The Amino Acid Sequences of Cytochrome *c* from Four Plant Sources

By RICHARD H. BROWN* and DONALD BOULTER

Department of Botany, University of Durham, Durham DH1 3LE, U.K.

(Received 3 July 1973)

Proposed amino acid sequences of cytochrome *c* from nasturtium (*Tropaeolum majus* L.), box-elder (*Acer negundo* L.), elder (*Sambucus nigra* L.) and parsnip (*Pastinaca sativa* L.) are presented. Because of the very limited amounts of cytochrome available from some plant sources, peptides derived from the cytochromes *c* have been sequenced by the semi-quantitative dansyl–Edman technique (Gray & Hartley, 1963) without supporting quantitative amino acid analyses. Because of the qualitative nature of the work, the sequences proposed must be regarded as tentative. Considerations of homology, although useful as a guide, have been kept to a minimum in the construction of sequences. Only the nasturtium sequence relies on considerations of homology for a complete ordering of the peptides. Where material permitted, each residue of a proposed sequence was determined at least once from both a tryptic and a chymotryptic peptide.

The comparison of amino acid sequences of specific proteins derived from different species has been of great value in studying the phylogenetic relationships among living organisms (Dayhoff, 1972). The protein studied most extensively in this connexion has been mitochondrial cytochrome *c*. A computer-generated phylogenetic tree for a range of higher plant species has been established (Boulter *et al.*, 1972) from the amino acid sequences of 15 plant cytochromes *c*.

Only limited amounts of cytochrome were available from these plant sources; peptides derived from the cytochromes *c* therefore were sequenced by the dansyl–Edman technique (Gray & Hartley, 1963). This method is applicable to the detection of amino acids at the nmol level (Brown & Boulter, 1973*b*), but has the disadvantage of being only semi-quantitative. The lack of material made it impractical to perform quantitative amino acid analyses of individual peptides, as was done previously (Thompson *et al.*, 1970); instead the semi-quantitative method of dansylation after hydrolysis was employed.

The sequences presented therefore do not have the reliability of those supported by quantitative measurements and the large amount of ‘redundant’ information that permits cross-checking of the results. Nevertheless, where possible, each residue of a presented sequence was determined from both a chymotryptic and a tryptic peptide, and a quantitative acid composition is presented for each total protein.

* Present address: Postgraduate School of Biological Sciences, University of Bradford, Bradford BD7 1DP, Yorkshire, U.K.

Materials and methods

Cytochrome *c* from nasturtium (*Tropaeolum majus* L.) and from box-elder (*Acer negundo* L.) were extracted from germinated seedlings. Seed was obtained from Hurst, Gunson and Tabor, Witham, Essex, U.K., and Barilli and Biagi, Bologna, Italy. Cytochrome *c* from elder (*Sambucus nigra* L.) was extracted from immature inflorescences collected locally. Cytochrome *c* from parsnip (*Pastinaca sativa* L.) was extracted from flowers and seed heads of parsnips grown in the University gardens.

The methods employed for the purification of each cytochrome and the sequence determinations of the purified cytochromes were those used previously in the sequence determination of *Nigella* cytochrome *c* (Brown & Boulter, 1973*b*).

Carboxypeptidase-C and Pronase digestions were carried out as described by Brown *et al.* (1973). Histidine, if suspected as C-terminal in a peptide, was confirmed as the bis-dansyl derivative by dansylation of the free amino acid remaining after the required number of Edman degradations of the peptide. If histidine was suspected in other positions, the α -dansyl-histidine product of the dansyl–Edman step was separated by paper electrophoresis at pH 1.9, re-dansylated and identified as the bis-dansyl derivative.

