The Amino Acid Sequence of Plastocyanin from *Cucurbita pepo* L. (Vegetable Marrow)

By MICHAEL D. SCAWEN* and DONALD BOULTER

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

(Received 9 April 1974)

The amino acid sequence of plastocyanin from marrow was determined. It consists of a single polypeptide chain of mol.wt. 10284 containing 99 amino acid residues. The sequence was determined by using a Beckman 890C automatic sequencer and by dansyl-phenyl isothiocyanate analysis of peptides obtained by the enzymic digestion of purified CNBr fragments. The sequence is in good agreement with the amino acid composition, except that fewer residues of glutamic acid were found in the sequence than were suggested by the composition. Evidence for histidine-37 was weaker than for the rest of the sequence. A 'tree' of phylogenetic affinities was constructed by using several higher-plant plastocyanin sequences.

Plastocyanin from marrow has been purified and some of its properties studied (Scawen & Hewitt, 1971). The aim of the present investigation was to determine its amino acid sequence and to use the data in conjunction with the amino acid sequences of other green-plant plastocyanins to construct a 'tree' relating the sequences.

Experimental

Materials

Plastocyanin was prepared from marrow leaves (cv. Asmer's Gourmet) by using the methods described by Scawen & Hewitt (1971). Other materials were as described by Thompson *et al.* (1970) and Scawen *et al.* (1974).

Methods

S-Carboxymethylcysteinyl plastocyanin was prepared and treated with CNBr as described previously (Scawen *et al.*, 1974). The fragments were separated by gel filtration on a column (1 cm × 110 cm) of Sephadex G-50 (fine grade) equilibrated in 70% (v/v) formic acid. The purified fragments were digested with trypsin or chymotrypsin for 10 min at 37°C by using a 2% (w/w) ratio of enzyme to substrate in 0.2M-N-ethylmorpholine adjusted to pH 8.5 with glacial acetic acid, and then with thermolysin for 1 h at 37°C by using a 2% (w/w) ratio of enzyme to substrate in the same buffer containing 5 mM-CaCl₂. The peptides were separated by high-voltage paper electrophoresis at pH 6.5 and 1.9 (Thompson *et al.*, 1970).

* Present address: Department of Biochemistry, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, U.K.

and by paper chromatography in butan-1-ol-acetic acid-water-pyridine (15:3:12:10, by vol.).

The amino acid sequences of the purified peptides were determined by using the dansyl-phenyl isothiocyanate method (Gray & Hartley, 1963) as described by 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.* (1970). C-Terminal amino acids were identified by using carboxypeptidase analysis as described by Ramshaw *et al.* (1974a).

Automated sequencer analysis was carried out on a Beckman 890C automatic protein sequencer by using the 'fast protein' and 'fast peptide' programmes, as described in the operation manual for the instrument (Beckman, 1972). A single analysis was performed in each case. The sequencer-grade chemicals used were obtained from Beckman-RIIC Limited, Glenrothes, Fife, U.K. The protein programme used a quadrol buffer during the coupling reaction, as described by Edman & Begg (1967). The volatile buffer used during the coupling reaction in the peptide programme was 0.5M-N-allyl-*NN*-dimethylamine in 60% (v/v) pyridine, adjusted to pH 9.5 with 50% (v/v) trifluoroacetic acid (Niall *et al.*, 1969).

Phenylthiohydantoin amino acids were identified by, (1) t.l.c. on precoated silica-gel sheets incorporating a fluorescent indicator (Eastman Kodak Ltd.) by using the solvent systems described by Jeppsson & Sjöquist (1967); (2) by g.l.c. on SP400-coated Chromosorb W AW-DMCS (100–120 mesh; Phase Separations Ltd., Queensferry, Flintshire, U.K.), as described in the instrument manual (Beckman, 1972); and (3) as their dansyl derivatives after hydrolysis in constant-boiling hydriodic acid (Inglis *et al.*, 1971). The phenylthiohydantoin derivatives of most residues

Fig. 1. Amino acid sequence of marrow plastocyanin

Peptides shown by a solid line (—) are those which were analysed quantitatively. Peptides shown by a dashed line (---) were analysed only qualitatively. Solid arrows (→) indicate that the residue was identified unambiguously; dashed arrows (---→) indicate that the residue could not be unambiguously identified. Arrows (---→) indicate that the C-terminal residue of a peptide was identified as the free amino acid. In peptides marked * methionine was identified as homoserine.

were identified both by g.l.c. and by t.l.c. The exceptions were the derivatives of threonine and serine, which were most easily identified after hydrolysis in hydriodic acid as α -aminobutyric acid and alanine respectively; the derivative of lysine was identified by t.l.c. and as the parent amino acid after hydrolysis; the derivative of asparagine could only be identified by t.l.c.

