Biochem. J. (1974) **143**, 691–701
Printed in Great Britain

The Amino Acid Sequence of Plastocyanin from French Bean (*Phaseolus vulgaris*)

By P. R. MILNE* and J. R. E. WELLS
Department of Biochemistry, University of Adelaide, Adelaide, S. Austral. 5001, *Australia*
and R. P. AMBLER
Department of Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, U.K.

(*Received* 14 *May* 1974)

The amino acid sequence of the plastocyanin from French bean (*Phaseolus vulgaris*) was determined. The protein consists of a single polypeptide chain of 99 residues, and the sequence was determined by characterization of CNBr, tryptic, chymotryptic and thermolysin peptides. When the sequence is compared with that from the plastocyanin of the unicellular green alga *Chlorella fusca*, the French-bean protein shows the deletion of the *N*-terminal residue, a two residue insertion and 53 identical residues. Detailed evidence for the sequence of the protein has been deposited as Supplementary Publication SUP 50037 (16pp., 1 microfiche) at the British Library (Lending Division) (formerly the National Lending Library for Science and Technology), Boston Spa, Yorks. LS23 7BQ, U.K., from whom copies may be obtained on the terms given in *Biochem. J.* (1973) **131**, 5.

Plastocyanin is a small blue copper protein, first isolated from the unicellular green alga *Chlorella ellipsoidea* by Katoh (1960). It has been found in a wide range of photosynthetic organisms, including uni- and multi-cellular algae (Katoh *et al.*, 1961; de Kouchkovsky & Fork, 1964; Gorman & Levine, 1966) and higher plants (Katoh *et al.*, 1961, 1962; de Kouchkovsky & Fork, 1964; Wells, 1965, 1966; Ramshaw *et al.*, 1973). It has been found in at least one blue-green bacterium (*Anabaena variabilis*; Lightbody & Krogmann, 1967), but not in any of the other classes of photosynthetic bacteria. It does not occur in the non-photosynthetic tissues of plants (Katoh *et al.*, 1961), although it is present in the leaves of etiolated barley seedlings (Plesincar & Bendall, 1970). Plastocyanin is involved in photosynthetic electron transport between photosystems I and II, although the precise details of its involvement are not yet clear (Boardman, 1968; Arnon *et al.*, 1970).

The plastocyanin from French bean (*Phaseolus vulgaris*) was isolated by Wells (1965) during the study of the plant carboxypeptidase phaseolain (Carey & Wells, 1972). French-bean plastocyanin consists of a single polypeptide chain, now known to contain 99 residues, to which one copper atom is bound. It is the smallest copper protein yet reported, with a molecular weight of about 10500 (Milne & Wells, 1970). The complete sequence of the plastocyanin from the green alga *Chlorella fusca* has been briefly reported (Kelly & Ambler, 1973), and the evidence for the amino acid sequence of this 98-residue protein is presented in an accompanying paper (Kelly & Ambler, 1974). The amino acid sequence of potato plastocyanin has been published by Ramshaw *et al.* (1974). French-bean plastocyanin has been crystallized and the tertiary structure is under investigation by X-ray methods (H. Freeman, personal communication).

We have been interested in the structure of plastocyanin for three reasons. First, its size and its involvement in photosynthesis indicate that the amino acid sequence will be useful for the investigation of evolutionary relationships between photosynthetic organisms. Secondly, we are interested to compare the primary structure of plastocyanin with that of the bacterial copper protein azurin (Ambler & Brown, 1967; Ambler, 1971). Preliminary results with spinach plastocyanin (R. P. Ambler, unpublished work) had indicated that the amino acid sequence around the single cysteine residue was similar to the sequence around the only free cysteine residue of the azurins. Evidence has been presented that the single free cysteine residues of both azurin and plastocyanin are involved in the binding of copper to the polypeptide chain (Katoh & Takamiya, 1964; Finazzi-Agro *et al.*, 1970). Finally, the small size of plastocyanin makes it appropriate for investigating the structural features that give rise to the unique blue colour and spectral characteristics of Type I Cu^{2+} proteins. All intensely blue copper proteins, such as caeruloplasmin, laccase, stellacyanin, azurin and plastocyanin, contain some copper in the 'Type I' form (Rist *et al.*, 1970; Andréasson & Vänngård, 1970;

* Present address: United Faculty of Theology, Ormond College, Parkville, Victoria 3052, Australia.

Malmström *et al.*, 1968). Plastocyanin, azurin and stellacyanin contain copper only in the 'Type I' form, and the Type I Cu^{2+} complex is apparently solely responsible for the unusual spectral properties of the intensely blue copper proteins. Although it has been the subject of much work and speculation, the identity of the copper-binding ligands has not yet been determined in any copper protein.

In this paper evidence is presented for the primary structure of the plastocyanin from French bean. Methods for preparing the protein, details of its properties and partial sequence studies have been published (Wells, 1965, 1966; Milne & Wells, 1970; Milne, 1971).

Experimental

The sequence work on French-bean plastocyanin was started in Adelaide and finished in Edinburgh. The methods and materials described below refer to the Australian part unless otherwise stated. The Edinburgh sequence experiments used protein prepared in Australia, examined by methods similar to those used by Kelly & Ambler (1974).

Materials

Plants. French bean (*Phaseolus vulgaris*) plants were obtained from a local Adelaide market gardener, and harvested when 6–8 weeks old.

Chromatographic materials. Whatman DEAE-cellulose DE-50 was obtained from W. & R. Balston Ltd., Maidstone, Kent, U.K. Sephadex was from Pharmacia, Uppsala, Sweden. Polyamide layers were from Cheng Chin Trading Co., Taipei, Taiwan.

