The Amino Acid Sequence of Cytochrome *c* from *Nigella damascena* L. (Love-in-a-Mist)

By R. H. BROWN* and D. BOULTER

Department of Botany, University of Durham, Durham City, U.K.

(Received 17 October 1972)

The amino acid sequence of cytochrome *c* from *Nigella damascena* L. was determined on $0.2\mu\text{mol}$ of protein. Peptides from a single chymotryptic digest were analysed by the dansyl-Edman procedure. These peptides were ordered by reference to the sequences of other plant cytochromes *c*, assuming that the *Nigella* cytochrome is homologous with the other cytochromes. Many of the *Nigella* peptides were one or two residues short when compared with the corresponding chymotryptic peptides from other plant cytochromes *c*. These residues are assumed to have been removed by an endogenous carboxypeptidase, and the identification and placing of these residues is entirely based on homology. These residues are numbered 3, 18, 42, 43, 44, 54, 67, 72, 73, 82 and 105. A number of other positions are almost entirely placed by homology. These are positions which could not be placed definitely by dansyl-Edman analysis or by dansylation after digestion with carboxypeptidase A, and are numbered 14, 15, 16, 39, 40, 85, 86, 87 and 88. Except for residue 15, all residues based entirely, or nearly so, on homology have been previously found invariant in sequences of plant cytochromes *c*. Experimental details are given in a supplementary paper that has been deposited as Supplementary Publication SUP 50017, 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.

The amino acid sequences of a large number of cytochromes *c* have now been determined, including sequences of 20 plant species (Boulter *et al.*, 1972). The determination of the sequence of *Nigella* cytochrome *c* forms part of a continuation of an investigation into plant phylogeny based on protein sequence studies.

Experimental

Materials

Seeds of *Nigella damascena* L. were obtained from Hurst, Gunson and Tabor, Witham, Essex, U.K. Other materials were as described by Thompson *et al.* (1971).

Methods

The isolation and purification of cytochrome *c* from *Nigella* followed the method previously used for seeding material (Richardson *et al.*, 1971). This method yielded approx. 12mg of cytochrome *c* from 100kg of dry seed, but the spectral ratios of the pre-

paration were very poor [E_{410} (oxidized)/ E_{280} (oxidized) = 1]. The material was considered insufficiently pure for the sequence analysis, and was further purified.

The material of spectral ratio 1 was applied to a column (3.1cm $\times$ 80cm) of Sephadex G-75 equilibrated in and eluted with 10mM- NaH_2PO_4 -NaOH buffer, pH 7.2, containing 1M-NaCl. Fractions containing cytochrome *c* were combined and dialysed against phosphate buffer from which NaCl was omitted. The cytochrome *c* solution was then applied to a column (1cm $\times$ 5cm) of Whatman CM-52 CM-cellulose equilibrated with 10mM- NaH_2PO_4 -NaOH buffer, pH 7.2, and the cytochrome *c* was oxidized on the column by application of the minimum volume of very dilute $\text{K}_3\text{Fe}(\text{CN})_6$ solution. The ferricytochrome *c* was then eluted from the column by an increasing pH gradient from 10mM- NaH_2PO_4 -NaOH buffer, pH 7.2, to 10mM- $\text{Na}_2\text{P}_2\text{O}_7$ -NaOH buffer, pH 11. The ferricytochrome *c* emerged from the column at pH 10.5 and had a spectral ratio E_{410} (oxidized)/ E_{280} (oxidized) of 4.0. The final yield (desalted) was 2.5mg.

The cytochrome *c* was digested with chymotrypsin (for 4h) as described previously (Thompson *et al.*, 1971). The resultant peptides were purified by electrophoresis on paper (Thompson *et al.*, 1970). The amino acid sequences of the peptides were

* Present address: Postgraduate School of Studies in Applied Biology, University of Bradford, Bradford, Yorks. BD7 1DP, U.K.