An identification by g.l.c. was accepted if the major peak represented, after correction for carry-over and background, at least 80% of the derivatized amino acid injected on to the column. Identification on t.l.c. and after hydrolysis was more subjective, but a positive identification was given if only one major spot was present on the chromatograms.

Samples of the purified CNBr fragments and peptides for amino acid analysis were hydrolysed in constant-boiling HCl at 105°C for 24 h. Samples of the protein were hydrolysed for 24 and 72 h. The analyses were carried out on a Locarte amino acid analyser.

Quantities of material used. For CNBr cleavage 40 mg of *S*-carboxymethylcysteinyl plastocyanin was used. Approx. 1 μ mol of the purified fragments was used for each enzymic digestion. Some 400 nmol of native protein and 300 nmol of fragment X-2 were used for the automatic sequencer analyses. Quantita-

tive amino acid analyses were carried out on 15–25 nmol of peptide or purified CNBr fragment and 50 μ g of total protein.

Nomenclature. The CNBr fragments are labelled X, and are numbered in Fig. 1 on the basis of their position in the complete sequence. Peptides obtained by enzymic digestion of the purified fragments are numbered according to their position in the parent fragment. The following abbreviations are employed: T, trypsin; C, chymotrypsin; H, thermolysin.

Results

The amino acid composition of marrow plastocyanin was determined by analysis and by calculation from the sequence (Table 1). The values are generally in good agreement, although those for glutamic acid, lysine and valine were higher by amino acid analysis. Plastocyanin was shown to have an *N*-terminal isoleucine and a *C*-terminal asparagine residue.

The treatment of *S*-carboxymethylcysteinyl marrow plastocyanin with CNBr gave three fragments, which were separated by gel filtration. The largest of the fragments, X-1, had *N*-terminal isoleucine and contained homoserine. In conjunction with the automatic sequencer analysis of the native

Table 1. *Amino acid composition of marrow plastocyanin and its CNBr fragments*

The results are expressed as residues/molecule and are based on 24 h and 72 h hydrolyses for the total protein and 24 h hydrolyses only in the case of CNBr fragments. For the complete protein the values taken were the means of the 24 h and 72 h values except that values for threonine, serine and tyrosine were corrected for losses, assuming first-order kinetics of destruction (Moore & Stein, 1963), and that for valine and isoleucine 72 h values were taken. Cystine was determined as cysteic acid or *S*-carboxymethylcysteine. The value for tryptophan was determined by using the method of Beaven & Holliday (1952) assuming the presence of three tyrosine residues.

	Total protein				X-1 fragment		X-2 fragment		X-3 fragment	
	24h	72h	Value taken	Sequence	24h	Sequence	24h	Sequence	24h	Sequence
Asp	15.4	15.2	15.3	15	9.6	10	4.2	4	1.1	1
Thr	1.8	1.1	2.3	2	0	0	0.9	1	0.8	1
Ser	6.3	5.1	7.1	7	3.5	4	2.7	3	0	0
Glu	11.3	11.8	11.5	9	4.3	4	5.9	5	0	0
Pro	5.2	5.4	5.3	5	3.3	3	2.3	2	0	0
Gly	12.6	11.2	11.9	12	6.9	7	4.1	4	1.0	1
Ala	7.2	7.4	7.3	7	5.3	5	2.1	2	0	0
Cys	0.9		0.9	1	0	0	0.9	1	0	0
Val	9.9	11.4	11.4	11	5.8	6	1.8	2	2.7	3
Met	1.7	1.8	1.8	2	0	0	0	0	0	0
Ile	4.4	5.3	5.3	5	4.8	5	0	0	0	0
Leu	5.9	6.5	6.2	6	3.2	3	3.1	3	0	0
Tyr	2.7	2.4	2.9	3	0.2	0	2.8	3	0	0
Phe	5.9	5.9	5.9	6	5.2	5	1.1	1	0	0
His	2.2	1.9	2.0	2	1.0	1	0.9	1	0	0
Lys	6.8	6.6	6.7	6	3.2	3	1.5	2	1.2	1
Arg	0	0	0	0	0	0	0	0	0	0
Trp			0.1	0		0		0		0
Hse	0	0	0	0	0.6	1	0.5	1	0	0
Totals				99		57		35		7

