The Amino Acid Sequence of Plastocyanin from Spinach (*Spinacia oleracea* L.)

By MICHAEL D. SCAWEN,* JOHN A. M. RAMSHAW†
and DONALD BOULTER

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

(Received 28 November 1974)

The amino acid sequence of spinach (*Spinacia oleracea* L.) plastocyanin was determined. It consists of a single polypeptide chain of 99 residues and has a sequence molecular weight of 10415. The sequence was determined by using a Beckman 890C automatic sequencer and by the dansyl-phenyl isothiocyanate analysis of peptides obtained by the enzymic digestion of purified CNBr fragments. Overlap through the two methionine residues was not shown. Sedimentation equilibrium in the ultracentrifuge gave a molecular weight for spinach plastocyanin of about 9000, in contrast with the value of 21000 reported previously by Katoh *et al.* (1962).

Spinach plastocyanin was the first higher-plant plastocyanin to be purified and studied (Katoh *et al.*, 1962). It was reported to have a molecular weight of 21000 and to contain two atoms of copper. More recently, plastocyanins isolated from a number of higher-plant species have been shown in every case to have molecular weights of about 10500, and to contain one atom of copper per molecule (Milne & Wells, 1970; Ramshaw *et al.*, 1973; Scawen *et al.*, 1975); even so, a value of 21000 is still often quoted as the molecular weight of plastocyanin.

The amino acid sequences of four higher-plant plastocyanins have been determined: potato (Ramshaw *et al.*, 1974a), broad bean (Ramshaw *et al.*, 1974b), elder (Scawen *et al.*, 1974) and marrow (Scawen & Boulter, 1974), and in each case the molecular weight calculated from the sequence was about 10500. The aim of the present investigation was to determine the complete amino acid sequence of spinach plastocyanin, in order to confirm its molecular weight, as well as to give more data for a project to deduce a plant phylogeny based on amino-acid-sequence data (Boulter, 1973).

Experimental

Materials

Plastocyanin was prepared from spinach leaves as described previously (Ramshaw *et al.*, 1973), except that the leaves were stored frozen at -20°C for several weeks before extraction instead of initially

making an acetone-dried powder. Other materials were as described previously (Thompson *et al.*, 1970; Scawen *et al.*, 1974).

Methods

Sedimentation-equilibrium measurements were made with a Beckman model E analytical ultracentrifuge operated at 20°C and 28000 rev./min. Spinach plastocyanin was dissolved in potassium acetate buffer, pH 5.5, $I = 0.1$ mol/litre, at a concentration of 1 mg/ml.

Spinach plastocyanin was denatured in 6M-guanidinium chloride, reduced and *S*-carboxymethylated as described by Milne & Wells (1970), except that the entire procedure was carried out in 6M-guanidinium chloride. Performic acid oxidation was carried out as described by Moore (1963).

The *S*-carboxymethylated protein was treated with a fivefold (w/w) excess of CNBr in 70% (v/v) formic acid for 24 h (Steers *et al.*, 1965). The fragments were separated by gel filtration on a column (1 cm $\times$ 110 cm; $V_0 = 72$ ml) of Sephadex G-50 (fine grade) equilibrated in 70% (v/v) formic acid. The eluate was collected in 3 ml fractions and the elution profile was followed by measuring the E_{280} and by *N*-terminal analysis by using the dansyl technique (Gray & Hartley, 1963).

Digestion by trypsin was for 10 min at 37°C in 0.2M-*N*-ethylmorpholine adjusted to pH 8.5 with acetic acid. Digestion by thermolysin was for 1 h at 37°C in the same buffer containing 5 mM- CaCl_2 . Digestion by papain was for 1 h at 37°C in p6.5H electrophoresis buffer (Thompson *et al.*, 1970) containing 0.1% (v/v) 2,3-dimercaptopropan-1-ol. For each digestion a 2% (w/w) ratio of enzyme to substrate was used.

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

† Present address: Department of Inorganic Chemistry, University of Sydney, Sydney, N.S.W., Australia 2006.

Peptides were separated by high-voltage paper electrophoresis at pH 6.5 or 1.9 (Thompson *et al.*, 1970) 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 as described by Thompson *et al.* (1970). Dansyl-amino acids were identified by chromatography on polyamide thin layers by using the solvent systems described by Ramshaw *et al.* (1970).

Samples for amino acid analysis containing 15–25 nmol of purified peptide or CNBr fragment, or 50 µg of protein, were hydrolysed *in vacuo* in constant-boiling HCl at 105°C for 24 or 72 h. The analyses were carried out on a Locarte amino acid analyser.

