Biochem. J. (1970) **119**, 535–539
Printed in Great Britain

The Amino Acid Sequence of *Helianthus annuus* L. (Sunflower) Cytochrome *c* Deduced from Chymotryptic Peptides

By J. A. M. RAMSHAW, E. W. THOMPSON and D. BOULTER
Department of Botany, University of Durham, South Road, Durham City, U.K.

(*Received* 30 *April* 1970)

Peptides derived from digestion of 1 μmol of sunflower cytochrome *c* with chymotrypsin were separated by paper electrophoresis. The sequences of these peptides were determined by using the dansyl–Edman method (Gray & Hartley, 1963) and confirmed by analysis of their amino acid composition. Comparison of the set of peptides with the chymotryptic peptides of mung-bean (Thompson, Laycock, Ramshaw & Boulter, 1970) and wheat germ (Stevens, Glazer & Smith, 1967) cytochrome *c* shows a clear homology. The complete sequence of sunflower cytochrome *c* was established by alignment of the sunflower peptides with the sequences of mung bean cytochrome *c* and wheat germ cytochrome *c*.

Protein sequence determinations on homologous sets of proteins from various species have enabled phylogenetic trees to be calculated that are in close agreement with those derived from fossil, morphological and cytological evidence (Dayhoff, 1969). The determination of the sequence of sunflower cytochrome *c* in this present investigation is part of a study into plant phylogeny using cytochrome *c* sequence determination.

MATERIALS

The sunflower seeds were obtained from the Tyneside Seed Co., Gateshead, Co. Durham, U.K. Chymotrypsin (three times recrystallized), papain (twice recrystallized) and trypsin (salt free, twice recrystallized) were obtained from the Worthington Biochemical Corp., Freehold, N.J., U.S.A. Carboxypeptidase A-phosphorofluoridate (crystalline suspension in water), was obtained from the Sigma Chemical Co., St Louis, Mo., U.S.A. Other chemicals used were of the highest purity readily available. The polyamide sheets were obtained from the Cheng Chin Trading Co. Ltd., Taipei, Taiwan.

METHODS

Preparation of cytochrome c. This is described by Richardson, Laycock, Ramshaw, Thompson & Boulter (1970). The E_{550} (reduced)/E_{280} (oxidized) ratio was 1.00 for the material used in this investigation.

Digestion with chymotrypsin. 1 μmol of cytochrome *c* was denatured with ethanol (Margoliash, Kimmel, Hill & Schmidt, 1962) but without addition of 5M-$(NH_4)_2SO_4$ and final dialysis. It was suspended in 2.5ml of water and titrated to pH8.0 with 0.025M-NaOH on a Radiometer TTT 1c autotitrator. Digestion was performed at 40°C by addition at zero time of 2% (w/w, enzyme: substrate) of α-chymotrypsin made up in water and a further 2% after 100min. The digestion was terminated after 150min by freezing.

Purification of peptides. Peptides purified by high-voltage paper electrophoresis at pH6.5 and 1.9 were detected and their mobilities were calculated as described by Thompson *et al.* (1970).

Amino acid sequence methods. The peptide sequences were determined by using the dansyl–Edman method (Gray & Hartley, 1963) as described by Thompson *et al.* (1970). Separation of dansyl-amino acids was on polyamide sheets (Woods & Wang, 1967). A new third dimension was adopted, running in the same direction as the second dimension with butyl acetate–methanol–acetic acid (30:20:1, by vol.). This enabled the complete separation of DNS-arginine, DNS-ϵ-*N*-trimethyl-lysine and ϵ-DNS-lysine. In this system, the separations of DNS-serine from DNS-threonine and of DNS-aspartic acid, DNS-glutamic acid and *O*-DNS-tyrosine were also better than those previously described (Frangione & Milstein, 1968; Thompson *et al.* 1970).

Removal of the haem moiety for sequence analysis. The haem peptide was taken up in 200 μl of 90% (v/v) formic acid and three 10 μl portions of 30% (w/v) H_2O_2 were added at 10min intervals to performic oxidize the thioether link (Nolan & Margoliash, 1966). The reaction proceeded at room temperature and was stopped after 30min by freezing and freeze-drying; the red colour had been completely discharged.