Enzymes. 1-Chloro-4-phenyl-3-L-tosylamido-butan-2-one (TPCK)-treated trypsin (TRTPCK 8DA and 8GA), chymotrypsin (CC 16148-9) and carboxypeptidase A (COA-DFP 6138) were from Worthington Biochemical Corp., Freehold, N.J., U.S.A.

Methods

Isolation of plastocyanin. Plastocyanin was isolated as previously described (Milne & Wells, 1970). The usual yield was 60–120mg of plastocyanin/10 litres of bean sap. This material was homogeneous by polyacrylamide-gel electrophoresis and *N*-terminal analysis, but to remove contaminating polyphenolic compounds, the preparations were normally further purified by repeated chromatography (Milne & Wells, 1970), until an absorption index value (E_{278}/E_{597}; Katoh *et al.*, 1962) of 1.1–1.2 was obtained.

The amount of native plastocyanin in preparations could be determined from the molar extinction coefficient of the copper chromophore at 597nm ($\varepsilon = 4.5 \times 10^3$ litre·mol^{-1}·cm^{-1}; Milne & Wells, 1970). Quantities of apoplastocyanin and carboxymethyl-plastocyanin were determined from the molar extinction coefficient at 278nm (5.0×10^3 litre·mol^{-1}·cm^{-1}) or by hydrolysis and amino acid analysis.

Carboxymethylation of plastocyanin. The method described by Milne & Wells (1970) was used.

Apoplastocyanin. Apoplastocyanin was prepared by trichloroacetic acid precipitation, as described by Kelly & Ambler (1974).

Amino acid analysis. In Australia, peptides were hydrolysed at 110°C for 20–24h under vacuum in 1ml of constant-boiling HCl. Large peptides were also analysed after longer times of hydrolysis, and from these analyses correction factors for the destruction/24h for threonine (6%) and serine (11%) were deduced. In all the later hydrolyses, thioglycollic acid [10μl of 10% (v/v) aqueous solution] was included, to decrease hydrolysis loss of tyrosine, methionine and *S*-carboxymethyl-cysteine (Sanger & Thompson, 1963; Dus *et al.*, 1970). In Scotland, peptides were hydrolysed at 105°C for 24h, and no correction made for serine or threonine destruction. Peptides which were believed to contain stable peptide bonds (Val-Val, Ile-Ile etc.) were generally also analysed after 96h hydrolysis.

Fractionation of tryptic peptides. Carboxymethyl-plastocyanin (3μmol) was digested with 0.6mg of trypsin (batch TRTPCK 8DA) in 3ml of 0.2M-NH_4HCO_3 at 37°C for 2h. The digest was then at once passed through a column (110cm×2cm diam.) of Sephadex G-50 (fine grade) equilibrated in 0.1M-NH_3 solution at room temperature, with a flow rate of 20ml/h, and fractions of 5ml volume were collected. Fractions were pooled on the basis of the u.v.-absorption profiles at 220nm and 276nm, and the examination of selected samples by high-voltage paper electrophoresis and *N*-terminal analysis. The smaller peptides were purified by high-voltage paper electrophoresis. The larger peptides were fractionated by chromatography on DEAE-cellulose (DE-50, 20cm×1.2cm diam., starting buffer 0.02M-sodium phosphate, pH 7.2, eluted with a 250ml linear gradient of NaCl from 0–1.0M), and finally purified by re-chromatography on a similar column with a 0–0.4M-NaCl linear gradient. The peptides were then desalted by gel filtration through Sephadex G-50 into 0.1M-NH_3 solution.

CNBr peptides. The three peptides which resulted from CNBr cleavage of carboxymethyl-plastocyanin were isolated as described by Milne & Wells (1970).

N-Terminal analysis of protein and peptides and dansyl–phenyl isothiocyanate degradation of peptides. The dansyl method (Gray, 1972*a*) was used for qualitative investigations of protein and peptides. The conditions for phenyl isothiocyanate degradations were as described by Gray (1972*b*). In Australia, dansyl-amino acids were identified by t.l.c. on 3.75cm square polyamide layers (as described by Hartley,

1970). In Scotland, primary identification was by high-voltage paper electrophoresis at pH 4.38, with confirmation by other paper systems (Ambler & Wynn, 1973).

Quantities of material used

Adelaide experiments. All experiments used carboxymethyl-plastocyanin. The amounts used were: tryptic digest I, 9 μmol; tryptic digest II (with high-specificity trypsin), 3 μmol; CNBr cleavage, 4 μmol; chymotryptic digest I, 2 μmol.

Edinburgh experiments. The amounts used were: thermolysin digest, 3 μmol of carboxymethyl-plastocyanin; chymotryptic digest II, 3 μmol of apoplastocyanin; tryptic digest III, 3 μmol of apoplastocyanin. Quantities of peptides used for analyses and sequence determination experiments were in the ranges given by Ambler & Wynn (1973).

Results

Previous work (Milne & Wells, 1970) had shown that French-bean plastocyanin consisted of a single polypeptide chain of about 100 amino acids to which was bound one copper atom. The protein had been shown to have the *N*-terminal sequence Leu-Glx-Val-Leu-, and to be split into three fragments by CNBr. These fragments were considered to contain 58, 35 and 7 residues, and the longest was *N*-terminal and the shortest *C*-terminal. The complete amino acid sequence of the protein has now been determined, and the proposed sequence is shown in Fig. 1.

When the sequence of the French-bean plastocyanin is compared with that of *Chlorella fusca* plastocyanin (Fig. 2), the algal protein is found to have an extra residue at the *N*-terminus. To enable corresponding residues to be given the same number, the *N*-terminal residue of the French-bean protein is numbered 2 and so on in the rest of this paper.