Automated sequencer analysis was carried out on a Beckman 890C protein sequencer as described by Scawen & Boulter (1974).

Quantities of material used. For CNBr cleavage 40 mg of *S*-carboxymethylated plastocyanin was used. About 1 µmol of each purified fragment was used for each enzymic digest. Automated sequence analysis was carried out on single samples

of about 0.4 µmol of native protein and fragment X-2.

Nomenclature. The purified CNBr fragments are labelled X and are numbered on Fig. 1 according to their positions in the complete sequence. Peptides obtained from these fragments are numbered according to their position in the parent fragment and are shown by H for thermolysin, T for trypsin and P for papain.

Results

The behaviour of spinach plastocyanin during its purification by chromatography on Sephadex G-75 indicated that its molecular weight was similar to the other higher-plant plastocyanins that have been prepared in this manner (Ramshaw *et al.*, 1973). A sedimentation-equilibrium experiment gave a molecular weight of about 9000, which is in reasonable agreement with the value of 10415 calculated from the amino acid sequence.

The amino acid composition of spinach plastocyanin as determined by analysis and as calculated from the sequence is shown in Table 1. These values are generally in good agreement, except that the value for glutamic acid was slightly, but consistently

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

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

	Total protein				X-1 fragment		X-2 fragment		X-3 fragment	
	24h	72h	Value taken	Sequence	24h	Sequence	24h	Sequence	24h	Sequence
Asp	11.4	10.9	11.2	11	8.4	8*	3.0	2	0.9	1
Thr	4.9	4.7	5.0	5	0.6	0	3.5	4	0.9	1
Ser	6.5	5.5	7.1	7	4.7	5	2.1	2	0.2	0
Glu	10.7	10.7	10.7	10	5.2	5	5.3	5	0	0
Pro	5.5	5.1	5.3	5	2.7	3	1.8	2	0	0
Gly	12.6	13.1	12.9	13	7.5	8	3.5	4	1.1	1
Ala	6.7	6.9	6.8	7	4.9	5	2.5	2	0	0
Cys	1.0†		1.0	1	0	0	*	1	0	0
Val	10.7	11.3	11.3	11	5.6	7	1.4	1	2.9	3
Met	2.1	1.7	1.9	2						
Ile	3.1	3.2	3.2	3	2.6	3	0.5	0	0	0
Leu	7.5	7.3	7.4	7	4.3	4	3.1	3	0	0
Tyr	2.7	2.6	2.7	3	0.3	0	2.6	3	0	0
Phe	6.4	5.7	6.1	6	4.7	5	1.2	1	0	0
His	1.9	1.9	1.9	2	1.1	1	1.0	1	0	0
Lys	6.1	6.3	6.2	6	2.3	2	3.3	3	1.1	1
Arg	0.2	0.3	0.3	0	0	0	0	0	0	0
Trp			0.11	0						
Hse					0.9	1	0.7	1	0	0
Totals				99		57		35		7

* *S*-Carboxymethylcysteine was not resolved from aspartic acid.

† Determined as cysteic acid.

high by analysis. Both arginine and tryptophan were absent and there was no evidence for the presence of any unusual amino acids. The protein was shown to have *N*-terminal valine and *C*-terminal asparagine residues.

CNBr cleaved *S*-carboxymethylated plastocyanin into three fragments, which were separated by gel filtration. The amino acid compositions of the purified fragments are shown in Table 1. The values determined by analysis are generally in good agreement with those calculated from the complete sequence (Fig. 1), except that the analysis value for valine in fragment X-1 was very low, probably because the several clusters of hydrophobic amino acids in this fragment prevented the complete release of this residue in the short hydrolysis time (24h) used. The small amounts of threonine and tyrosine present in the analysis of fragment X-1 and of isoleucine in fragment X-2 probably result from slight cross-contamination of the fragments. Otherwise the compositions of the three CNBr fragments accounted quantitatively for the composition of the complete protein.

The largest fragment, X-1 ($V_e = 84\text{--}102\text{ml}$), had an *N*-terminal valine residue and contained homoserine. These findings, together with the results of the automated sequencer analysis of the intact protein, show that this fragment represents the *N*-terminal region of the protein, residues 1–57 (Fig. 1). The second fragment to be eluted from the Sephadex G-50 column, X-2 ($V_e = 105\text{--}126\text{ml}$), had an *N*-terminal serine residue and also contained homoserine, and so must be internal in the sequence, residues 58–92 (Fig. 1). The smallest fragment, X-3 ($V_e = 141\text{--}168\text{ml}$), had an *N*-terminal valine residue but did not contain homoserine; it had a *C*-terminal asparagine residue and must therefore represent the *C*-terminal region of the protein, residues 93–99 (Fig. 1). In addition to these three major fragments, a small amount of material was eluted from the column before fragment X-1 ($V_e = 60\text{--}81\text{ml}$). This material had an *N*-terminal valine residue and contained tyrosine, a residue which was only present in fragment X-2. Probably it was present owing to incomplete cleavage at the first methionine residue; it was not analysed further.

