The Amino Acid Sequence of Cytochrome *c* from *Allium porrum* L. (Leek)

By R. H. BROWN and D. BOULTER

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

(Received 20 July 1972)

The amino acid sequence of leek cytochrome *c* was determined with 0.4 μ mol of protein. The sequence was deduced solely from a chymotryptic digest. The cytochrome was homologous with other plant cytochromes *c* of mitochondrial origin. Leek cytochrome *c* has an *N*-acetylated 'tail' as compared with mammalian cytochrome *c*, and two residues of ϵ -*N*-trimethyl-lysine. Unlike other plant cytochromes *c*, leek cytochrome *c* has glutamic acid or glutamine in position 11, leucine in position 20 and alanine in position 51. 4-Hydroxyproline partially substitutes for proline in position 79. Experimental details are given in a supplementary paper that has been deposited as Supplementary Publication SUP 50012, 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).

Young mature tissue yielded lower quantities of cytochrome *c* than etiolated seedlings, the former giving approximately one-tenth of the yield of cytochrome compared to the latter, per given weight of fresh wet material. It was of interest therefore, to see if the investigation could be completed on less than 0.5 μ mol of protein by limiting the study to a single chymotryptic digest. It was considered that the plant sequences determined to date were homologous to an extent that would enable all the chymotryptic peptides to be placed in sequence without the necessity for overlapping tryptic peptides.

Materials

Approximately 100 kg of leeks that had been grown from seed to marketable size locally, were generously supplied by Mr. D. Sayers. A further 100 kg of leeks were obtained from Crawfords, Durham City, Durham, U.K. The leeks from both sources were deep frozen at -20°C until required. Other materials were as described by Thompson *et al.* (1971a).

Methods

The isolation and purification of the cytochrome *c* from leek were carried out essentially as described by Richardson *et al.* (1971). The leeks were hand-chopped to small pieces before homogenizing them in the Waring Blender. The filtration step at pH 4.5

employed for seedling material was omitted. From 200 kg of starting material, 30 mg of crude cytochrome was obtained after the concentration step on CM-Sephadex. The final yield of material of spectral ratio $E_{410}(\text{oxidized})/E_{280}(\text{oxidized}) = 4.0$ was 5 mg.

The purified cytochrome was digested with chymotrypsin (4h) as described previously (Thompson *et al.*, 1971a). 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). Dns-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 the haem peptide with performic acid (Ramshaw *et al.*, 1971). The *N*-terminal blocking group was identified by the method of Schmer & Kreil (1969) and the 1-acyl-2-dansyl-hydrazine derivative identified by chromatography on polyamide sheets (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). Certain peptides were digested with carboxypeptidase A (Thompson *et al.*, 1970; 1h except as stated otherwise) and released amino acids were identified as the Dns derivatives.

Results

The amino acid compositions of leek cytochrome *c* determined by analysis and calculated from the

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

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

	Amino acid composition (residues/molecule)			Sequence
	24h	72h	Average	
Asp	9.5	9.6	9.5	10
Thr	6.7	6.5	6.8*	6
Ser	6.1	5.6	6.3*	5
Glu	10.1	9.9	10.0	11
Pro	7.6	7.4	7.5	7
Gly	11.0	11.3	11.2	11
Ala	12.7	12.8	12.7	13
Cys	2.6	2.6	2.6	2
Val	4.9	4.9	4.9	4
Met	1.7	1.9	1.8	2
Ile	3.7	3.7	3.7	3
Leu	8.5	8.3	8.4	9
Tyr	5.5	5.4	5.4	6
Phe	3.6	3.5	3.5	4
His	1.8	1.7	1.7	2
Lys	10.5	11.7	11.1	11
Me ₃ Lys	1.6	1.7	1.6	2
Arg	2.3	2.0	2.2	2
Trp	n.d.	n.d.		1
			Total	111

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

sequence, are shown in Table 1. These values are in reasonable agreement, except for threonine, serine, glutamic acid, isoleucine and valine. Fig. 1 gives the complete sequence of leek cytochrome *c*, together with an indication of the chymotryptic peptides identified. The evidence for assigning the sequence of the residues is given. The results of the dansyl-Edman analysis of peptides are given in Table 2.

Consideration of the mobilities of the non-haem peptides indicate that there are eleven amide groups. Of these, eight can be placed unambiguously, but the remainder cannot be placed on present evidence. Unplaced amide positions are at residues 5, 10 or 11 (two amides) and 68, 69 or 70 (one amide). An unexpected chymotryptic break at glutamine-97 enabled positions 97 and 98 to be placed unambiguously as glutamine and aspartic acid respectively.

Comparison of the mobilities of the chymotryptic haem peptides of leek and horse heart (Margoliash *et al.*, 1961) reveals that the leek peptide has a lower charge. A single amide position occurs therefore, at residue 24 or 29. The total number of amides is twelve.