Experimental approach

The initial sequence work was done in Adelaide. The protein contains no arginine and only 5 residues of lysine, so only 6 peptides are formed when the protein is digested with good trypsin. As the CNBr peptides X-1, X-2 and X-3 contained respectively 3, 1 and 1 residues of lysine, all the 6 tryptic peptides except T-2 and T-3 could be ordered by using simply the *N*-terminal and amino acid analysis results for the CNBr and tryptic peptides (Milne, 1971). In a preliminary experiment (trypsin digest I), with unsatisfactory trypsin (batch TRTPCK 8GA), several chymotrypsin-susceptible bonds were split, and secondary cleavage products of peptides T-1, T-4, and T-5 were isolated. The sequences of several of these peptides were either partially or completely determined by the dansyl–phenyl isothiocyanate method. The isolation of some lysine-containing chymotryptic peptides enabled the relative order of peptides T-2 and T-3 to be determined. These experiments gave sequence information on 66 of the 98 peptide bonds in the molecule, residues 2–11, 20–36, 55–64 and 72–100 (Milne, 1971). The sequence of residues 2–26 was subsequently determined with an automatic sequenator (P. Edman & G. Begg, personal communication).

The Adelaide sequence work was then completed in Edinburgh, by a thorough characterization of the peptides from a thermolysin digest of carboxymethyl-plastocyanin, and by a tryptic digest (III) and a chymotryptic digest (II) of apoplastocyanin.

The amino acid composition of French-bean plastocyanin is shown in Table 1, which shows the values taken from the sequence (Fig. 1) together with the values deduced from a series of hydrolysis experiments (Milne & Wells, 1970). The values are in very good agreement except for one extra glutamic acid residue in the acid hydrolysis results. This may be due to an erroneously low colour factor for glutamic acid having been used, as this amino acid is the least stable component during storage of amino acid analyser calibration solutions.

Amide groups were assigned from a consideration of the electrophoretic mobilities at pH 6.5 of simple peptides containing the residue under examination. The results of carboxypeptidase digestion of appropriate peptides, and, for residues 2–26, the sequenator results (P. Edman & G. Begg, personal communication) were also used for these assignments.

Peptide nomenclature

Peptides are identified by a capital letter indicating the method used for the degradation of the protein, the letters used being: T, trypsin; X, CNBr; C, chymotrypsin; H, thermolysin; N, papain; S, subtilisin. For the tryptic and CNBr peptides, the letter is followed by a number indicating the order of the peptide in the polypeptide chain (thus peptide T-1 is the *N*-terminal tryptic peptide). A lower-case letter after the number for a tryptic peptide indicates that the peptide was formed by a secondary enzymic activity in the trypsin used, with peptide 'a' forming the *N*-terminal part of the primary peptide. For the more complex chymotryptic and thermolysin digests, and for the products of secondary digests with different enzymes, the nomenclature of Ambler & Wynn (1973) is used.

Evidence for the proposed sequence

The evidence for the proposed amino acid sequence is summarized in Fig. 1. This shows all the peptides isolated from each digest, together with peptides