The automated sequence analysis of the native protein enabled most of the first 40 residues in the sequence to be identified (Fig. 1). Confirmation of these residues was obtained from the results of a dansyl-phenyl isothiocyanate analysis of the peptides derived from a thermolytic digest of fragment X-1. The sequencer results were used to establish the order of all these thermolytic peptides, since they provided overlaps for every peptide except peptide XIH8. As this was the only peptide present in the digest which contained homoserine it must have

been derived from the *C*-terminus of the fragment. Residues 53 and 54 were confirmed by isolating specifically the papain 'overlap' peptide containing homoserine, as described previously (Scawen *et al.*, 1974), and carrying out a dansyl analysis.

Automated sequencer analysis of fragment X-2, by using the peptide programme, enabled most of the first 21 residues to be identified. The sequence of the remaining 14 residues was established by the dansyl-phenyl isothiocyanate analysis of purified thermolytic and tryptic peptides (Fig. 1). These peptides also gave the complete sequence of the fragment, so confirming the sequencer results. No overlap was established between residues 81 and 82 because both thermolysin and trypsin cleaved this peptide bond. However, peptides X2T4 and X2H5 were the only ones to contain homoserine and so must be the *C*-terminal peptides of the fragment. In addition, the sequencer result provided an overlap with peptides X2H1 to X2H4, leaving peptide X2H5 as the only one remaining unplaced. The properties and amino acid compositions of the peptides obtained from the different digests of these two CNBr fragments are shown in Table 2.

The smallest fragment, X-3, was further purified by high-voltage paper electrophoresis at pH 6.5 before its sequence was established by dansyl-phenyl isothiocyanate analysis. As in previous investigations (Ramshaw *et al.*, 1974a; Scawen & Boulter, 1974) the electrophoretic separation gave a neutral peptide in low yield which had the same composition as the major basic peptide and which was presumably the result of the deamidation of the latter at asparagine-99.

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 pH 6.5 (Offord, 1966).

Discussion

The sequences of the three CNBr fragments account completely for the composition of the total protein except for the discrepancies discussed above. Since no other major CNBr fragments were isolated it is considered that the sequences of these three fragments provide sufficient evidence to establish the complete sequence of the protein, even though peptides that overlap the two methionine residues were not isolated.

The amino acid compositions of peptides representing every residue in the sequence were determined and there was generally good agreement between the values calculated from the sequence and those determined by analysis. The analysis values for isoleucine and valine in peptide XIH6 and for valine in peptide XIH7 were low, probably because of

the incomplete release of these amino acids in the short hydrolysis time used (24h). In addition, the hydrolysates of peptides X1H2 and X1P contained glutamic acid, although no evidence for the presence of this amino acid was found during the dansyl-phenyl isothiocyanate analysis of these peptides. The recoveries of homoserine and of tyrosine from peptide hydrolysates were often low.

The majority of the residues in this sequence have been identified from the results of at least two independent analyses. In some cases the two independent identifications were as dansyl derivatives after the phenyl isothiocyanate degradation of peptides obtained from two separate enzymic digests (e.g.

residues 50-57). In other cases the results obtained from the automatic sequencer have been used in conjunction with those obtained from the dansyl-phenyl isothiocyanate analysis of peptides (e.g. residues 1-40). However, some residues were not identified with the usual degree of certainty. Thus of the residues identified from the sequencer analysis of the total protein, the identification of residues 20, 23, 36, 37 and 38 is weaker than the others. Since only a single thermolysin digest was carried out on the X-1 fragment, residues 41-49 were identified only once, i.e. during the dansyl degradation of peptide X1H7. However, the single papain peptide analysed, X1P, overlapped residues 50-57, therefore con-

Fig. 1. Amino acid sequence of spinach plastocyanin

Peptides and fragments indicated by a solid line (—) are those for which amino acid-analysis results were obtained. Other purified peptides are shown by dashed lines (---). Residues that were unambiguously identified are indicated by solid arrows (→); those for which some ambiguity existed, e.g. because of excessive carry-over, are indicated by dashed arrows (---→). Arrows (→) indicate that the free amino acid was identified as its dansyl derivative. In peptides marked * methionine was identified as homoserine.

firming the positioning of peptide X1H7. The region comprising residues 50-57 was established with redundant information, except that alanine-53 was not clearly identified in peptide X1H7 because of carry-over of alanine-52, and lysine-54 could not be identified in this peptide. The amino acid composition of peptide X1H7 was in good agreement with the proposed sequence, so it was

considered that there was sufficient evidence for these two residues.