Enzymic digestion of peptides. Certain peptides were digested with 5% (w/w) trypsin in pH8.4, 0.2M-NH_4HCO_3 at 37°C for 60min. The peptide for papain digestion was taken up in 200 μl of 0.01M-sodium phosphate buffer, pH6.5, containing 20 μg of papain and 0.05% of 2,3-dimercaptopropan-1-ol (Smith & Kimmel, 1960). Digestion was at 37°C and was terminated after 200min by freezing and freeze-drying. Digestion by carboxypeptidase A-phosphorofluoridate was by using 0.2–0.5mg of enzyme per mol of peptide in 0.2M-NH_4HCO_3 buffer,

pH 8.4, at 37°C. The amino acids liberated were detected by the dansyl method.

Detection of acetyl group. The method used for the detection of the acetyl group was based on that of Schmer & Kreil (1969), but scaled down to the amount of peptide available. Separation of the 1-acyl-2-dansylhydrazine product was on polyamide sheets by using the same solvent system as was used to separate the dansyl-amino acids.

Peptide amino acid compositions. These were carried out as described by Thompson *et al.* (1970).

Nomenclature of peptides. Peptides are numbered on the basis of their occurrence in the deduced sequence starting at the *N*-terminus. Peptides derived by further cleavage are given by a subscript to the parent peptide. Residue numbers used refer to Fig. 1.

RESULTS

The sunflower cytochrome *c* (1 μmol, measured spectrophotometrically) precipitated immediately on denaturation and was readily digested with α-chymotrypsin. The results of the dansyl–Edman analysis and the electrophoretic mobilities of the following peptides are summarized in Table 1. The electrophoretic mobilities were used to determine the amide content of the peptides (Offord, 1966). The results of the quantitative amino acid analyses are given in Table 2.

Peptide C1 (acetyl-Ala-Ser-Phe). This was ninhydrin negative, but total analysis by the dansyl method gave the composition as (Ala, Ser, Phe). Carboxypeptidase A digestion released phenylalanine and traces of serine after 3h and equal amounts of phenylalanine and serine after 18h. The *N*-terminal block was determined as an acetyl group.

Peptide C2 (Ala-Glu-Ala-Pro-Ala-Gly-Asp-Pro-Thr-Thr-Gly-Ala-Lys-Ile-Phe). This gave an *N*-terminal alanine. The peptide was then digested with papain and only two major peptides were located after electrophoresis at pH 1.9. The original peptide had a charge of −1 at pH 6.5, indicating that residues −5 and +2 are both acidic. Insufficient material was available for an amino acid analysis of this peptide.

Peptide C3 (Lys - Thr - Lys - $CySO_3$* - Ala - Glx -

* Abbreviations: $CySO_3$, cysteic acid; TML, ε-*N*-trimethyl-lysine.

Table 1. *Results of dansyl–Edman analysis*

Peptide	Mobility* at pH 6.5	Mobility† at pH 1.9	Dansyl–Edman results
C1	−1.50	0	Does not react
C2	−0.67		Ala-Digested with papain
C2 P1		0.47	Ala-Glx-Ala-Pro-Ala-Gly
C2 P2		0.52	Asx-Pro-Thr-Thr-Gly-Ala-Lys-Ile-Phe
C3	0.54		Lys-Performic oxidized and digested with trypsin.
C3 T1		1.61	Lys-Thr-Lys
C3 T2		1.36	Thr-Lys
C3 T3		0.24	$CySO_3$-Ala-Glx-$CySO_3$-His-Thr-Val-Glx-Lys
C3 T4		1.14	Gly-Ala-Gly-His
C4	0.86	0.67	Lys-Glx-Gly-Pro-Asx-Leu-Asx
C5	0	0.60	Gly-Leu-Phe
C6	0.65	0.60	Gly-Arg-Glx-Ser-Gly-Thr-Thr-Ala-Gly-Tyr
C7	0	0.63	Ser-Tyr
C8	0.90	0.73	Ser-Ala-Ala-Asx-Lys-Asx-Met
C9	0	0.46	Ala-Val-Ile
C10	−1.78	0.32	Glx-Glx-Asx-Thr-Leu-Tyr
C11	−1.35	0.44	Asx-Tyr-Leu
C11A	−1.65	0.49	Asx-Tyr
C12	1.70	1.02	Leu-Asx-Pro-TML-Lys-Tyr
C13	1.02	0.78	Ile-Pro-Gly-Thr-Lys-Met
C14	0	0.42	Val-Phe-Pro-Gly-Leu
C15	1.70	1.31	TML-Lys-Pro-Glx-Glx-Arg
C16	−0.98		Ala-Asx-Leu-Ile-Ala-Tyr
C17	1.02	0.84	Leu-Lys-Thr-Ser-Thr-Ala

* From DNS-arginine relative to DNS-Arg-Arg.