Table 2. Properties and amino acid compositions of peptides from marrow plastocyanin

Electrophoretic mobilities at pH 6.5 were measured from dansyl-arginine relative to dansyl-arginylarginine (the mobility of aspartic acid in this system is -0.365) and at pH 1.9 from 5-dimethylaminonaphthalene-1-sulphonic acid relative to dansyl-arginine. BAWP, *R_f* of peptide on paper chromatography in butan-1-ol-acetic acid-water-pyridine (15:3:12:10, by vol.). The compositions are expressed as residues/molecule and are based on the results of single 24 h hydrolyses. Amino acids present in quantities less than 0.2 residues/molecule were ignored. +ve or -ve indicate that the peptides were positively or negatively charged, but that their mobilities could not be accurately determined.

Peptide	Mobility at			BAWP	Asp	Thr	Ser	Glu	Pro	Gly	Ala	Val	Ile	Leu	Tyr	Phe	His	Lys	Hse	Cys
	pH6.5	pH1.9																		
X1H1	-1.2	0.45		0.85				1.0				1.0	0.9	1.0						
X1H2	-2.0	0.42		0.25	2.0		1.0			3.2		1.0	0.9	0.9						
X1H3	-0.85	0.33		0.67	2.0				1.3		1.1		0.9	1.0		0.9				
X1H3a	0	0.72		0.67																
X1H3b	-1.0	0.42		0.55																
X1H4	0	0.65		0.53			0.9									1.0				
X1H5	0	0.79		0.20				1.0		0.9	1.8	0.9						1.2		
X1H6	+1.0	0.71		0.40	2.9			0.9	0.9	1.3	1.2	0.9	0.9			1.7	0.8	0.8		
X1H7	-1.7	0.40		0.33	2.9		1.1	1.9	1.2	2.2	1.1	2.5	1.1			1.1		1.0		
X1H8	+1.7	0.60		0.44			1.1						1.0						0.6	
X1C1	-1.20	0.21		—	2.0		0.7	0.9		2.5	1.1	0.9	1.0	2.9		1.0				
X1C2	-0.95	0.35		—																
X1C3	0	0.58		—																
X1C4	+ve	—		—																
X1C5	-ve	—		—	2.7		1.8	2.2	1.0	2.3	0.9	1.2	2.1					1.1	0.6	
X2H1	-3.1	0.39		0.17	1.9			2.1						1.0						
X2H2	-1.4	0.30		0.45	0.8			0.8	1.0	1.1	1.0	0.9		0.9	0.7			0.8		
X2H3	0	0.70		0.42	1.1							0.9								
X2H4	0	0.63		0.20				1.0		1.1				1.1	0.9			0.9		
X2H5	0	0.51		0.21		0.7	1.5	1.2	1.2	2.1	1.1				0.8	0.8	1.1		0.6	1.0
X2T1	-1.8	—		—			0.9													
X2T2	0	0.81		—																
X2T3	0	0.39		—																
X3	0.85	0.75		—																
							See Table 1													

See Table 1

protein this fragment was shown to represent the *N*-terminal region of the sequence, residues 1–57 (Fig. 1). The second fragment to be eluted from the column, X-2, had *N*-terminal asparagine and contained homoserine, and so was internal in the sequence (residues 58–92, Fig. 1). The smallest fragment, X-3, had *N*-terminal valine, but did not contain homoserine. Instead it had *C*-terminal asparagine, and so represents the *C*-terminal region of the sequence, residues 93–99 (Fig. 1). The amino acid compositions of the purified fragments, determined by analysis and calculated from their sequences, are shown in Table 1. The results are generally in good agreement, except that in fragment X-2 the analytical value for glutamic acid was high whereas that for lysine was low; the recoveries of homoserine were consistently low.

The sequence of the fragment X-1 was established partly from the automated sequencer analysis of the native protein, which enabled the first 24 residues to be identified, and partly from the dansyl-phenyl isothiocyanate analysis of peptides obtained by digestion with chymotrypsin and thermolysin. The properties and amino acid compositions of these peptides are given in Table 2.

The sequence of the fragment X-2 was established by dansyl-phenyl isothiocyanate analysis of peptides obtained by digestion with trypsin and thermolysin. In addition automated sequence analysis of this fragment, by using the fast peptide programme, enabled most of the first 21 residues to be identified. The properties and amino acid compositions of these peptides are shown in Table 2.