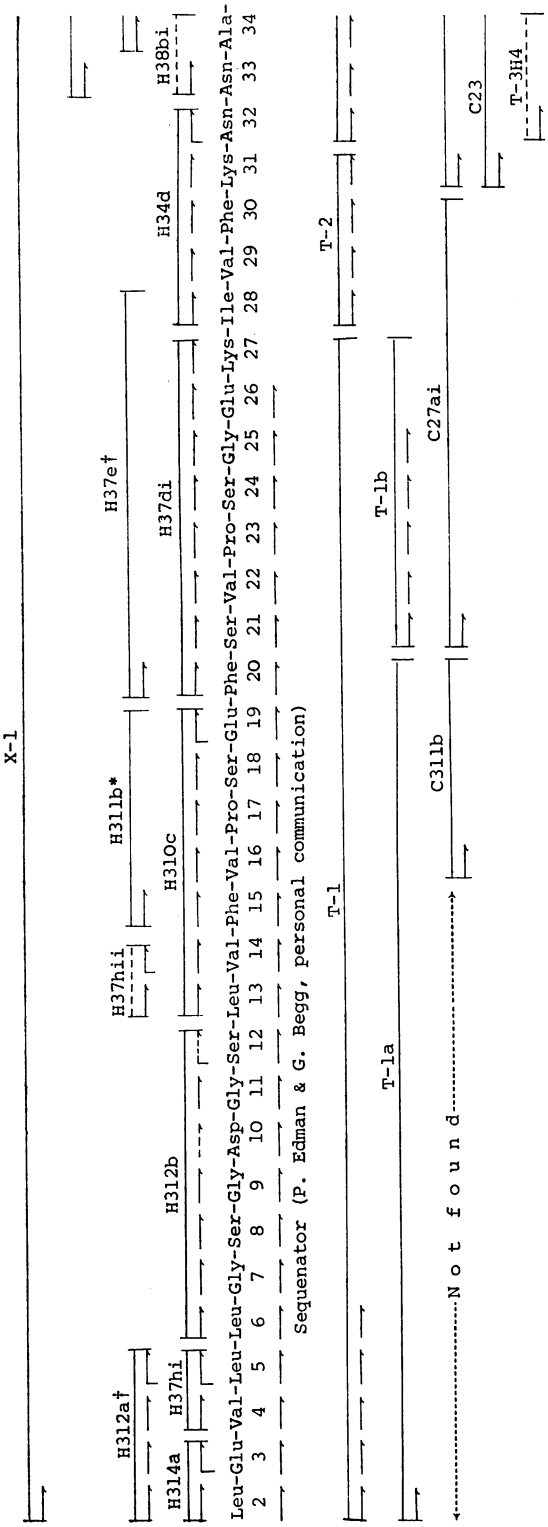

Fig. 1. *Amino acid sequence of French-bean plastocyanin*

Peptides derived by CNBr cleavage (X) and thermolysin digestion (H) are shown above the sequence, and by tryptic (T) and chymotryptic (C) digestion below the sequence. Vertical arrows show peptides formed by further digestion. For each digest, peptides formed in best yield are shown nearest the sequence. Full lines indicate quantitative amino acid analyses (substandard if marked *). Broken lines indicate qualitative analyses. Under the sequence, — indicates residues identified (P. Edman & G. Begg, personal communication) after degradation of the protein in a sequenator. Under peptide lines, → indicates end-groups and subsequent residues revealed by phenyl isothiocyanate degradation, and identified by the dansyl method (substandard if shown as - -›). ⌐→ indicates final residues of peptides identified as the free amino acid after removal of the remainder by phenyl isothiocyanate degradation. Peptides marked † were examined by carboxypeptidase A digestion. References to the text are given near regions that are specifically discussed. In the sequence, the *N*-terminal residue is numbered 2, so that corresponding residues in the French bean and *Chlorella fusca* plastocyanins should have the same numbers (Fig. 2).

