Modification of the Automated Sequence Determination as Applied to the Sequence Determination of the *Spirulina maxima* Ferredoxin†

Masaru Tanaka, Mitsuru Haniu, Kerry T. Yasunobu,* Krishna K. Rao, and David O. Hall

ABSTRACT: The amino acid sequence of the *Spirulina maxima* ferredoxin was shown to be: H_2N-Ala-Thr-Tyr-Lys-Val-Thr-Leu-Ile-Ser-Glu-Ala-Glu-Gly-Ile-Asn-Glu-Thr-Ile-Asp-Cys-Asp-Asp-Asp-Thr-Tyr-Ile-Leu-Asp-Ala-Ala-Glu-Glu-Ala-Gly-Leu-Asp-Leu-Pro-Tyr-Ser-Cys-Arg-Ala-Gly-Ala-Cys-Ser-Thr-Cys-Ala-Gly-Lys-Ile-Thr-Ser-Gly-Ser-Ile-Asp-Gln-Ser-Asp-Gln-Ser-Phe-Leu-Asp-Asp-Asp-Gln-Ile-Glu-Ala-Gly-Tyr-Val-Leu-Thr-Cys-Val-Ala-Tyr-Pro-Thr-Ser-Asp-Cys-Thr-Ile-Gln-Thr-His-Gln-Glu-Glu-Gly-Leu-Tyr-COOH. The *S. maxima* ferredoxin is the first procaryote ferredoxin of the plant-algal type to be reported. A modification of the automated sequence determination of a peptide, which was extracted by the organic solvents used to remove excess reagents and the amino acid thiazoline, was utilized to complete the sequence of a 36 residue tryptic peptide.

Ferredoxins are iron-sulfur proteins which act as redox proteins in bacteria, plants, and algae (Lovenberg, 1973). For purposes of establishing structure-function relationships, for providing essential structural information for crystal x-ray diffractionists, and for obtaining evolutionary and genetic data, our laboratory has embarked on a program to examine the sequence of a large number of different iron-sulfur proteins (Yasunobu and Tanaka, 1973). In the present investigation, the goal of the project was to determine the amino acid sequence of the *Spirulina maxima* ferredoxin. This blue-green algae is a procaryote. The other ferredoxins of the plant-algal type which have been sequenced to date include the proteins from spinach (Matsubara et al., 1967), alfalfa (Keresztes-Nagy et al., 1969), *Leucaena glauca* (Benson and Yasunobu, 1969), *Colocasia esculenta* (taro) (Rao and Matsubara, 1970), and *Scenedesmus* (Sugeno and Matsubara, 1969). These ferredoxins have all been isolated from eucaryotes. Ferredoxins from plant and algae contain 2 g-atoms of iron and sulfide per mol of protein (Lovenberg, 1973). During the sequence determination of the *S. maxima* ferredoxin, a useful modification of the automated Edman degradation was observed. This technique is described in the present report. A brief report which describes the sequence of the *S. maxima* ferredoxin has been published (Tanaka et al., 1975).

Experimental Procedures

The *S. maxima* ferredoxin was isolated in a purified form as described in a previous report (Hall et al., 1972). The iron and sulfide were removed and the *S*-β-carboxymethylcysteine derivative was prepared as reported previously (Yasunobu and Tanaka, 1973). All of the sequencer reagents were of sequanal quality and were purchased from Pierce Chemical Co. Trypsin was purchased from the Worthington Biochemical Corp. as a three times crystalized preparation and was treated with L-1-tosylamido-2-phenylethyl chloromethyl ketone.

Methods

Amino Acid Composition. The amino acid compositions of the protein and peptides were determined on 5.7 *N* HCl hydrolysates in a Beckman-Spinco 120C automatic amino acid analyzer as described by Spackman et al. (1958). The instrument was equipped with high sensitivity cuvettes and a 4–5 mV full scale range card.

End Group Analyses. NH_2-terminal end group analyses of the protein and peptides were determined by the manual Edman degradation procedure (Edman, 1970). The carboxyl terminal amino acid analyses were determined by hydrazinolysis (Bradbury, 1958) and by the use of carboxypeptidase A and B (Ambler, 1967).

Sequence Determinations. The manual Edman degradation (Edman, 1970) or the automated Edman degradation procedures (Edman and Begg, 1967) were used to sequence the peptides and proteins (100–350 nmol). All of the automated runs were carried out in the Beckman-Spinco Model 890 protein/peptide sequencer and utilized the protein double cleavage program. The amino acid phenylthiohydantoins were identified by gas chromatography in a Beckman GC-45 gas chromatograph as described by Pisano and Bronzert (1969), or by thin-layer chromatography as described by Edman and Begg (1967), or by amino acid analysis of 6 *N* HCl hydrolysates of the amino acid phenylthiohydantoin (Van Orden and Carpenter, 1964).