determined by the dansyl-Edman method (Gray & Hartley, 1963) as described previously (Thompson *et al.*, 1970). Dns-amino acids were identified by chromatography on polyamide sheets (Woods & Wang, 1967) with the solvent systems described by Ramshaw *et al.* (1970). The haem group was removed from haem peptides by performic acid oxidation (Ramshaw *et al.*, 1970). Certain peptides were further digested with chymotrypsin or carboxypeptidase A, as described previously (Thompson *et al.*, 1970). Released amino acids were identified as the dansyl derivatives. The *N*-terminal blocking group (acetyl) was identified as the 1-acyl-2-dansyl hydrazide as described previously (Thompson *et al.*, 1970). The quantitative method of amino acid analysis used previously for peptides (Thompson *et al.*, 1970) was used only for the complete protein. Instead, a semi-quantitative method of dansylation after hydrolysis (Thompson *et al.*, 1970) was used.

For the identification of residues in minor or long peptides, the dansyl-Edman method described by Thompson *et al.* (1970) was modified. *N*-Terminal amino acids were identified from one-tenth of the

usual portion, i.e. 1–5 nmol, the Dns-amino acids being chromatographed on polyamide sheets (5 cm × 5 cm).

Results

The amino acid compositions of *Nigella* cytochrome *c*, determined by analysis and calculated from the sequence, are shown in Table 1. These values are in reasonable agreement, although values obtained by analysis for glycine, alanine and lysine differ by approximately one residue from the sequence values. Fig. 1 gives the complete sequence of *Nigella* cytochrome *c*, together with an indication of the chymotryptic peptides identified.

The sequence was deduced from a single chymotryptic digest, homology with other plant cytochromes being assumed to place the peptides in order. In the sequence studies on chymotryptic peptides from *Nigella* cytochrome *c*, many of the peptides were one or two residues short when compared with the corresponding chymotryptic peptides from other plant cytochromes *c*. This shortening was ascribed to

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

Tryptophan was not determined. The best spectral ratios of the cytochrome indicate that one tryptophan residue is present (Richardson *et al.*, 1971).

	Amino acid composition (residues/mol)			
	Single hydrolysis of 0.1 mg		Average	Sequence
	20h	70h		
Asp	9.5	10.3	9.9	10
Thr	5.9	5.5	6.1*	6
Ser	6.9	6.2	7.1*	7
Glu	11.1	11.9	11.5	11
Pro	6.0	5.6	5.8	6
Gly	12.2	12.1	12.2	11
Ala	12.4	11.4	11.9	13
Cys	2.5	2.9	2.7	2
Val	4.4	4.4	4.4	4
Met	1.4	1.4	1.4	1
Ile	1.5	2.0	2.0†	2
Leu	8.9	8.7	8.8	9
Tyr	5.2	6.2	5.7	6
Phe	4.0	4.0	4.0	4
His	2.7	2.4	2.5	3
Lys	13.1	12.0	12.6	11
Me ₃ Lys	1.6	1.2	1.4	2
Arg	2.0	1.9	2.0	2
Trp				1
	Total ...			111

* Extrapolated to zero time, assuming first-order kinetics of destruction.

† Maximal value.

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

Chymotryptic peptides are indicated. —, Residues determined by dansyl-Edman analysis. Fragments assigned to positions 3, 18, 42–44, 54, 67, 72, 73, 82 and 105 are placed solely by homology. —, Residues determined or confirmed by digestion with carboxypeptidase A; *, 16h digest. —, Residues determined qualitatively and, except for position 1, ordered by homology. ↑, Major cleavage; ↑, partial cleavage; ↑, secondary chymotryptic cleavage after removal of haem; X, suspected positions of attack of assumed endogenous carboxypeptidase activity. Abbreviation: Me₃Lys, ϵ -*N*-trimethyl-lysine. For full details of the individual peptides see the Supplementary Paper SUP 50017.

the activity of a carboxypeptidase of specificity similar to that of carboxypeptidase A. Thus phenylalanine and tryptophan appeared to have been removed from all peptides containing one of these amino acids as C-terminus; tyrosine appeared to have been removed from Gly-Tyr, Leu-Tyr and Ala-Tyr sequences, but not from Asp-Tyr or Ser-Tyr; leucine appeared to have been removed from Gly-Leu-Phe and Thr-Leu-Tyr sequences. The removed amino acids were recovered in the neutral zone of the pH6.5 electrophoresis of the digest. In each case the amino acids were removed from positions in the sequence that had been previously found invariant in all plant cytochrome *c* sequences.