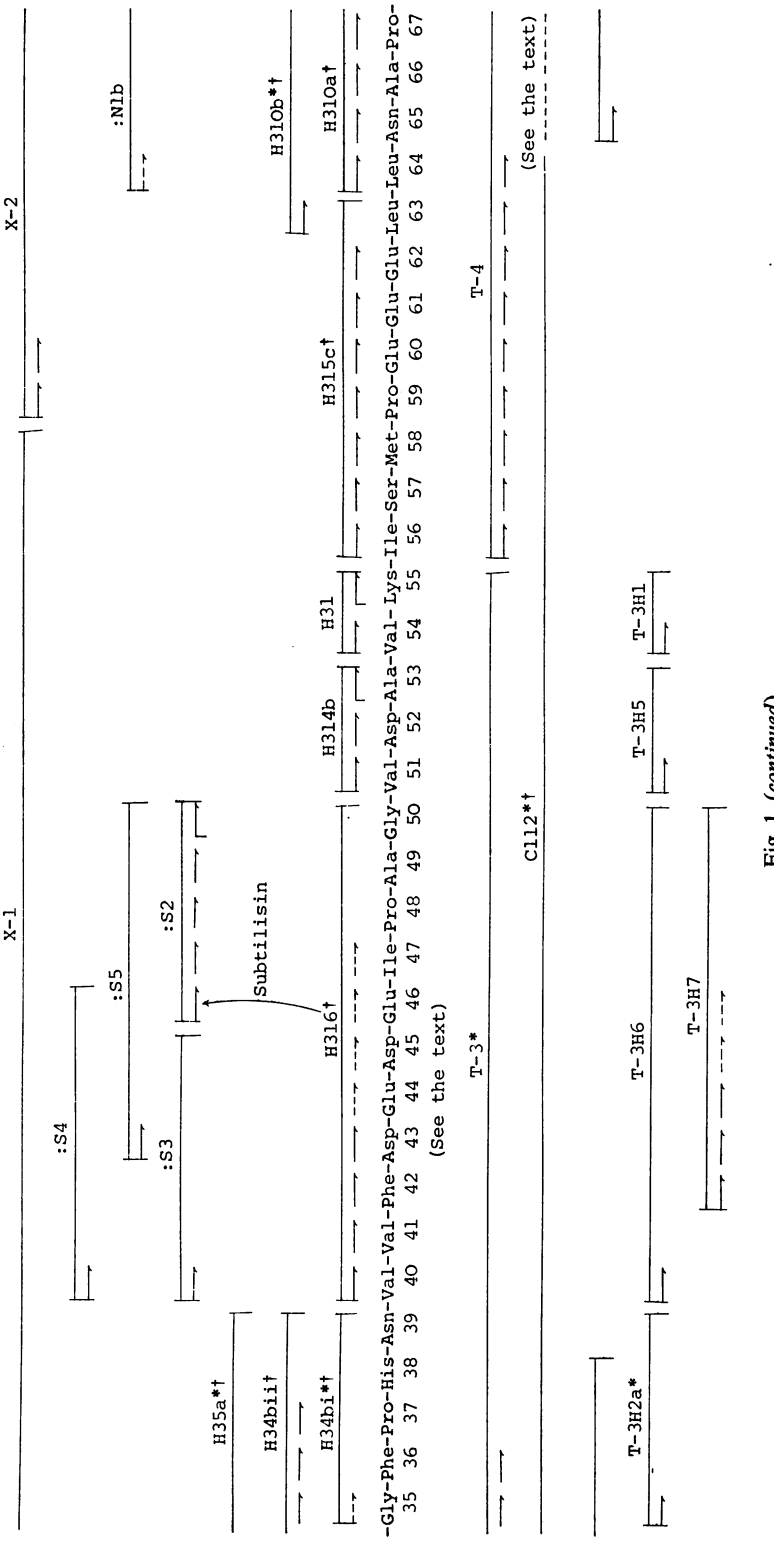

Fig. 1. *(continued)*

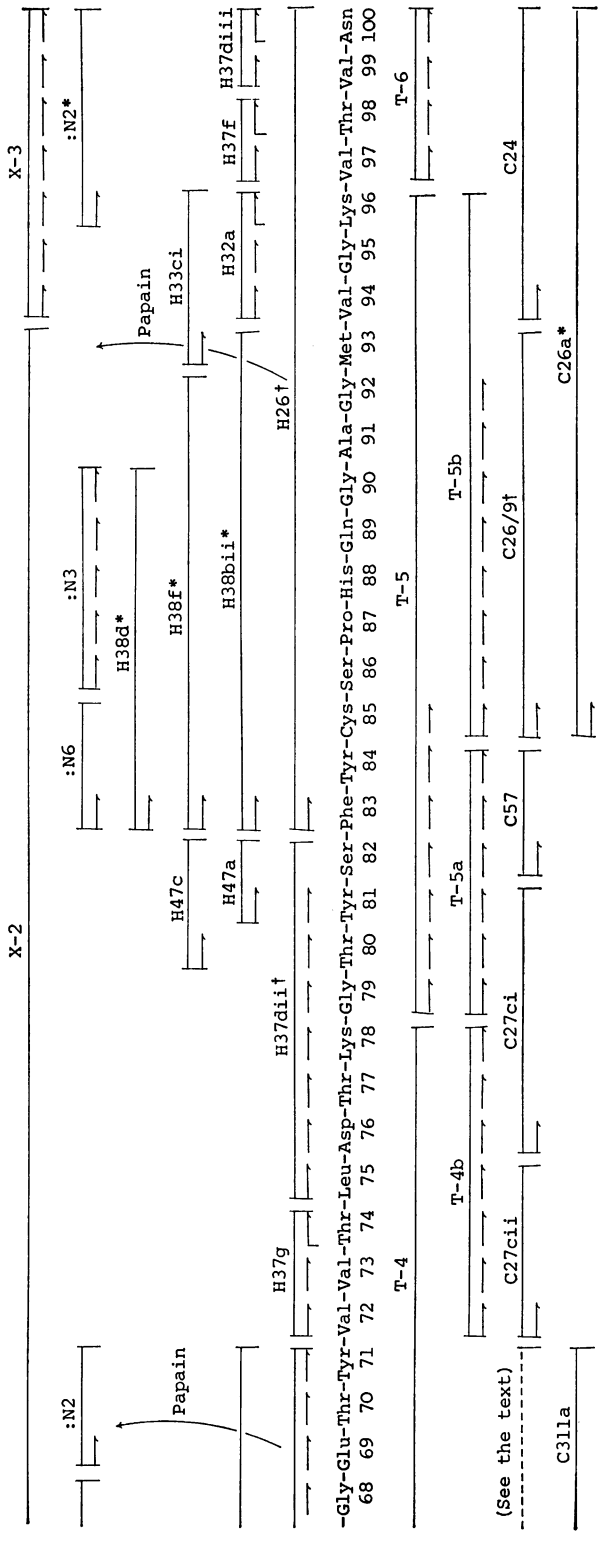

Fig. 1. (*continued*)

```
     1  2  3  4  5  6  7  8  9  10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
(a)     Leu-Glu-VAL-Leu-LEU-GLY-Ser-Gly-Asp-GLY-Ser-LEU-Val-PHE-Val-PRO-Ser-Glu-Phe-Ser-Val-Pro-Ser-GLY-GLU-Lys-Ile-
(b)     Leu-Asp-VAL-Leu-LEU-GLY-Gly-Asp-Asp-GLY-Ser-LEU-Ala-PHE-Ile-PRO-Gly-Asn-Phe-Ser-Val-Ser-Ala-GLY-GLU-Lys-Ile-
(c) Asp-Val-Thr-VAL-Lys-LEU-GLY-Ala-Asp-Ser-GLY-Ala-LEU-Val-PHE-Glu-PRO-Ser-Ser-Val-Thr-Ile-Lys-Ala-GLY-GLU-Thr-Val-

     29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56
(a) -Val-Phe-Lys-ASN-ASN-ALA-GLY-PHE-PRO-HIS-ASN-Val-VAL-PHE-ASP-GLU-ASP-GLU-Ile-PRO-Ala-GLY-Val-Asp-ALA-Val-Lys-Ile-
(b) -Thr-Phe-Lys-ASN-ASN-ALA-GLY-PHE-PRO-HIS-ASN-Val-VAL-PHE-ASP-GLU-ASP-GLU-Ile-PRO-Ala-GLY-Val-Asp-ALA-Ser-Lys-Ile-
(c) -Thr-Trp-Val-ASN-ASN-ALA-GLY-PHE-PRO-HIS-ASN-Ile-VAL-PHE-ASP-GLU-ASP-GLU-Val-PRO-Ser-GLY-Ala-Asn-ALA-Glu-Ala-Leu-

     57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84
(a) -SER-Met-Pro-Glu-GLU-Glu-Leu-LEU-ASN-ALA-Pro-GLY-GLU-Thr-TYR-Val-Val-Thr-Leu-Asp-Thr-Lys-GLY-THR-TYR-Ser-Phe-Tyr-
(b) -SER-Met-Ala-Glu-GLU-Asp-Leu-LEU-ASN-ALA-Ala-GLY-GLU-Thr-TYR-Ser-Val-Thr-Leu-Ser-Glu-Lys-GLY-THR-TYR-Thr-Phe-Tyr-
(c) -SER-----His-----GLU-Asp-Tyr-LEU-ASN-ALA-Pro-GLY-GLU-Ser-TYR-Ser-Ala-Lys-Phe-Asp-Thr-Ala-GLY-THR-TYR-Gly-Tyr-Phe-

     85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
(a) -CYS-Ser-PRO-HIS-GLN-GLY-ALA-GLY-MET-Val-GLY-Lys-Val-THR-VAL-Asn
(b) -CYS-Ala-PRO-HIS-GLN-GLY-ALA-GLY-MET-Val-GLY-Lys-Val-THR-VAL-Asn
(c) -CYS-Glu-PRO-HIS-GLN-GLY-ALA-GLY-MET-Lys-GLY-Thr-Ile-THR-VAL-Gln
```

Fig. 2. *Comparison of amino acid sequences of plastocyanins from* (*a*) *French bean,* (*b*) *potato* (*Ramshaw et al.*, 1974) *and* (*c*) *Chlorella fusca* (*Kelly & Ambler*, 1974)
Residues that are the same in all three sequences are shown in capitals.