Tryptic Peptides. About 2.5 μmol of the Cys(Cm)-ferredoxin was hydrolyzed with 5% trypsin for 22 hr at 28°. The sample was then immediately dried and then dissolved in 1.0 ml of 0.1 *M* NH_4OH and the sample was applied to a Sephadex G-50 column (1.9 × 55 cm). 0.1 *N* ammonium hydroxide was used as the eluent. The flow rate was 60 ml/hr and fractions of 3.85 ml were collected for measuring the absorbance. The peak fractions were collected, pooled, and concentrated. Peptides were further purified by paper chromatography using several different solvent systems.

† From the Department of Biochemistry-Biophysics, University of Hawaii School of Medicine, Honolulu, Hawaii 96822, and the Botany Department, University of London King's College, London, England. *Received June 18, 1975.* Research was supported in part by Grants GM 16784 and GM 16228 from the National Institutes of Health, Grant GB 43448 from the National Science Foundation, Hawaii Heart Association, and the U.K. Science Research Council.

Table I: Amino Acid Composition of the *Spirulina maxima* Ferredoxin.

Amino Acid	From Acid Hydrolysates[a]	From the Sequence
Lysine	2.00 (2)	2
Histidine	0.94 (1)	1
Arginine	1.03 (1)	1
Carboxymethylcysteine	5.83 (6)	6
Aspartic acid	12.98 (13)	13[b]
Threonine	9.70 (10)	10
Serine	7.56 (8)	8
Glutamic acid	13.07 (13)	13[c]
Proline	1.94 (2)	2
Glycine	6.97 (7)	7
Alanine	10.07 (10)	10
Valine	2.98 (3)	3
Isoleucine	7.80 (8)	8
Leucine	7.03 (7)	7
Tyrosine	5.83 (6)	6
Phenylalanine	1.05 (1)	1
Total residues	98	98

[a] Acid hydrolyses were carried out on carboxymethylferredoxin for 24, 48, and 72 hr at 110° with 6 *N* HCl. The amino acid residues were calculated on the basis of a lysine content of 2.00 mol/mol of protein. The values of threonine and serine were obtained by extrapolating to zero time. The values of valine, isoleucine, and leucine were the maximum values (72 hr). Numbers in parentheses indicate values rounded off to the nearest whole number. [b] Sum of 12 aspartic acids and one asparagine. [c] Sum of eight glutamic acids and five glutamines.

Results

Purity and Amino Acid Composition of the S. maxima Ferredoxin. Manual Edman degradation and carboxypeptidase A digestion of the protein disclosed that alanine and tyrosine were the NH_2- and COOH-terminal amino acids of the Cys(Cm)-ferredoxin. The yields were quantitative and no other end groups were detected. The amino acid composition of the *S. maxima* Cys(Cm)-ferredoxin was determined on the pure sample and the results of these analyses are summarized in Table I. No tryptophan was detected when the procedure of Liu (1972) was used and no methionine sulfone was detected in performic acid treated ferredoxin (Hirs, 1967). The ferredoxin contained 98 amino acids in agreement with the total amino acid content determined by sequence determination of the protein.

NH_2-Terminal Sequence of the Cys(Cm)-ferredoxin. As a first step in the sequence determination of the protein, about 225 nmol of the Cys(Cm)-ferredoxin were analyzed in the protein sequencer. The results of these analyses are summarized in Figure 1. Excellent results were obtained and it was possible to clearly determine the sequence of the first 50 residues from the NH_2-terminal end of the protein. The average repetitive yield of Pth-alanine was 95% and the recovery of Pth-alanine at the fiftieth step was about 10%.

Isolation and Purification of the Tryptic Peptides. In order to complete the sequence of the ferredoxin, the Cys(Cm)-ferredoxin was hydrolyzed with trypsin. The hydrolysate was then separated into different molecular weight fractions by passing the sample on a Sephadex G-50 column. Three peaks were observed in the chromatogram and the samples under each peak were pooled. The first peak contained a mixture of peptides T-2 and T-4. The second peak contained peptide T-3 plus impurities and peak 3 contained peptide T-1 in an impure state.