Results

Nasturtium cytochrome *c*

Fig. 1 shows the proposed sequence and summarizes the evidence for it. A single chymotryptic digest

Fig. 1. Proposed structure of cytochrome c from *nasturtium*, showing peptides isolated from tryptic (T) and chymotryptic (C) digestions

Peptides were sequenced by the dansyl-Edman procedure, with enzymic sub-digestions when necessary. A dashed line (----) indicates that a definite identification could not be made during the dansyl-Edman degradation. Residues 2, 3 and 67 were assigned from sequential digests of chymotryptic peptides with carboxypeptidase A. Amides were located by examining the mobilities of peptides or from digests of tryptic peptides with carboxypeptidase C (marked *). The acetyl (Ac) blocking group was identified as the 1-acetyl-2-dansyl-hydrazide. Any one of the lysine residues shown may be an arginine residue, or an arginine residue may be missing (see the text).

of 0.4 μ mol of cytochrome yielded peptides which, when ordered by homology with other known sequences, accounted for the complete sequence.

Each residue was positively identified in the dansyl-Edman analyses and could be confirmed by a positive identification in the peptide total hydrolyses. The

only exceptions were tryptophan-67, which was destroyed in the analysis, and residues 1-3 of the protein, which yielded a blocked *N*-terminal peptide, which was not amenable to dansyl-Edman analysis. These remaining residues were identified by the use of carboxypeptidase A.

A single tryptic digest of 0.2 μmol of cytochrome yielded less information than the chymotryptic digest, although most of the residues could be positively confirmed. The tryptic peptide 64-81 was combined with the chymotryptic peptide 64-75 before analysis of that part of the sequence. The peptides isolated from an enzymic sub-digest of the mixture therefore did not yield two independent sequence determinations, so that the results are presented only once in Fig. 1 as a chymotryptic peptide.

Overlaps between tryptic and chymotryptic peptides enabled some ordering of peptides in the sequence without recourse to considerations of homology. However, bonds between lysine-12 and alanine-13, lysine-63 and alanine-64, and between tyrosine-56 and serine-57 were hydrolysed in high yield by both enzymes. Evidence for these bonds, together with bonds between residues 67 and 68 and 73 and 74 from the combined tryptophan peptide

sub-digest (see above), is therefore from considerations of homology only.

The amino acid composition of the protein (Table 1) agrees with the proposed sequence except for the numbers of threonine, serine and arginine residues. Of these, arginine is the most suggestive of an error in the proposed sequence, in view of the difficulty of distinguishing Dns-arginine from ε-Dns-lysine on chromatograms. Therefore any one of the lysine residues shown in Fig. 1 may be an arginine residue, or, alternatively, an arginine residue may be missing from the sequence. The sequence Asn-Lys occurs both in positions 20-21 and positions 62-63 in the sequence. The evidence for the identification of residues 62-63 therefore is wholly from the chymotryptic digest.

Acer cytochrome c

Fig. 2 shows the proposed sequence and summarizes the evidence for it. Single digests of 0.5 μmol of cytochrome each with trypsin and chymotrypsin yielded peptides that accounted for the complete sequence. Each residue was positively identified in the dansyl-Edman analysis of at least one peptide

Table 1. Amino acid compositions of cytochrome c from nasturtium, elder, parsnip and Acer

Portions of each cytochrome were used for 24h and 72h hydrolyses in duplicate. Values for serine and threonine were extrapolated to zero time, assuming first-order kinetics of destruction. Values for valine and isoleucine are the maximal (72h) values. Values for other amino acids are the average of the 24h and 72h analyses. Tryptophan was not determined directly. The spectral ratios of the cytochromes indicate that one tryptophan residue is present for each cytochrome (Richardson *et al.*, 1970, 1971).