The smallest fragment, X-3, was further purified by high-voltage paper electrophoresis at pH 6.5 and 1.9 before its sequence was determined by dansyl-phenyl isothiocyanate analysis. Electrophoresis at pH 6.5 in addition to the main peptide gave a neutral peptide in low yield with the same composition; this peptide was probably formed from the parent peptide by deamidation at asparagine-99, as was also found to be the case for potato (Ramshaw *et al.*, 1974a), broad-bean (Ramshaw *et al.*, 1974b) and *Chlorella* plastocyanins (Kelly & Ambler, 1973).

Discussion

The sequences of the three CNBr fragments account for the composition of the whole protein. Since no other CNBr fragments were found, it is felt that the analysis of these fragments provides sufficient evidence for the sequence of the protein, even though overlapping peptides through the two methionine residues were not analysed. The evidence for the complete sequences of the two fragments X-1 and X-2 is based on the results obtained from the automatic sequencer in conjunction with the dansyl-phenyl isothiocyanate analysis of the purified peptides obtained from these fragments. In the case of fragment X-1 the sequencer result overlaps peptides X1H1 to X1H5 and X1C1 to X1C3. Only peptides X1H8 and X1C5 contained homoserine so these must be the *C*-terminal peptides of the fragment. Peptide X1C5 provides an overlap between peptides X1H7 and X1H8 and peptide X1C4, the only remaining chymotryptic peptide not to contain homoserine, provides an overlap between peptides X1H6 and X1H7, so giving a complete overlapping sequence for this fragment. Most of the first 24 residues in this fragment were identified from the sequencer run and from the results of two separate digests, whereas most of the remaining 33 residues were identified from the results of the two digests. In the case of fragment X-2 the sequencer result overlaps peptides X2H1 to X2H4 and X2T1 to X2T3. In addition only peptides X2H5 and X2T3 contained homoserine, so therefore these must be the *C*-terminal peptides in the fragment. Thus, of the 92 residues in these two fragments, 48 were identified from the automatic sequencer results. It is our view that modern automatic sequencers are sufficiently reliable so that the results obtained from them can be used as the basis of evidence for a sequence as well as to provide the necessary overlaps with purified peptides that have been sequenced by the classical dansyl-phenyl isothiocyanate method.

Evidence for histidine-37 was weaker than for the rest of the sequence. No clear evidence was found for histidine-37 during the dansyl-phenyl isothiocyanate analysis of peptides X1H6 and X1C4, although the

Table 3. *A matrix of amino acid differences between plastocyanins*

The non-integral values result from the micro-heterogeneity found in the elder plastocyanin (Scawen *et al.*, 1974). Other data are taken from potato (Ramshaw *et al.*, 1974a), broad bean (Ramshaw *et al.*, 1974b) and *Chlorella* (Kelly & Ambler, 1973).

	Marrow	Broad bean	Elder	Potato	Chlorella
Marrow	—				
Broad bean	20	—			
Elder	19½	21½	—		
Potato	17	22	18	—	
<i>Chlorella</i>	49	43	48½	48	—

Fig. 2. Ancestral amino acid sequence topology of five green-plant plastocyanins

quantitative analysis of peptide X1H6 showed the presence of one histidine residue. The amino acid composition of the whole protein and CNBr fragments showed the presence of two histidines, one in fragment X-1 and one in fragment X-2. That in fragment X-2 was positively identified at position 87. There was no evidence for the presence of a histidine residue in any other peptide that was analysed, either quantitatively or qualitatively. The electrophoretic mobility of peptide X1H6 at pH6.5 indicated the presence of a greater positive charge than could be accounted for by the presence of lysine-30 alone. Every plastocyanin examined so far has a histidine residue in position 37, potato (Ramshaw *et al.*, 1974a), broad bean (Ramshaw *et al.*, 1974b); elder (Scawen *et al.*, 1974) and *Chlorella fusca* (Kelly & Ambler, 1973), and considerations of homology therefore suggest that marrow plastocyanin contains a histidine residue here, although homology with other plastocyanin sequences has not been used to assign residues in this sequence.

All the amide residues in the sequence have been assigned, either from the automatic sequencer results or from a consideration of the electrophoretic mobilities of the peptides at pH6.5 (Offord, 1966). Two peptides, X1C4 and X1C5, did not migrate as discrete zones during high-voltage paper electrophoresis at pH6.5, so consequently no definite mobilities could be assigned to these peptides.