The mixture of peptides T-2 and T-4 were separated by paper chromatography in 1-butanol–pyridine–acetic acid–water (60:40:12:48, v/v) to yield partially purified T-2 and T-4. The peptides were individually further purified in pyridine–isoamyl alcohol–0.1 *N* ammonium hydroxide (60:30:50, v/v). Peptides T-1 and T-3 were sufficiently pure after chromatography in 1-butanol–pyridine–acetic acid–water.

Table II: Amino Acid Composition[a] and Properties of Tryptic Peptides of Carboxymethylferredoxin.

Amino Acid	T-1	T-2	T-3	T-4	Total Residues
Lysine	0.95 (1)		0.96 (1)		2
Histidine				0.86 (1)	1
Arginine		0.85 (1)			1
Carboxymethylcysteine		1.97 (2)	1.84 (2)	1.88 (2)	6
Aspartic acid		7.04 (7)		5.99 (6)	13
Threonine	0.88 (1)	2.84 (3)	0.92 (1)	4.92 (5)	10
Serine		1.77 (2)	0.86 (1)	4.75 (5)	8
Glutamic acid		5.06 (5)		8.10 (8)	13
Proline		0.97 (1)		0.90 (1)	2
Glycine		1.99 (2)	1.94 (2)	2.93 (3)	7
Alanine	1.00 (1)	4.00 (4)	3.00 (3)	2.00 (2)	10
Valine		1.05 (1)		1.98 (2)	3
Isoleucine		4.06 (4)		4.09 (4)	8
Leucine		4.04 (4)		3.10 (3)	7
Tyrosine	0.94 (1)	1.98 (2)		2.87 (3)	6
Phenylalanine				0.99 (1)	1
Total residues	4	38	10	46	98
Recovery (%)	74	71	86	71	
R_f in BPAW[b]	0.45	0.00	0.20	0.41–0.69	
R_f in PIN[b]	0.63	0.23	0.41	0.33	
Color reaction with ninhydrin	Purple	Grayish violet	Violet	Purple	
Pauly reaction	Orange brown	Brown		Reddish brown	
Purification method[b]	BPAW	BPAW PIN	BPAW	BPAW PIN	

[a] Results from 6 *N* HCl hydrolysates (24 and 48 hr). The numbers in parentheses refer to the assumed stoichiometric number of residues per molecule of pure peptide. In the table, the amounts of amino acids present in less than 10% are not reported. [b] Abbreviations used are: BPAW, paper chromatography in the solvent system, 1-butanol–pyridine–acetic acid–water (60:40:12:48, v/v); and PIN, paper chromatography with pyridine–isoamyl alcohol–0.1 *N* ammonium hydroxide (60:30:50, v/v).

The amino acid compositions of the tryptic peptides are summarized in Table II and some of the properties of the peptides are presented. The four tryptic peptides contained 98 amino acid residues in excellent agreement with the amino acid composition of the protein.

Sequence Determinations of Tryptic Peptides. Residues 43–52. Peptide T-3: Ala-Gly-Ala-Cys(Cm)-Ser-Thr-Cys-(Cm)-Ala-Gly-Lys. Nine steps of the manual Edman degradation and direct amino acid analysis of the residue after the ninth step of the Edman reaction established the sequence of peptide T-3 as shown in Figure 1. The yield of Pth-glycine at the ninth step was 20%.

Residues 53–98. Peptide T-4: Ile-Thr-Ser-Gly-Ser-Ile-Asp-Gln-Ser-Asp-Gln-Ser-Phe-Leu-Asp-Asp-Asp-Gln-Ile-Glu-Ala-Gly-Tyr-Val-Leu-Thr-Cys(Cm)-Val-Ala-Tyr-Pro-Thr-Ser-Asp-Cys(Cm)-Thr-Ile-Gln-Thr-His-Gln-Glu-Glu-Gly-Leu-Tyr. The peptide was analyzed in the protein sequencer and the protein program was used. After about step 29, the yield of the amino acid phenylthiohydantoin dropped precipitously. Therefore, in the second determination, after step 28 the residual peptide was removed from the cup with 20% pyridine and transferred to a test tube and