† From 1-dimethylaminonaphthalene-5-sulphonic acid relative to DNS-arginine.

Table 2. *Results of quantitative amino acid analyses*

Amino acids found in molar ratios of less than 0.25 are ignored.

	C1	C3	C4	C5	C6	C7	C8(A)	C10	C11(A)	C12	C13	C14	C15	C16	C17
Asp			1.87 (2)				0.98 (1)	0.95 (1)	1.00 (1)	1.03 (1)				1.00 (1)	
Thr		2.06 (2)			1.92 (2)			0.68 (1)			0.85 (1)				1.66 (2)
Ser	0.85 (1)				0.90 (1)	1.00 (1)	1.00 (1)								0.89 (1)
Glu		2.24 (2)	1.13 (1)		1.04 (1)			1.76 (2)					1.92 (2)		
Pro			* (1)							0.97 (1)	0.73 (1)	1.09 (1)	1.00 (1)		
Gly		* (2)	1.25 (1)	1.10 (1)	* (3)						1.25 (1)	1.06 (1)			
Ala	1.12 (1)	1.92 (2)			0.97 (1)		1.64 (2)							1.22 (2)	0.94 (1)
Val		0.98 (1)										0.73 (1)			
Met											1.00 (1)				
Cys		0.83 (2)													
Ile											0.76 (1)			0.79 (1)	
Leu			1.00 (1)	1.00 (1)				1.00 (1)		1.00 (1)		1.00 (1)		1.23 (1)	1.00 (1)
Tyr					0.92 (1)	0.79 (1)		* (1)	0.65 (1)	0.91 (1)				* (1)	
Phe	1.00 (1)			0.94 (1)								0.95 (1)			
TML										0.96 (1)			0.76 (1)		
Lys		* (3)	1.32 (1)				1.10 (1)			1.05 (1)	1.02 (1)		* (1)		1.02 (1)
His		2.00 (2)													
Trp															
Arg					1.00 (1)								1.24 (1)		

* Amino acid present but peak not suitable for calculation.

$CySO_3$-His-Thr-Val-Glx-Lys-Gly-Ala-Gly-His). This was the haem peptide and gave an *N*-terminal lysine. It was digested with trypsin and then performic oxidized. After electrophoresis at pH 1.9, four major peptides were located, together with free lysine. The amide content could not be inferred directly.

Peptide C4 (Lys-Gln-Gly-Pro-Asn-Leu-Asn). The charge at pH 6.5 of +1 indicates that residues 28, 31 and 33 are all amides. A peptide due to partial cleavage at asparagine-33 was also found in reasonable yield.

Peptide C5 (Gly-Leu-Phe).

Peptide C6 (Gly-Arg-Gln-Ser-Gly-Thr-Thr-Ala-Gly-Tyr). The new solvent system described allowed a positive identification of arginine-38 in the dansyl–Edman analysis. A charge of +1 at pH 6.5 indicates that residue 39 is glutamine.

Peptide C7 (Ser-Tyr).

Peptide C8 (Ser-Ala-Ala-Asn-Lys-Asn-Met). The peptide was difficult to separate from peptide C13 and was recovered in only low yield. Sufficient was available for dansyl–Edman analysis only. A peptide resulting from partial cleavage of residue lysine-53 was also located and this was used to determine the amino acid composition on the automatic analyser. The charge of both these peptides at pH 6.5, +1 indicates that both residues 52 and 54 are amides.

Peptide C9 (Ala-Val-Ile-Trp). This was the only Ehrlich-positive peptide, and gave a pink colour in the trifluoroacetic acid stage of the first Edman

degradation, indicating tryptophan (Uphaus, Grossweiner, Katz & Kopple, 1969). Sufficient peptide could not be separated from peptide C14 for an amino acid analysis.

Peptide C10 (Glx-Glx-Asx-Thr-Leu-Tyr). The charge of −2 at pH 6.5 indicates that one of the acidic residues is in the amide form.

Peptide C11 (Asp-Tyr-Leu). The charge of −1 at pH 6.5 indicates residue 66 is aspartic acid. The peptide breaking after tyrosine-67 was obtained in major yield.

Peptide C12 (Leu-Asn-Pro-TML-Lys-Tyr). The charge of +2 at pH 6.5 indicates residue 70 is asparagine. The use of the new solvent system allowed a positive identification of ϵ-*N*-trimethyl-lysine in position 72.