There is more redundant evidence for the sequence of fragment X-2, because both tryptic and thermolytic digests were carried out in addition to a single sequencer analysis. Only residues 81, 84, 87, 91 and 92 were not clearly identified at least twice, although because the amino acid compositions of the

Table 2. *Mobilities and compositions of peptides from spinach plastocyanin*

Electrophoretic mobilities of peptides at pH 6.5 are given relative to aspartic acid (the mobility of Dns-arginylarginine = $0.365 \times$ the mobility of aspartic acid) and at pH 1.9 were measured from 1-dimethylaminonaphthalene-5-sulphonic acid relative to Dns-arginine. The amino acid compositions are given as residues/molecule and are based on single 24h hydrolyses. Residues present in a molar ratio of less than 0.2 have been ignored. Abbreviations: BAWP, butan-1-ol-acetic acid-water-pyridine (15:3:12:10, by vol.); Cm-Cys, S-carboxymethylcysteine.

Peptide	Mobilities		R_F in BAWP	Amino acid composition																
	pH6.5	pH1.9		Asp	Thr	Ser	Glu	Pro	Gly	Ala	Val	Ile	Leu	Tyr	Phe	His	Lys	Hse	Cm-Cys	
X1H1	-0.44	0.47	0.74	—	—	—	1.0	—	—	—	1.9	—	1.1	—	—	—	—	—	—	
X1H2	-0.36	0.37	0.25	1.2	—	1.1	0.5	—	3.8	—	—	—	0.9	—	—	—	—	—	—	
X1H3	0	0.70	0.68	—	—	—	—	—	—	1.0	—	—	1.0	—	—	—	—	—	—	
X1H4	-0.31	0.30	0.56	1.0	—	0.9	—	1.1	1.2	—	—	—	1.1	—	1.9	—	—	—	—	
X1H5	-0.73	0.35	0.17	—	—	0.9	2.0	—	1.0	0.8	0.9	—	—	—	—	—	—	—	—	
X1H6	+0.36	0.67	0.23	2.7	—	—	—	1.1	1.2	1.1	0.7	0.7	—	—	2.1	0.9	1.3	—	—	
X1H7	-0.69	0.40	0.26	3.0	—	1.0	2.3	1.1	1.2	2.0	2.6	1.0	1.0	—	1.0	—	1.1	—	—	
X1H8	0, +0.73	0.60	—	—	—	1.0	—	—	—	—	—	1.0	—	—	—	—	—	0.8	—	
X1P	+0.36	—	—	1.0	—	1.1	0.5	—	—	1.7	0.8	0.9	—	—	—	1.1	0.6	—	—	
X2T1	-0.51	0.46	—	2.0	0.9	0.8	2.9	1.0	1.2	1.1	—	—	2.0	0.7	—	—	1.1	—	—	
X2T2	0	0.72	—	—	1.8	—	1.0	—	—	—	1.0	—	1.1	—	—	—	1.1	—	—	
X2T3	+0.44	0.94	—	—	0.9	—	—	—	1.0	—	—	—	—	0.9	—	—	1.1	—	—	
X2T4	0	0.50	—	—	—	—	—	1.0	2.2	1.0	—	—	—	0.5	0.8	0.9	—	0.5	0.4	
X2H1	-0.95	0.40	0.26	0.9	—	0.8	2.0	—	—	—	—	—	0.9	—	—	—	—	—	—	
X2H2	0	0.54	0.21	1.0	0.8	—	1.1	1.0	1.2	1.0	—	—	1.0	0.7	—	1.1	—	—	—	
X2H2a	-0.36	0.35	0.34	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
X2H2b	+0.58	1.20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
X2H3	0	0.69	0.45	—	0.8	—	—	—	—	—	1.0	—	—	—	—	—	—	—	—	
X2H4	+0.28	0.88	0.16	—	1.7	—	1.0	1.1	1.1	1.0	—	—	1.0	0.8	1.0	1.0	2.1	—	—	
X2H5	0	0.53	0.27	—	—	1.1	1.0	1.1	2.1	1.0	—	—	—	0.6	1.0	1.0	0.6	1.0	—	
X3	+0.32	—	—	See Table 1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	

tryptic peptides are in good agreement with the proposed sequence, it is considered that there is sufficient evidence for it.