Table 1. *Amino acid composition of French-bean plastocyanin*

Results are shown as residues per molecule. Analysis values (from Milne & Wells, 1970) were the average of 21, 48 and 69h hydrolyses at 110°C, except for the following amino acids: serine and threonine values were derived by extrapolation to zero time of hydrolysis; for valine and isoleucine, the 69h values were taken; cysteine (as cysteic acid) was from an independent hydrolysis of a performic acid-oxidized sample; methionine was from an independent hydrolysis with HCl containing thioglycollic acid, the result from the average of the first three hydrolyses being 1.4 residues per molecule; tryptophan was measured in the intact protein by the method of Spies & Chambers (1949).

	Analysis	Sequence (Fig. 1)
Glycine	11.0	11
Alanine	5.1	5
Valine	13.8	14
Leucine	6.8	7
Isoleucine	2.9	3
Serine	8.0	8
Threonine	5.1	5
Aspartic acid	10.2	5
Asparagine		5
Glutamic acid	10.7	9
Glutamine		1
Phenylalanine	6.1	6
Tyrosine	2.9	3
Tryptophan	0.0	0
Cysteine	1.0	1
Methionine	1.8	2
Proline	7.0	7
Lysine	5.0	5
Histidine	1.9	2
Arginine	0.0	0
Total		99

formed by further digestion of selected peptides with other enzymes. Symbols show how much of the sequence of each peptide has been determined by the dansyl–phenyl isothiocyanate method, as well as peptides that were examined with carboxypeptidase. Cases in which the results of dansyl–phenyl isothiocyanate degradation or quantitative amino acid analysis were not considered wholly satisfactory are indicated.

Dansyl–phenyl isothiocyanate degradation results were considered unsatisfactory when more than one dansyl-amino acid was identified as present in significant amount (subjective visual estimate of more than 20%) at a step of the degradation, or when for some reason or other (such as accident or shortage of material) the dansyl-amino acid was not positively identified.

The criteria for satisfactory amino acid analyses are: (1) that no impurity should be present in amount larger than 0.20 mol/mol and (2) that when the relative amounts of amino acids present are calculated on the basis that the average amount is integral, no values should fall outside the limits 0.8–1.2, 1.8–2.2, 2.7–3.3, 3.7–4.3 etc. The peptides that did not meet these analytical criteria (marked * on Fig. 1) failed for the following reasons:

(1) Peptide C112 (mixture of residues 31–64 and 31–71): analysis gave Gly 2.3, Ala 3.3, Val 4.0, Leu 2.0, Ile 2.0, Ser 0.8, Asp 6.3, Glu 5.5, Phe 2.0, Met 0.9, Pro 3.0, Lys 1.9, His 0.9 (Thr 0.26, Tyr 0.19), for hydrolysis for 96h at 105°C. The amino acid analysis and the amounts of amino acids released by carboxypeptidase digestion (Leu 1.5, Thr 0.31, Tyr 0.23, Val 0.13; no Glu or Gln) are consistent with the fraction being a mixture of about 75% of a peptide comprising residues 31–64 and 25% of a peptide comprising residues 31–71. The extra *C*-terminal sequence of the latter peptide (also found as peptide C311a) was ...Asn-Ala-Pro-Gly-Glu-Thr-Tyr. These two big peptides would be expected to have very similar electrophoretic properties. As this part of the sequence had been well characterized by other experiments, it was not considered necessary completely to purify these peptides.

(2) Peptide T-3 (residues 32–55). From other evidence this peptide should contain four valine residues, including the sequence -Val-Val- at positions 40/41. Both the Australian and the Scottish analyses were of samples hydrolysed for only 24h. The former (hydrolysed at 110°C) showed Val 3.6, and also failed the criteria by showing Glu 2.3 (=2), whereas the latter (at 105°C) showed Val 3.2.

(3) Peptide X-1 (residues 2–58; Milne & Wells, 1970, peptide CNBr-1): the analysis showed Glu 5.8, whereas all the other sequence evidence for this region indicated that there were only five residues of glutamic acid or glutamine. The only other piece of evidence that supports this analysis is the glutamic acid value for hydrolyses of the whole protein (Table 1 and Milne & Wells, 1970). A possible explanation for these high recoveries of glutamic acid has been suggested above.

(4) A few peptides slightly transgressed the stoicheiometry or impurity criteria. Most of these offenders were minor peptides isolated in low yield, or the products of secondary digestions, but peptide T-5, a major tryptic peptide, showed Ser 1.7 (=2). The other offenders were peptides: H34bi (Ala 0.34 =0); H311b (Pro 1.3 =1); H310b (Ala 0.42 =0, Ser 0.24 =0, Phe 0.23 =0); T-1b (Ser 1.7 =2); T-5b (Pro 1.3 =1); T-3H2a (Ser 0.21 =0); H26N2 (Gly 0.21 =0, Ser 0.28 =0). The worst case, peptide H310b, could not be simply explained as contamination with a single known peptide of similar properties. The analysis for peptide H35a was satisfactory except that, through an instrumental fault, no quantitative value was obtained for histidine.

(5) Several peptides derived from carboxymethylplastocyanin gave bad analysis values for *S*-carboxymethyl-cysteine (Cmc) and methionine. Thus peptide X-2 showed Cmc 0.6 (=1), T-5b Cmc 0.5 (=1) (also Pro 1.3 =1; see above). In some cases both cysteine and *S*-carboxymethyl-cysteine, and both methionine and *S*-carboxymethyl-methionine were detected, although the combined values still did not give an adequate recovery. These peptides were: H38f (Cys+Cmc 0.6 =1); H38bii (Cys+Cmc 0.34 =1; Met+*S*-carboxymethyl-methionine 0.7 =1). Peptide C26a gave an extremely low recovery of methionine (0.3 =1), as well as defaulting values of Gly (2.6 =3), Lys (0.7 =1) and His (0.7 =1). Completely satisfactory values were obtained for the peptide (C26/9) from the same region of the sequence from chymotryptic digest II, in which apoplastocyanin was digested, and the peptide oxidized with performic acid vapour on paper in the course of its purification.