Peptide C13 (Ile-Pro-Gly-Thr-Lys-Met). This peptide was strongly platinic iodide-positive (methionine).

Peptide C14 (Val-Phe-Pro-Gly-Leu).

Peptide C15 (TML-Lys-Pro-Glx-Glx-Arg). The three basic residues were positively identified in positions 86 (ϵ-*N*-trimethyl-lysine), 87 (lysine) and 91 (arginine) by using the new solvent system described. The charge of +2 at pH 6.5 indicates one amide residue, which cannot be placed. The break at arginine-91 was the major break although a peptide in minor yield breaking at leucine-94 was also found.

Peptide C16 (Ala-Asp-Leu-Ile-Ala-Tyr). The charge of −1 at pH 6.5 indicates residue 93 is aspartic acid. A low yield of alanine in the peptide analysis (Table 2) was probably due to incomplete hydrolysis of the Ile-Ala bond. A peptide Ile-Ala-Tyr was also found in low yield.

Peptide C17 (Leu-Lys-Thr-Ser-Thr-Ala). This peptide was suspected to be the *C*-terminal peptide of the protein since alanine is not a residue normally susceptible to chymotryptic cleavage.

When compared with the chymotryptic peptides of mung-bean cytochrome *c* and wheat germ cytochrome *c*, the equivalent number of sunflower peptides of identical length and sequence, with the exception of a limited number of amino acid substitutions, clearly indicates the close homology that exists between these three proteins. Thus it is possible, by alignment with the complete sequences of the other two cytochromes, to deduce an unambiguous sequence for sunflower cytochrome *c* by using the chymotryptic peptides alone (Fig. 1). No other peptides were discovered during the sequence analysis. A quantity of core material remained on the pH 6.5 electrophoresis origin. The electrophoretic mobilities of the peptides indicate ten amides in the set of peptides examined of which seven can be unambiguously placed: glutamine in positions 28, 39 and asparagine in positions 31, 33, 52, 54, 70. The remaining three cannot be placed

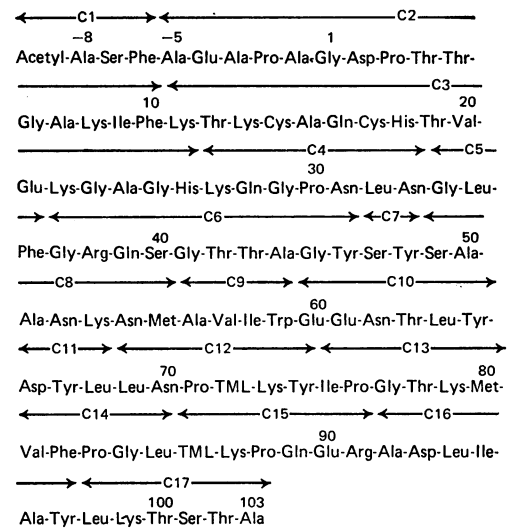

Fig. 1. Amino acid sequence of sunflower cytochrome *c*. Residue numbering is based on alignment with mammalian cytochrome *c*.

on the evidence given but are known to be in peptides C3, C10 and C15, and have been assigned by homology. Thus in peptide C3, the haem peptide, the amide content was determined by the method used with mung-bean cytochrome *c* (Thompson *et al.* 1970), and glutamine placed in position 16 and glutamic acid in position 21 by homology with cytochrome *c* from other species. In peptide C10 glutamic acid is placed in positions 60 and 61 and asparagine in position 62 by homology with all plant and fungal cytochrome *c*. In peptide C15 glutamine is placed in position 89 and glutamic acid in position 90 by homology with all other cytochromes (Dayhoff, 1969).

Cleavage at the *C*-terminal of all residues normally susceptible to chymotrypsin was observed except for phenylalanine-82 which is adjacent to a proline residue and tyrosine-67 where only incomplete cleavage was observed. Cleavage also occurred at histidine-26 and both methionine residues (55, 80). Cleavage occurred at leucine-68 and leucine-85 and partially at both asparagine-33 and leucine-94, whereas no other break involving either of these residues was found. A major tryptic-type specificity was observed in the cleavage at the *C*-terminal of arginine-91. Cleavage at lysine-53 was also found.