The sequence data reported in this paper show that spinach plastocyanin has a molecular weight of about 10500. This value is similar to those obtained for all the other higher-plant plastocyanins [potato (Ramshaw *et al.*, 1974a), broad bean (Ramshaw *et al.*, 1974b), elder (Scawen *et al.*, 1974) and marrow (Scawen & Boulter, 1974)] for which amino acid-sequence data are available. The results of the chromatographic and sedimentation-equilibrium investigations reported here also give a molecular weight of this order. Katoh *et al.* (1962) reported that spinach plastocyanin had a molecular weight of about 21000, as calculated from sedimentation and diffusion data, although the copper content of their preparation indicated a minimum molecular weight of 10800. With the finding that other higher-plant plastocyanins had molecular weights of about 10500 there was some speculation that spinach, and possibly other plastocyanins, could exist in a dimeric form under certain conditions (Milne & Wells, 1970), although these workers found no evidence of association in French-bean plastocyanin.

However, the value for $s_{20,w}^0$ of spinach plastocyanin, 1.72S (Katoh *et al.*, 1962), is very close to that determined for marrow plastocyanin, 1.69S (Scawen *et al.*, 1975), and similar to values of 1.36–1.42S and of 1.4S found for the $s_{20,w}$ of French-bean and parsley plastocyanins respectively (Milne & Wells, 1970; Graziani *et al.*, 1974). In contrast with the similar sedimentation values obtained for different plastocyanins, marrow plastocyanin had a $D_{20,w}$ value of $14.6 \times 10^{-7} \text{ cm}^2/\text{s}$ (Scawen *et al.*, 1975), whereas the D value for spinach plastocyanin determined by Katoh *et al.* (1962) was $6.6 \times 10^{-7} \text{ cm}^2/\text{s}$, which could account for the higher molecular-weight value of the latter protein as determined by Katoh *et al.* (1962). These values of D may be compared with those found by Graziani *et al.* (1974), who reported values of D ranging from 6×10^{-7} to $11 \times 10^{-7} \text{ cm}^2/\text{s}$ for parsley plastocyanin, which was unstable under the conditions of measurement.

We thank Dr. J. Easterby, Department of Biochemistry, University of Liverpool, for performing the sedimentation-equilibrium experiment, and Dr. R. P. Ambler, Department of Molecular Biology, University of Edinburgh, for sending us unpublished results on the partial sequence of spinach plastocyanin. We are grateful to Mrs. E. Jones, Mrs. A. Dolby and Mr. J. Gilroy for invaluable technical assistance and to the Science Research Council and Nuffield Foundation for financial support.

References

- Beaven, G. H. & Holliday, E. R. (1952) *Advan. Protein Chem.* **7**, 319–386
- Boulter, D. (1973) *Nobel Symp.* **25**, 211–216
- Gray, W. R. & Hartley, B. S. (1963) *Biochem. J.* **89**, 59P
- Graziani, M. T., Agro, A. F., Rotilio, G., Barra, D. & Mondovi, B. (1974) *Biochemistry* **13**, 804–809
- Katoh, S., Shiratori, I. & Takamiya, A. (1962) *J. Biochem. (Tokyo)* **51**, 32–40
- Milne, P. R. & Wells, J. R. E. (1970) *J. Biol. Chem.* **245**, 1566–1574
- Moore, S. (1963) *J. Biol. Chem.* **238**, 235–237
- Moore, S. & Stein, W. H. (1963) *Methods Enzymol.* **6**, 817–831
- 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., Brown, R. H., Scawen, M. D. & Boulter, D. (1973) *Biochim. Biophys. Acta* **303**, 269–273
- 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. & Boulter, D. (1974) *Biochem. J.* **143**, 257–264
- Scawen, M. D., Hewitt, E. J. & James, D. M. (1975) *Phytochemistry* **14**, in the press
- Scawen, M. D., Ramshaw, J. A. M., Brown, R. H. & Boulter, D. (1974) *Eur. J. Biochem.* **44**, 299–303
- Steers, E., Craven, G. R., Anfinsen, C. B. & Bethune, J. L. (1965) *J. Biol. Chem.* **240**, 2478–2484
- Thompson, E. W., Laycock, M. V., Ramshaw, J. A. M. & Boulter, D. (1970) *Biochem. J.* **117**, 183–192