Amino acid composition (residues/molecule)								
	Nasturtium		Elder		Parsnip		Acer	
	Analysis	Sequence	Analysis	Sequence	Analysis	Sequence	Analysis	Sequence
Asp	12.8	13	11.8	12	12.0	12	12.2	12
Thr	6.6	6	7.0	7	6.9	7	11.8	7
Ser	6.0	5	5.5	5	3.9	4	12.3	4
Glu	8.4	8	9.4	9	6.5	6	9.7	9
Pro	6.0	6	8.0	8	7.5	7	12.7	8
Gly	11.4	11	11.1	11	12.4	12	14.9	11
Ala	15.3	15	12.3	12	14.3	14	21.0	12
Cys	1.1	2	1.4	2	1.6	2	1.1	2
Val	3.4	3	3.3	3	4.5	4	7.5	3
Met	1.1	1	2.1	2	1.0	1	2.1	2
Ile	3.2	3	3.1	3	3.4	3	6.8	3
Leu	9.4	9	8.0	8	9.4	10	9.8	8
Tyr	5.7	6	6.2	6	5.7	6	6.2	6
Phe	4.2	4	3.9	4	3.3	4	4.0	4
His	2.0	2	1.9	2	2.9	3	2.3	2
Lys	11.7	12	11.5	12	11.5	11	11.4	12
Me ₃ Lys	2.1	2	1.3	2	1.7	2	1.4	2
Arg	2.8	2	1.8	2	2.3	2	5.9	2
Trp	—	1	—	1	—	1	—	1

Fig. 3. Proposed structure of cytochrome c from elder showing peptides isolated from tryptic (T) and chymotryptic (C) digestions

and in most cases was identified in a peptide from both digests. As for nasturtium, the only residues not identified by dansyl-Edman analysis were residues

1-3 and 67, which were identified by the use of carboxypeptidase A.

Overlaps between tryptic and chymotryptic pep-

tides enabled the ordering of peptides in the sequence without recourse to considerations of homology. Bonds between tyrosine-56 and serine-57 and between

asparagine-70 and threonine-71 were hydrolysed in high yield by both enzymes, so there is no direct evidence for these bonds from overlaps. However,

Fig. 4. Proposed structure of cytochrome c from parsnip, showing peptides isolated from tryptic (T) and chymotryptic (C) digestions

Peptides were sequenced by the dansyl-Edman procedure, with enzymic sub-digestions when necessary. A dashed line (----) indicates that a definite identification could not be made during the dansyl-Edman degradation. Residues 2, 3 and 67 were assigned from sequential digests of chymotryptic peptides with carboxypeptidase A. Amides were located by examining the mobilities of peptides or from digests of tryptic peptides with carboxypeptidase C (marked *). The acetyl (Ac) blocking group was identified as the 1-acetyl-2-dansyl-hydrazide.

since only three fragments remained to be ordered, one being blocked at the *N*-terminus and one having *C*-terminal serine or alanine, they may be ordered without recourse to homology.

A total amino acid composition of the preparation used in this investigation is presented in Table 1. Most amino acids were present in much larger quantities than could be accounted for in the proposed sequence, except notably tyrosine, phenylalanine and histidine. It is probable that the preparation contained an impurity, which was not immediately obvious from its spectral ratio, which depends primarily on the absorption of aromatic amino acids at 280nm. No evidence was found for peptides that could not be placed in the presented sequence from either the tryptic or the chymotryptic digest.

Residue 112 is either alanine or serine; the cytochrome exists in two forms that differ in the residue in this position. The *C*-terminal peptides from both digests were obtained in high yield and gave a very clear indication of peptide polymorphism in the dansyl-Edman analysis. Alanine-112 was not a carry-over from an incomplete degradation at alanine-111, since the dansyl derivative without hydrolysis was obtained only at residue 112. The two forms were present in roughly equal amounts.

Elder cytochrome c

Fig. 3 shows the proposed sequence and summarizes the evidence for it. Single digests of 0.5 μ mol of cytochrome each with trypsin and chymotrypsin yielded peptides that could be ordered from overlaps to a sequence of 111 residues. Each residue was identified at least once by dansyl-Edman analysis except for residues 1-3 and 67, which were identified by the use of carboxypeptidase A (see results for nasturtium).

The amino acid composition of the protein (Table 1) is consistent with the proposed sequence.