An unusual chymotryptic split was observed in the cleavage of the alanine-glycine bond in positions 13 and 14. A change in position 51, a previously unvaried position in plants, from threonine to alanine introduced a second unusual split for chymotrypsin. A third unusual split after glutamine-97 was also observed. There was no evidence of tryptic-type activity in the digest.

Table 2. *Sequence evidence for chymotryptic peptides of leek cytochrome c*

Amino acid compositions of peptides were determined semi-quantitatively by total hydrolysis, followed by preparation and separation of the Dns-amino acids. Mobilities at pH6.5 are given relative to Dns-Arg-Arg and at pH1.9, relative to Dns-arginine. Forward arrows indicate positions confirmed by dansyl-Edman degradation; reverse arrows indicate positions confirmed by carboxypeptidase-A digestion. The mobility at pH6.5 was estimated to within ± 0.1 .

Peptide	Mobility	
	pH 6.5	pH 1.9
1 Acetyl(Ala) [*] Thr-Phe ←←	-1.3	0
2 Ser-Glx-Ala-Pro-Pro-Gly-Asx-Glx-Lys-Ala →→→→→→→→	0	0.72
2(a)† Ser-Glx-Ala-Pro-Pro-Gly-Asx-Glx-Lys-Ala →→→→→→→→	-0.7	0.60
3 Gly-Glx(Lys)Ile-Phe →→→	0	0.85
4 Lys-Leu →→	1.5	1.12
5 Lys(Cys,Ala,Glx,Cys,His,Thr,Val,Glx,Lys,Gly,Ala,Gly,His) →	0.3	

Table 2—continued

		Mobility	
		pH 6.5	pH 1.9
	Haem group removed, digested with chymotrypsin		
5C1	$\begin{array}{c} \text{O}_3\text{H} \quad \text{O}_3\text{H} \\ \quad \\ \text{Lys}-\text{Cys}-\text{Ala}-\text{Glx}-\text{Cys}-\text{His} \\ \hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \end{array}$		0.75
5C2	$\text{Thr}-\text{Val}-\text{Glx}-\text{Lys}-\text{Gly}-\text{Ala}-\text{Gly}-\text{His}$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$		1.05
6	$\text{Lys}-\text{Glx}-\text{Gly}-\text{Pro}(\text{Asx}, \text{Leu}, \text{Asx})$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	0.9	0.75
7	$\text{Leu}-\text{Asx}(\text{Gly}, \text{Leu}, \text{Phe})$ $\hline \text{---} \quad \text{---}$	0	0.46
7(a)	$\text{Leu}-\text{Asx}$ $\hline \text{---}$	0	0.72
7(b)	$\text{Gly}-\text{Leu}-\text{Phe}$ $\hline \text{---} \quad \text{---} \quad \text{---}$	0	0.63
8	$\text{Gly}-\text{Arg}-\text{Glx}-\text{Ser}-\text{Gly}-\text{Thr}$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	1.0	0.85
9	$\text{Ala}-\text{Ala}-\text{Gly}-\text{Tyr}$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	0	0.57
10	$\text{Ser}-\text{Tyr}$ $\hline \text{---} \quad \text{---}$	0	0.63
11	$\text{Ser}-\text{Ala}-\text{Ala}-\text{Asx}-\text{Lys}-\text{Asn}$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	1.0	0.85
11(a)	$\text{Ser}-\text{Ala}-\text{Ala}-\text{Asn}$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	0	0.63
11(b)	$\text{Lys}(\text{Asx})$ $\hline \text{---}$	1.8	1.27
12	$\text{Met}(\text{Ala}, \text{Val}, \text{Val}, \text{Trp})$ $\hline \text{---}$	0	—
12(b)	$\text{Ala}-\text{Val}-\text{Val}-\text{Trp}$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	0	—
13	$\text{Glx}-\text{Glx}-\text{Asx}-\text{Thr}-\text{Leu}-\text{Tyr}$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	-1.3	0.30
14	$\text{Asx}-\text{Tyr}-\text{Leu}$ $\hline \text{---} \quad \text{---} \quad \text{---}$	-1.3	0.30
14(a)	$\text{Asx}-\text{Tyr}$ $\hline \text{---} \quad \text{---}$	-1.3	0.50
15	$\text{Leu}-\text{Leu}-\text{Asx}-\text{Pro}(\text{Me}_3\text{Lys}, \text{Lys}, \text{Tyr})$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	1.6	0.93
15(b)	$\text{Leu}-\text{Asx}-\text{Pro}(\text{Me}_3\text{Lys}, \text{Lys}, \text{Tyr})$ $\hline \text{---} \quad \text{---} \quad \text{---}$	1.8	1.00
16	$\text{Ile}-\text{Pro}-\text{Gly}-\text{Thr}-\text{Lys}-\text{Met}$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	1.0	0.85
17	$\text{Val}-\text{Phe}-\text{Pro}-\text{Gly}-\text{Leu}$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	0	0.52
18	$\text{Me}_3\text{Lys}-\text{Lys}-\text{Pro}-\text{Gln}$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	2.3	1.35
19	$\text{Asx}-\text{Arg}-\text{Ala}-\text{Asx}-\text{Leu}$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	1.2	0.75
20	$\text{Ile}-\text{Ala}-\text{Tyr}$ $\hline \text{---} \quad \text{---} \quad \text{---}$	0	0.59
21	$\text{Leu}-\text{Lys}-\text{Glx}-\text{Ser}-\text{Thr}-\text{Ala}$ $\hline \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---}$	0	0.83