Supplementary publication

Detailed evidence for the amino acid sequence of the protein has been deposited as Supplementary Publication SUP 50037 with the British Library (Lending Division) for storage on microfiche. The evidence comprises:

(1) Tables showing the properties of all the peptides in Fig. 1. The successive steps used in the purification of each peptide are given, and where applicable the values of V/V_0 [elution volume/void volume for gel filtration through Sephadex G-25 in 5% (v/v) formic acid] and electrophoretic mobility on paper at pH6.5. Absolute percentage yields are given for peptides from primary digests, and relative yields for peptides from secondary digests. Amino acid analyses are given for all peptides, which show all impurities present in amounts greater than 0.1 mol/mol peptide.

(2) Further tables show the individual sequence evidence for each peptide. This consists of the *N*-terminal group analysis results, and the results of dansyl–phenyl isothiocyanate degradation experiments. Details are given when these results are substandard.

(3) The evidence for the presence or absence of amide groups on each aspartic acid or glutamic acid residue in the protein is given.

(4) Details are given of experiments with carboxypeptidase A on the peptides marked † in Fig. 1, and of the effect of phaseolain [a carboxypeptidase C (Carey & Wells, 1972)] on the whole protein.

Discussion

Problems in the elucidation of the sequence

The determination of the amino acid sequence of the French-bean plastocyanin presented few major problems, as adequate amounts of protein of high purity were available. The protein contained sufficient well-situated residues of methionine (two) and lysine (five) to allow the isolation and characterization of all the CNBr and tryptic peptides. This was in contrast with the *Chlorella fusca* protein (Kelly & Ambler, 1974), which had to be investigated by chymotryptic and thermolysin digestion.

At the beginning of the investigation there was difficulty in getting satisfactory tryptic digestion, as the batch of commercial chymotrypsin-inhibitor-treated trypsin available had much chymotrypsin-like activity, so yields of the larger tryptic peptides were low. This may have been due to actual contamination with chymotrypsin, or to the presence of the autolysis product pseudotrypsin (Keil-Dlouhá *et al.*, 1971). A second batch of the same product proved satisfactory, and adequate results were also obtained with trypsin that had been treated before use with diphenylcarbamyl chloride (Erlanger & Cohen, 1962), even though the crystalline trypsin had been stored for several years at 4°C with no special attempts made to keep it dry.

French-bean plastocyanin lacks arginine, and contains only a low proportion of lysine, so some of the tryptic peptides are large and cannot be separated on paper in good yield. Preliminary electrophoretic experiments showed that these large peptides all had net negative charges at pH6.5, and so would be expected to be separable on an anion-exchange resin. This proved to be the case, and after initial removal of the small peptides by gel filtration, the large peptides were readily purified in satisfactory yield by chromatography on DEAE-cellulose eluted with a linear NaCl gradient. The purification of the CNBr peptides has been reported elsewhere (Milne & Wells, 1970). After an initial gel filtration, the smaller tryptic peptides, the thermolysin peptides and most of the expected chymotryptic peptides could be fractionated satisfactorily by paper methods. The final tryptic digest (III) was also fractionated by paper methods, but the yields of two of the three large peptides were much lower than by the ion-exchange method. In digest III, peptide T-1 was recovered in 7% yield and peptide T-4 in 5% yield, compared with 53% and 40% respectively in digest II. Surprisingly, the similar-sized peptide T-3 was recovered in 21% yield in digest III compared with only 16% by the ion-exchange method. The *N*-terminal chymotryptic peptide (residues 2–15) was not isolated, presumably because it was too hydrophobic to separate on paper despite its small size.

Some difficulty was experienced in the investigation of the very acidic regions of the sequence. It proved difficult to get unambiguous results for the region of residues 43–46 (Asp-Glu-Asp-Glu), as the phenyl isothiocyanate degradation tended to get ragged through this sequence. It was necessary to study peptides from secondary digests of peptide H316

(residues 40–50) and T-3 (residues 32–55) to establish this sequence conclusively. The proof of absence of amides in this sequence and for residues 60–62 (Glu-Glu-Glu) also posed problems. Evidence for the presence or absence of amide groups was generally based on the qualitative mobility at pH 6.5 of peptides containing only a single acidic residue, but this was not possible for these multiple acidic residue sequences. In these cases the mobilities of peptides formed by phenyl isothiocyanate degradation of peptides T-3H7 (residues 42–50) and H315c (residues 56–63) were measured, and the residues identified as acidic by the progressive decrease in mobility of the residual peptides as they were successively removed. The quantitative mobility values were compared with predicted values (Offord, 1966), but no assignments were made solely on the basis of such numerical results.

Accuracy of proposed sequence

Most of the peptide bonds in the sequence have been established by phenyl isothiocyanate degradation, many of them in two or more independent experiments. The method appears to have a generally high reliability, but errors (especially transpositions) are always possible, particularly from mislabelling. Except where specifically indicated, dansyl-amino acids were positively identified, no reliance being put on the known amino acid composition of the peptide under investigation to distinguish between dansyl-amino acids of similar chromatographic or electrophoretic properties.

The peptide bonds not established by this method were 27–28, 37–38–39–40, 50–51, 53–54, 55–56, 71–72 and 92–93–94. The evidence for these bonds is primarily from amino acid compositions, in particular the composition differences between related peptides (e.g. between peptides H37e and H37di, and between peptides T-1b and C27ai for bond 27–28). Evidence from carboxypeptidase digestion was also used (e.g. to provide evidence for bonds 50–51 and 53–54 by the amino acids released from peptide T-3). The evidence for all the bonds seems to be adequate, and no region of the sequence can be singled out as being particularly likely to contain an error.

