrome *c* from *Spinacea oleracea* L. (Spinach)

By R. H. BROWN, M. RICHARDSON, R. SCOGIN* and D. BOULTER
Department of Botany, University of Durham, Durham City, U.K.

(Received 31 July 1972)

The amino acid sequence of spinach (*Spinacea oleracea* L., var. *Monster Viroflay*) cytochrome *c* was determined on 1 μ mol of protein. The molecule consists of 111 residues and is homologous with other mitochondrial cytochromes *c*. Experimental details are given in a supplementary paper that has been deposited as Supplementary Publication SUP 50013, 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.* (1973), 131, 5.

This investigation is a continuation of a project aimed at the reconstruction of a phylogenetic tree derived from a comparison of amino acid sequences in plant cytochromes *c* of mitochondrial origin (Boulter *et al.*, 1972).

Materials and Methods

Spinach seeds were obtained from Hurst, Gunson and Tabor, Witham, Essex, U.K.

The isolation and purification of cytochrome *c* from spinach were done essentially as described by Richardson *et al.* (1971). Dry seed (300 kg), after germination and extraction of cytochrome, yielded 15 mg of cytochrome of purity index (E_{550} , reduced/ E_{280} , oxidized) 0.9. The cytochrome *c* was digested with chymotrypsin (4 h) or trypsin (2 h) as described previously (Thompson *et al.*, 1971). The resulting peptides were purified by electrophoresis at pH 6.5 and 1.9 (Thompson *et al.*, 1970). The amino acid sequences of the peptides were determined by using the dansyl-Edman method (Gray & Hartley, 1963) as described previously (Thompson *et al.*, 1970). Dansyl amino acids were identified by chromatography on polyamide sheets (Woods & Wang, 1967) by using the solvent systems described by Ramshaw *et al.* (1971). The haem moiety was removed from haem peptides with performic acid (Ramshaw *et al.*, 1971). Certain peptides were further digested with chymotrypsin (Thompson *et al.*, 1970), carboxypeptidase A (Thompson *et al.*, 1970; 1 h except where stated otherwise), carboxypeptidase C (Tschesche & Kupfer, 1972; 16 h digestion at 37°C) or a protease mixture (Pronase, Sigma type VI from *Streptomyces griseus*;

16 h digestion with 5%, w/w, enzyme/peptide ratio, at 37°C). Released amino acids were identified as the Dns derivatives. The *N*-terminal blocking group was identified by using the method of Schmer & Kreil (1969), and the 1-acyl-2-dansyl-hydrazine derivative identified by t.l.c. (Thompson *et al.*, 1970). The quantitative method of amino acid analysis of peptides (Thompson *et al.*, 1970) was replaced by the semi-quantitative method of dansylating an hydrolysate also described by Thompson *et al.* (1970).

Results

The amino acid compositions of spinach cytochrome *c* determined by analysis and calculated from the sequence, are shown in Table 1. These values are in reasonable agreement. The inconsistencies between averaged 24 h and 72 h values for serine, proline and alanine, are ascribed to high variation between individual analysis values for these amino acids. Averaged analyses for these amino acids are therefore, open to some uncertainty. Tryptophan was not determined, but the best spectral ratios of the cytochrome indicate that one tryptophan residue is present (Richardson *et al.*, 1971).

Fig. 1 gives the complete sequence of spinach cytochrome *c*, together with an indication of the tryptic and chymotryptic peptides identified. The evidence for the sequence of the residues is indicated.

Consideration of the mobilities of the non-haem peptides indicate that there are ten amide groups. Of these, eight may be placed unambiguously. The remaining two amides occur in positions 5 or 10 and positions 97 or 98. Residues 5 and 10 were shown to be glutamic acid and asparagine respectively by total enzymic hydrolysis of peptide T1C2 with Pronase. Residue 10 was confirmed by carboxypeptidase-C

* Present address: Rancho Santa Ana Botanic Gardens, Claremont, Calif., U.S.A.

Table 1. *Amino acid composition of spinach cytochrome c*

The 24 and 72h analyses are the average of two determinations. n.d., Not determined.

	Amino acid composition (residues/molecule)			
	24h	72h	Average	Sequence
Asp	12.6	11.4	12.5	12
Thr	6.7	6.8	6.7	7
Ser	6.0	7.5	6.8	7
Glu	8.2	8.5	8.4	8
Pro	6.2	9.0	7.6	7
Gly	9.6	11.0	10.3	10
Ala	14.7	11.0	12.8	12
Cys	1.8	1.8	1.8	2
Val	3.5	3.2	3.4	4
Met	0.7	0.7	0.7	1
Ile	4.0	3.4	3.7	4
Leu	8.4	8.9	8.7	9
Tyr	5.4	5.4	5.4	6
Phe	4.0	3.7	3.8	4
His	2.2	1.9	2.1	2
Lys	11.8	10.7	11.2	11
Me ₃ Lys	2.0	2.0	2.0	2
Arg	2.4	2.2	2.3	2
Trp	n.d.	n.d.	—	1

digestion of the same peptide. Residues 97 and 98 were shown to be glutamine and aspartic acid respectively, by carboxypeptidase-C digestion of peptide T11, and residue 98 was confirmed by Pronase digestion. Acid/amide positions in the haem region were determined by Pronase digestion of peptide C3C1 and the basic tryptic haem peptide (T4+T5). A single amide was located at residue 24, residue 29 being aspartic acid. The total number of amides is therefore eleven.