Comparison of the spectral ratios of the material used for this investigation with the spectral ratios of other plant cytochrome *c* (Richardson *et al.* 1970) indicated that the preparation might not have been pure. A total amino acid composition of

the preparation showed an abnormally high glycine value but peptides due to an impurity were not evident during the examination of the chymotryptic digest. Probably the impurity was not digested by chymotrypsin and represents the material that stayed on the origin in the pH 6.5 electrophoresis. However, electrophoresis at pH 6.5 of a tryptic digest of the same material clearly revealed the presence of an impurity. The electrophoretogram contained three major basic peptide zones, which were eluted and again subjected to electrophoresis at pH 1.9. As most of the peptides, when separated, gave dansyl–Edman analyses (Gly-Gly-Gly-Gly-), a preparation of higher purity than that used here would be required to examine tryptic peptides and to obtain a complete amino acid composition.

DISCUSSION

The unambiguous sequence of sunflower cytochrome *c* has been determined from the chymotryptic peptides and their close homology with other plant cytochrome *c*. The determination of the sequence was quite possible even in the presence of impurity in the preparation; of the ten amide residues present those in peptides C3, C10 and C15 were assigned to positions 16, 62 and 89 by homology with other cytochrome *c*. The sunflower cytochrome *c* sequence is given in Fig. 1 as a single chain of 111 amino acids with a molecular weight of 12702 including the prosthetic group. The unusual amino acid, ϵ-*N*-trimethyl-lysine, first reported in wheat germ cytochrome *c* (Delange, Glazer & Smith, 1969) has been found in positions 72 and 86 in the sunflower cytochrome *c* sequence and thus is found in these positions in all the plant cytochrome *c* so far examined. Similarly, an *N*-acetylated 'tail' of identical length starting acetyl-Ala-Ser-Phe is a common feature of this group. Of the 20 residues found in those positions only in wheat germ cytochrome *c* and mung-bean cytochrome *c*, 17 are the same in the sunflower cytochrome *c* sequence. The *C*-terminal peptide, like that of mung-bean cytochrome *c*, but unlike that of wheat germ cytochrome *c* is one residue short compared with the majority of other cytochrome *c* (Dayhoff, 1969).

The determination of the sequence of sunflower cytochrome *c* is a further step in an investigation into the ordination of the major plant orders by using primary sequence data (Boulter, Laycock, Ramshaw & Thompson, 1970).

We thank Dr M. E. Boulter and Mrs L. Mayo for help with the dansyl–Edman analysis, and the Science Research Council and the Nuffield Foundation for financial support.

REFERENCES

Boulter, D., Laycock, M. V., Ramshaw, J. A. M. & Thompson, E. W. (1970). In *Phytochemical Phylogeny*, p. 180. Ed. by Harborne, J. B. London and New York: Academic Press.

Dayhoff, M. O. (1969). *Atlas of Protein Sequence and Structure*, vol. 4. Silver Spring, Md.: National Biomedical Research Foundation.

Delange, R. J., Glazer, A. N. & Smith, E. L. (1969). *J. biol. Chem.* **244**, 1385.

Frangione, B. & Milstein, C. (1968). *J. molec. Biol.* **33**, 893.

Gray, W. R. & Hartley, B. S. (1963). *Biochem. J.* **89**, 379.

Margoliash, E., Kimmel, J. R., Hill, R. L. & Schmidt, W. R. (1962). *J. biol. Chem.* **237**, 2148.

Nolan, C. & Margoliash, E. (1966). *J. biol. Chem.* **241**, 1049.

Offord, R. E. (1966). *Nature, Lond.*, **211**, 591.

Richardson, M., Laycock, M. V., Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1970). *Phytochemistry*, (in the Press).

Schmer, G. & Kreil, G. (1969). *Analyt. Biochem.* **29**, 186.

Smith, E. L. & Kimmel, J. R. (1960). In *The Enzymes*, vol. 4, p. 133. Ed. by Boyer, P. D., Lardy, H. & Myrbäck, K. New York: Academic Press Inc.

Stevens, F. C., Glazer, A. N. & Smith, E. L. (1967). *J. biol. Chem.* **242**, 2764.

Thompson, E. W., Laycock, M. V., Ramshaw, J. A. M. & Boulter, D. (1970). *Biochem. J.* **117**, 183.

Uphaus, R. A., Grossweiner, L. I., Katz, J. J. & Kopple, K. D. (1959). *Science, N.Y.*, **129**, 641.

Woods, K. R. & Wang, K. T. (1967). *Biochim. biophys. Acta*, **133**, 369.