As mentioned above, the amino acid composition of the protein derived from hydrolysis experiments agrees well with that from the sequence (Table 1), apart from an extra residue of glutamic acid detected by hydrolysis in the whole plastocyanin and in the CNBr peptide X-1. The discrepancy is not much larger than the expected errors in the analytical method, and, as suggested above, may be due to inaccuracy in the amino acid analyser calibration mixture. There is no obvious site in the sequence into which an extra residue of glutamic acid or glutamine could be fitted that would not be inconsistent with sequence determination experiments and with what is known about protease specificity.

Comparison of known plastocyanin sequences

The amino acid sequences of the plastocyanins from potato (*Solanum tuberosum*; Ramshaw *et al.*, 1974) and the unicellular green alga *Chlorella fusca* (Kelly & Ambler, 1973, 1974) are known (Fig. 2). The potato protein is of exactly the same length as that from French bean, and 80 of the 99 residues are identical. As compared with the higher plant proteins, the algal protein contains an extra residue at the *N*-terminus, and also has a two residue deletion at about residues 58–59. The plastocyanins show some sequence similarity to the bacterial azurins (Ambler, 1971; Ambler & Brown, 1967) around the single free cysteine residue.

We are very grateful to Professor Hans Freeman, School of Chemistry, University of Sydney, for the gift of 100 mg of plastocyanin which enabled the sequence to be finished. The isolation of plastocyanin in Sydney was made possible by research grant 65/1552 from the Australian Research Grants Committee to H. C. Freeman. We also thank Professor E. O. P. Thompson for the gift of the good trypsin. P. R. M. was the recipient of a C.S.I.R.O. Scholarship. R. P. A. was partially supported by the (British) Medical Research Council.

References

Ambler, R. P. (1971) in *Recent Developments in the Chemical Study of Protein Structures* (Previero, A., Pechère, J.-F. & Coletti-Previero, M.-A., eds.), pp. 289–305, Inserm, Paris

Ambler, R. P. & Brown, L. H. (1967) *Biochem. J.* **104**, 784–825

Ambler, R. P. & Wynn, M. (1973) *Biochem. J.* **131**, 485–498

Andréasson, L.-E. & Vänngård, T. (1970) *Biochim. Biophys. Acta* **200**, 247–257

Arnon, D. I., Chain, R. K., McSwain, B. D., Tsujimoto, H. Y. & Knaff, D. B. (1970) *Proc. Nat. Acad. Sci. U.S.* **67**, 1404–1409

Boardman, N. K. (1968) *Advan. Enzymol. Relat. Areas Mol. Biol.* **30**, 1–79

Carey, W. F. & Wells, J. R. E. (1972) *J. Biol. Chem.* **247**, 5573–5579

de Kouchkovsky, Y. & Fork, D. C. (1964) *Proc. Nat. Acad. Sci. U.S.* **52**, 232–239

Dus, K., Flatmark, T., de Klerk, H. & Kamen, M. D. (1970) *Biochemistry* **9**, 1984–1996

Erlanger, B. F. & Cohen, W. (1962) *J. Amer. Chem. Soc.* **85**, 348–349

Finazzi-Agro, F., Rotilio, G., Avigliano, L., Guerrieri, P., Boffi, V. & Mondovi, B. (1970) *Biochemistry* **9**, 2009–2014

Gorman, D. S. & Levine, R. P. (1966) *Plant Physiol.* **41**, 1637–1642

Gray, W. R. (1972*a*) *Methods Enzymol.* **25**, 121–138

Gray, W. R. (1972*b*) *Methods Enzymol.* **25**, 333–344

Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822

Katoh, S. (1960) *Nature (London)* **186**, 533–534

Katoh, S. & Takamiya, A. (1964) *J. Biochem.* (*Tokyo*) **55**, 378–387
Katoh, S., Suga, I., Shiratori, I. & Takamiya, A. (1961) *Arch. Biochem. Biophys.* **94**, 136–141
Katoh, S., Shiratori, I. & Takamiya, A. (1962) *J. Biochem.* (*Tokyo*) **51**, 32–40
Keil-Dlouhá, V., Zylber, N., Imhoff, J.-M., Tong, N.-T. & Keil, B. (1971) *FEBS Lett.* **16**, 291–295
Kelly, J. & Ambler, R. P. (1973) *Biochem. Soc. Trans.* **1**, 164–166
Kelly, J. & Ambler, R. P. (1974) *Biochem. J.* **143**, 681–690
Lightbody, J. J. & Krogmann, D. W. (1967) *Biochim. Biophys. Acta* **131**, 508–515
Malmström, B. G., Reinhammar, B. & Vänngård, T. (1968) *Biochim. Biophys. Acta* **156**, 67–76
Milne, P. R. (1971) Ph.D. Thesis, University of Adelaide
Milne, P. R. & Wells, J. R. E. (1970) *J. Biol. Chem.* **245**, 1566–1574
Offord, R. E. (1966) *Nature* (*London*) **211**, 591–593
Plesincar, M. & Bendall, D. S. (1970) *Biochim. Biophys. Acta* **216**, 192–199
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. (1974) *Biochem. J.* **139**, 583–592
Rist, G. H., Hyde, J. S. & Vänngård, T. (1970) *Proc. Nat. Acad. Sci. U.S.* **67**, 79–86
Sanger, F. & Thompson, E. O. P. (1963) *Biochim. Biophys. Acta* **71**, 468–471
Spies, J. R. & Chambers, D. C. (1948) *Anal. Chem.* **20**, 30–39
Wells, J. R. E. (1965) *Biochem. J.* **97**, 228–235
Wells, J. R. E. (1966) *Nature* (*London*) **212**, 896–898