Residue 24 gave some difficulty in the analysis, because during the Edman degradation of the tryptic haem peptides cyclization of glutamine-24 occurred after removal of asparagine-23. After opening of the pyroglutamyl ring by the method of Dekker *et al.* (1949), the degradation of the peptides proceeded normally.

Parsnip cytochrome c

Fig. 4 shows the proposed sequence and summarizes the evidence for it. Single digests of 0.5 μ mol of cytochrome each with trypsin and chymotrypsin yielded peptides that could be ordered from overlaps to a sequence of 111 residues. Each residue was identified at least once by dansyl-Edman analysis except for residues 1-3 and 67, which

were identified by the use of carboxypeptidase A (see results for nasturtium).

The amino acid composition of the protein (Table 1) is consistent with the proposed sequence except that the analysis for leucine is suggestive of one residue fewer than the value calculated from the sequence. However, all leucine residue in the sequence were identified from both digests, in particular the 'extra' ones (compared with other plant sequences) at residues 30 and 66.

Discussion

Nasturtium cytochrome *c* has a number of positions in the sequence at which substitutions occur that have not previously been observed in plant cytochromes *c*. These are asparagine-11, aspartic acid-15, asparagine-20, leucine-66 and alanine-70. All these changes occur in positions that have been seen to vary previously (Boulter *et al.*, 1972; Brown & Boulter, 1973a).

A change in a residue that has been previously unvaried in plant sequences has been observed in position 23 of the elder sequence. Asparagine replaces alanine in this position.

The *Acer* sequence contains three positions in which the amino acids found differ from those previously observed in these positions in plant sequences. Residue 12 is alanine, observed also in sequences of *Enteromorpha* (B. T. Meatyard, unpublished work) and *Nigella* (Brown & Boulter, 1973b). Residue 112 is an additional residue compared with most plant sequences and is either serine or alanine (see above). The only other plant cytochromes with more than 111 residues are wheat (Stevens *et al.*, 1967), barley (D. L. Richardson, unpublished work) and *Ginkgo* (Ramshaw *et al.*, 1971). Residue 69 is tyrosine; previously this position has been an unvaried glutamic acid, although a further change occurs in the parsnip sequence, where the residue is aspartic acid or asparagine.

The sequence Lys-Asp-Val (residues 11-13) is common to the parsnip and spinach sequences (Brown *et al.*, 1973), as also is leucine-30. Leucine-66 of the parsnip sequence occurs also in nasturtium cytochrome *c*. Glycine-15, alanine-68 and histidine-108 have not previously been observed in these positions in sequences of plant cytochromes.

The phylogenetic implications of the elder sequence have been presented by the authors (Boulter *et al.*, 1972).

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

References

- 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. (1973a) *Biochem. J.* **131**, 247-251
- Brown, R. H. & Boulter, D. (1973b) *Biochem. J.* **133**, 251-254
- Brown, R. H., Richardson, M., Scogin, R. & Boulter, D. (1973) *Biochem. J.* **131**, 253-256
- Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, vol. 5, National Biomedical Research Foundation, Silver Spring, Md.
- Dekker, C. A., Stone, D. & Fruton, J. S. (1949) *J. Biol. Chem.* **181**, 719-729
- Gray, W. R. & Hartley, B. S. (1963) *Biochem. J.* **89**, 379-380
- Ramshaw, J. A. M., Richardson, M. & Boulter, D. (1971) *Eur. J. Biochem.* **23**, 475-483
- Richardson, M., Laycock, M. V., Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1970) *Phytochemistry* **9**, 2271-2280
- Richardson, M., Richardson, D. L., Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1971) *J. Biochem. (Tokyo)* **69**, 811-813
- Stevens, F. C., Glazer, A. N. & Smith, E. L. (1967) *J. Biol. Chem.* **242**, 2764-2779
- Thompson, E. W., Laycock, M. V., Ramshaw, J. A. M. & Boulter, D. (1970) *Biochem. J.* **117**, 183-192