A partial split after tryptophan-67 was evidence of chymotryptic activity in the tryptic digest. A partial split after threonine-50 occurred in the chymotryptic digest.

Discussion

In the present investigation, the sequence of spinach cytochrome *c* has been deduced from analysis of peptide fragments by the dansyl-Edman technique (Gray & Hartley, 1963). The complete sequence, except for valine-13, was obtained from a single chymotryptic digest of 0.5 μ mol of cytochrome. The chymotryptic peptides were ordered by homology with other known sequences. Analyses of overlapping chymotryptic fragments resulting from partial splits (e.g. peptide C10a), were consistent with this ordering.

Peptides from the tryptic digest, also of 0.5 μ mol of cytochrome, were analysed with respect to their *N*-terminal amino acids, semi-quantitative amino acid compositions and mobilities. In addition, some peptides were subjected to dansyl-Edman degradation in order to confirm positions in the sequence known to vary between plant cytochromes. Residues 2, 63 and 109-113 were confirmed only by semi-quantitative amino acid analysis of peptides T1C1, T8 and T13. In each case, the peptides were short and present in high yield and the analyses presented no ambiguity of composition. Positions 8, 29, 58, 59 and 74 were each confirmed by two independent identifications by the dansyl-Edman method of either separate fractions of the same peptide (eluted from adjacent strips of the electrophoresis paper), or overlapping peptides [e.g. peptides C10 and C10(a)] from the chymotryptic digest. Residues 11 and 98 were confirmed by carboxypeptidase-C digestion of peptides T1C2 and T11. Valine-13, which was not seen during the dansyl-Edman analysis of peptide C2, possibly because of incomplete hydrolysis of Val-Gly, was confirmed in the tryptic digest by dansyl-Edman analyses of two separate fractions of the peptide T2. In each case, therefore, positions which are known to vary between plant cytochromes have been confirmed by at least two separate dansyl-Edman analyses. Other positions, where the supporting evidence is based on semi-quantitative methods and reliance on

Fig. 1. Amino acid sequence of spinach cytochrome *c*

---, Composition determined qualitatively; →, sequence determined by dansyl-Edman degradation; ←, sequence determined by using carboxypeptidase A; ⇌, amino acids determined by using carboxypeptidase C; *, acid/amide positions determined by Pronase digestion; †, 16h digestion; ↓↑, major enzymic cleavages; ↓↑, partial enzymic cleavages; ‡, chymotryptic cleavage of tryptic peptide. Abbreviations: Me₃Lys, ϵ -*N*-trimethyl-lysine; T, tryptic peptide; C, chymotryptic peptide. For full details of the individual peptides, see the Supplementary paper SUP 50013.

homology, are open to a degree of uncertainty, and there is the possibility that small deletions, or insertions could have been missed.

Positions in the sequence not previously observed in sequences of plant cytochromes *c* are lysine-11, valine-13, leucine-30, serine-53 and serine-68. Residue 51 has been threonine in all plant sequences published to date (Boulter *et al.*, 1972), although it is alanine also in leek (Brown & Boulter, 1973) and *Enteromorpha* (B. T. Meatyard, unpublished work). Residue 53 has been glycine in all eukaryotic mitochondrial cytochrome *c* sequences published to date. This decreases the number of 'invariant' residues to 28, the value of 34 (Boulter & Ramshaw, 1972) being decreased to 29 by the publication of the sequences of snail (Brown *et al.*, 1972) and *Crithidia* (Pettigrew, 1972) cytochromes *c*.

A detailed phylogenetic analysis of the sequences has been presented by Boulter *et al.* (1972). In that article, the change to leucine-30 was omitted from the information considered, but does not affect the topologies obtained.

We thank Mr. G. Bainbridge, Mrs. E. Innes, Mrs. L. Mayo and Mrs. L. Colpin, for invaluable technical assistance, and the Nuffield Foundation and the Science Research Council for financial support.

References

- Boulter, D. & Ramshaw, J. A. M. (1972) *Phytochemistry* **11**, 553-561
- Boulter, D., Ramshaw, J. A. M., Thompson, E. W., Richardson, M. & Brown, R. H. (1972) *Proc. Roy. Soc. Ser. B* **181**, 441-455
- Brown, R. H. & Boulter, D. (1973) *Biochem. J.* **131**, 247-251

- Brown, R. H., Richardson, M., Boulter, D., Ramshaw, J. A. M. & Jefferies, R. P. S. (1972) *Biochem. J.* **128**, 971-974
- Gray, W. R. & Hartley, B. S. (1963) *Biochem. J.* **89**, 378-380
- Pettigrew, C. H. (1972) *FEBS Lett.* **22**, 64-66
- Ramshaw, J. A. M., Richardson, M. & Boulter, D. (1971) *Eur. J. Biochem.* **23**, 475-483
- Richardson, M., Richardson, D. L., Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1971) *J. Biochem. (Tokyo)* **69**, 811-813
- Schmer, G. & Kreil, G. (1969) *Anal. Biochem.* **29**, 186-192
- Thompson, E. W., Laycock, M. V., Ramshaw, J. A. M. & Boulter, D. (1970) *Biochem. J.* **117**, 183-192
- Thompson, E. W., Richardson, M. & Boulter, D. (1971) *Biochem. J.* **121**, 439-446
- Tschesche, H. & Kupfer, S. (1972) *Eur. J. Biochem.* **26**, 33-36
- Woods, K. R. & Wang, K. T. (1967) *Biochim. Biophys. Acta* **133**, 369-370