Other amino acid positions were identified by

dansyl-Edman analysis of peptides, with the following exceptions.

Residue 1, alanine, in peptide 1, acetyl-Ala-Ser, was identified from the total hydrolysis of the peptide.

Residues 14–16 (peptide 3), 39–41 (peptide 6) and 85–88 (peptide 22) were placed from the total hydrolysis of the peptides and homology with other known sequences. None of these positions has to date been found to vary in plant cytochromes *c* whose sequences have been determined, except for residue 15, which is unvaried except in sunflower (Ramshaw *et al.*, 1970), and residue 41, which has been asparagine in all previously observed plant cytochrome *c* sequences except that of *Ginkgo*, where it is also histidine (Ramshaw *et al.*, 1971). Residue 41,

although not identified by dansyl-Edman analysis, was confirmed by carboxypeptidase A digestion of peptide 6.

The peptide following peptide 6 (Gly-Leu-Phe) appeared to have been completely digested by the presumed carboxypeptidase activity during the chymotryptic digestion. The sequence at these positions was assigned by homology. (Leucine and phenylalanine were positively identified in the neutral zone of the pH 6.5 electrophoresis, but glycine could not be distinguished from traces of the amino acid normally eluted from the electrophoresis paper.)

The electrophoretic mobilities of the peptides at pH 6.5 indicated the presence of 11 amides, of which seven could be placed unambiguously. Single unplaced amide positions occurred in peptide 3 (position 3, 4, 9 or 15) and peptide 24 (97, 98 or 101). Two unplaced amides occurred in peptide 5 (positions 24, 29 or 30).

Discussion

In the present investigation the amino acid sequence of *Nigella* cytochrome *c* has been deduced from analysis of peptide fragments by the dansyl-Edman technique (Gray & Hartley, 1963). Peptides from a single chymotryptic digest of 2.4 mg of protein (and some free amino acids liberated by an endogenous carboxypeptidase during the digestion), were ordered by homology with other known plant sequences.

Variations at positions in the sequence not previously observed in sequences of plant cytochromes *c* were glutamine/glutamic acid-30, valine-51 and leucine-103. Residue 51 has been threonine in most plant sequences published to date, although it is alanine in spinach (Brown *et al.*, 1973) and leek (Brown & Boulter, 1973). Residue 103 has been isoleucine in all plant sequences published to date and in all other eukaryotic cytochromes *c* except tuna fish and *Candida* (Dayhoff, 1969) and bonito (Nakayama *et al.*, 1971), where the residue is valine.

There was insufficient material for quantitative peptide analyses or for a tryptic digest. As most of the

residues have been identified by a single dansyl-Edman analysis and have been ordered with considerable reliance on homology, the possibility arises that certain residues may have been incorrectly assigned. This possibility is supported by inconsistencies in the total amino acid compositions determined by analysis and by calculation from the published sequence. These inconsistencies may, however, arise from impurities in the preparation, as spectral ratios for plant cytochromes *c* purified previously (Richardson *et al.*, 1971) have been generally 10–20% higher.

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., 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., Scogin, R. & Boulter, D. (1973) *Biochem. J.* **131**, 253–256
- Dayhoff, M. O. (1969) *Atlas of Protein Sequence and Structure*, vol. 4, National Biomedical Research Foundation, Silver Spring, Md.
- Gray, W. R. & Hartley, B. S. (1963) *Biochem. J.* **89**, 379–380
- Nakayama, T., Titani, K. & Narita, K. (1971) *J. Biochem. (Tokyo)* **70**, 311–326
- Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1970) *Biochem. J.* **119**, 535–539
- Ramshaw, J. A. M., Richardson, M. & Boulter, D. (1971) *Eur. J. Biochem.* **23**, 475–483
- Richardson, M., Richardson, D., Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1971) *J. Biochem. (Tokyo)* **69**, 811–813
- 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
- Woods, W. R. & Wang, K. T. (1967) *Biochim. Biophys. Acta* **133**, 369–370