* Digestion with carboxypeptidase for 16h.

† Deamidated form of peptide 2.

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

Chymotryptic peptides are indicated by the designations under the sequence. ---, Composition determined qualitatively; —, sequence determined by dansyl-Edman degradation; —, sequence determined by using carboxypeptidase A (*, 16h digestion); †, major chymotryptic cleavage; ‡, partial cleavage; †, secondary chymotryptic cleavage after removal of haem; Me₃Lys, ϵ -*N*-trimethyl-lysine. For full details of the individual peptides see the Supplementary Paper SUP 50012.

Discussion

In the present investigation, the sequence of leek cytochrome *c* has been deduced from analyses of peptide fragments from a single chymotryptic digest. There was insufficient material for confirmatory quantitative amino acid analyses to be made on each peptide. The semi-quantitative analyses that were performed served mainly for the preliminary identification of peptides and cannot in general be relied upon as supporting evidence for the sequence. There was also insufficient material for a tryptic digest to be performed, so that the chymotryptic peptides were aligned by homology with other known plant cytochrome *c* sequences (Boulter *et al.*, 1972). As most of the residues have been identified on the evidence of a single dansyl-Edman analysis, the possibility arises that certain residues may have been incorrectly assigned, particularly in view of certain 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 purified previously (Richardson *et al.*, 1971, except *Arum*) have been 10–20% higher. The reliance on homology for ordering the peptides in the sequence introduces the possibility that deletions or insertions in the sequence

might have been missed. However, the high degree of homology between plant cytochromes *c* (Boulter *et al.*, 1972) makes these possibilities less likely.

The unusual splits observed in the chymotryptic digest are almost certainly non-specific cleavages. The longer time of digestion compared with previous investigations (normally 2h) would favour such cleavages. The break at threonine-alanine (residues 50–51) has also been observed in a 4h digestion of spinach cytochrome *c* (Brown *et al.*, 1973) and a break at the same position has been observed in a 4h digestion of *Nigella* cytochrome *c*, where the sequence is Thr-Val (R. H. Brown, unpublished work).

Positions in the leek sequence not previously observed in the sequences of plant cytochromes *c*, are glutamine (or glutamic acid)-11, leucine-20 and alanine-51. Position 11 is highly variable (Boulter & Ramshaw, 1972), but positions 20 and 51 have not been found to vary before in plants, although they do vary when all previously observed sequences are considered.

Position 79 is partially substituted by 4-hydroxyproline, in common with cytochromes *c* from castor, *Abutilon* and cotton (Thompson *et al.*, 1971a,b). The sequence also contains two residues of ϵ -*N*-trimethyl-lysine, in common with all other plant cytochromes *c*.

Partial deamidation at positions 10 or 11 during the

purification of the cytochrome was demonstrated by the presence of both peptides 2 and 2(a), having the same sequence but differing by a unit charge. Only one sequence of cytochrome *c* from a monocotyledon has been previously published; that of wheat (Stevens *et al.*, 1967). A detailed phylogenetic analysis of cytochrome sequences will therefore, have to wait further sequences of monocotyledon cytochromes *c*.

We thank Mr. G. Bainbridge, Mrs. E. Innes, Dr. M. E. Boulter and Mrs. L. Mayo, 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., Richardson, M., Scogin, R. & Boulter, D. (1973) *Biochem. J.* **131**, 253–256
- Gray, W. R. & Hartley, B. S. (1963) *Biochem. J.* **89**, 379–380
- Margoliash, E., Smith, E. L., Kreil, G. & Tuppy, H. (1961) *Nature (London)* **192**, 1121–1127
- 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
- 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
- Thompson, E. W., Richardson, M. & Boulter, D. (1971a) *Biochem. J.* **121**, 439–446
- Thompson, E. W., Notton, B. A., Richardson, M. & Boulter, D. (1971b) *Biochem. J.* **124**, 787–791
- Woods, W. R. & Wang, K. T. (1967) *Biochim. Biophys. Acta* **133**, 369–370