Where quantitative amino acid compositions of peptides were determined, the results were in reasonable agreement with the sequences of the peptides except that recoveries of hydrophobic amino acids were sometimes low, particularly where there was a cluster as in peptide X1H7. The recoveries of tyrosine and homoserine were often low.

The molecular weight of marrow plastocyanin as calculated from the sequence is 10284, which compares favourably with the values of 10300 to 12500 given earlier by Scawen & Hewitt (1971). The sequence of marrow plastocyanin is similar to the sequences of potato (Ramshaw *et al.*, 1974a), broad-bean (Ramshaw *et al.*, 1974b) and *Chlorella* plasto-

cyanins (Kelly & Ambler, 1973). In common with the other higher-plant plastocyanins both arginine and tryptophan are absent. All the isoleucine residues are in fragment X-1, whereas all the tyrosine residues are in fragment X-2, which was also found to be the case with the other higher-plant plastocyanins listed above.

When the four higher-plant sequences are compared with the *Chlorella* plastocyanin sequence the regions 31–49 and 78–92, which have been noted as being highly conserved regions (Ramshaw *et al.*, 1974a,b; Scawen *et al.*, 1974), are still found to be conserved and may therefore be of importance in the function of the protein. There are 43 positions which are invariant between these five sequences, and of these 10 are glycine residues. In the marrow sequence positions 71 and 79, which had previously been considered invariant, are no longer so. In position 71 lysine replaces serine and in position 79 serine replaces threonine. Table 3 shows a matrix of amino acid differences between the five plastocyanin sequences.

Fig. 2 relates the sequences by using the ancestral amino acid sequence method (Dayhoff, 1972). This topology does not coincide with the relationships proposed for the same or related species of cytochrome *c* sequences (Boulter, 1973). However, it is approximately the same, and, in view of the few sequences considered and the fact that the distances between the nodes are small compared with the branch length of the sequences, changes are expected to occur when more data is available.

We thank Dr. D. Peacock for analysing the data by using the ancestral-sequence method, John Gilroy and Elizabeth Jones for technical help and Anne Dolby for the amino acid analyses.

References

- Beaven, G. H. & Holliday, E. R. (1952) *Advan. Protein Chem.* **25**, 319–386
- Beckman (1972) *Model 890C Sequencer Instruction Manual*, Spinco Division, Beckman Instruments Inc., Palo Alto

- Boulter, D. (1973) *The Use of Amino Acid Sequence Data in the Classification of Higher Plants*, Nobel Symp. **25**, *Chemistry in Botanical Classification* (Bendz, G. & Santesson, J., eds.), pp. 211–216, Nobel Foundation, Stockholm, and Academic Press, New York
- Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, pp. 7–16, National Biomedical Research Foundation, Silver Springs
- Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1**, 89–91
- Gray, W. R. & Hartley, B. S. (1963) *Biochem. J.* **89**, 378–380
- Inglis, A. S., Nicholls, P. W. & Roxburgh, C. M. (1971) *Aust. J. Biol. Sci.* **24**, 1247–1250
- Jeppsson, J. O. & Sjöquist, J. (1967) *Anal. Biochem.* **18**, 264–269
- Kelly, J. & Ambler, R. P. (1973) *Biochem. Soc. Trans.* **1**, 164–166
- Moore, S. & Stein, W. H. (1963) *Methods Enzymol.* **6**, 817–831
- Niall, H. D., Keutmann, H. T., Copp, D. H. & Potts, J. T. (1969) *Proc. Nat. Acad. Sci. U.S.* **64**, 771–778
- Offord, R. E. (1966) *Nature (London)* **211**, 591–593
- Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1970) *Biochem. J.* **119**, 535–539
- Ramshaw, J. A. M., Scawen, M. D., Bailey, C. J. & Boulter, D. (1974a) *Biochem. J.* **139**, 583–592
- Ramshaw, J. A. M., Scawen, M. D. & Boulter, D. (1974b) *Biochem. J.* **141**, 835–843
- Scawen, M. D. & Hewitt, E. J. (1971) *Biochem. J.* **124**, 32P
- Scawen, M. D., Ramshaw, J. A. M., Brown, R. H. & Boulter, D. (1974) *Eur. J. Biochem.* **44**, 299–303
- Thompson, E. W., Laycock, M. V., Ramshaw, J. A. M. & Boulter, D. (1970) *Biochem. J.* **117**, 183–192
- Woods, W. R. & Wang, K. T. (1967) *Biochim. Biophys. Acta* **133**, 369–370.