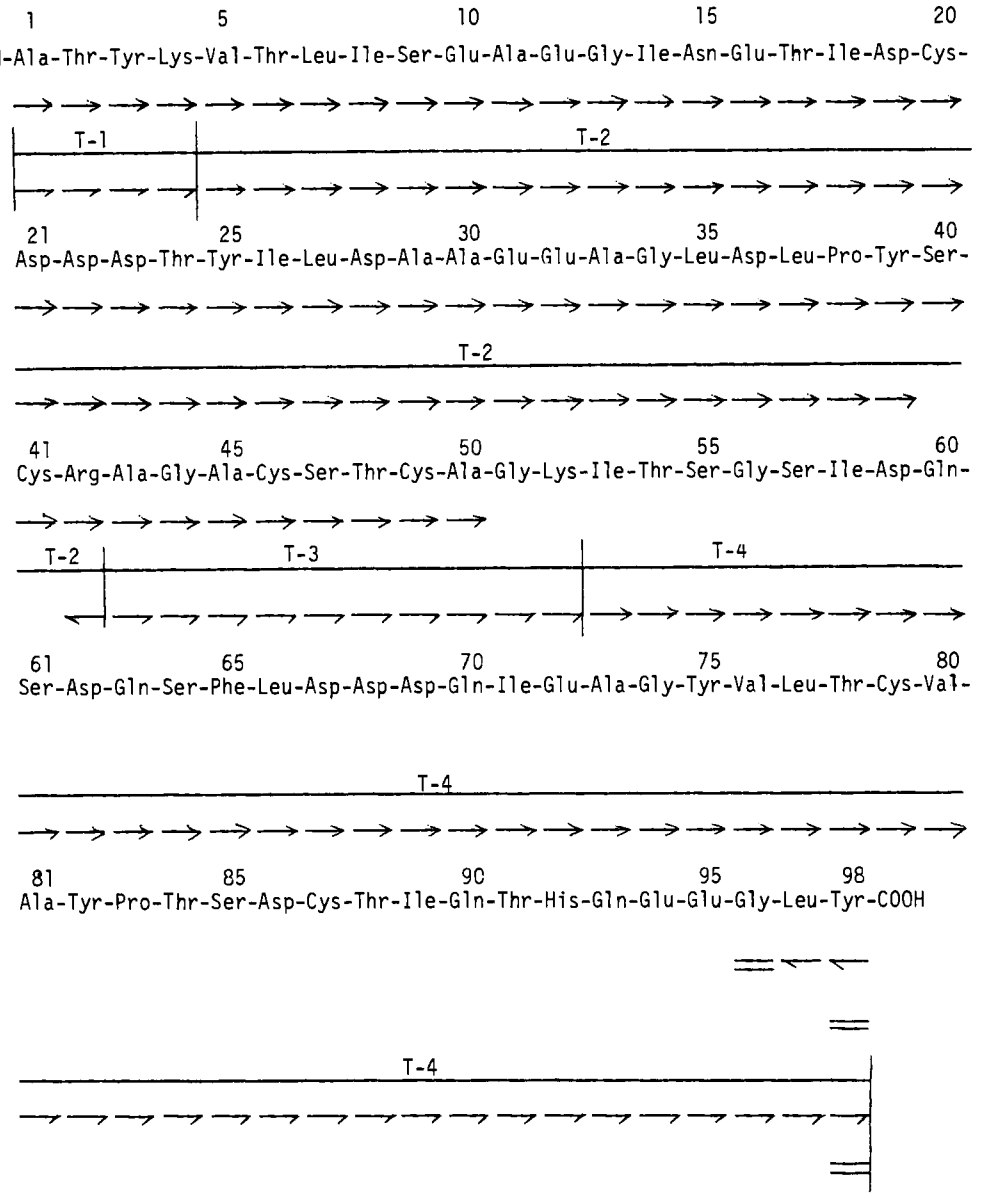

FIGURE 1: Reconstruction of the complete amino acid sequence of *Spirulina maxima* ferredoxin and sequence data of peptide fragments. In the figure, the symbols →, ⇀, ↼, = represent sequences determined by use of the Beckman sequencer, direct manual Edman degradation, carboxypeptidase A or B, and hydrazinolysis experiments, respectively.

dried. Thus, from steps 29 to 45, manual Edman degradation was used to complete the sequence of the peptide as shown in Figure 1. Although the yield of tyrosine was low, 6% after the 45th step of Edman degradation, however, hydrazinolysis of the peptide also confirmed that tyrosine (81% yields) was the C-terminal amino acid.

Complete Sequence. The protein sequencer run established the sequence of the first 50 residues from the NH_2-terminal end of the protein and indicated that peptides T-1 and T-2 and a part of T-3 were from the NH_2-terminal region of the protein (Figure 1). In order to establish the identity of residues 51–52, peptide T-3 was sequenced. Peptide T-4 was the COOH-terminal peptide based on the hydrazinolysis experiments which showed that tyrosine was the COOH-terminal amino acid.

Discussion

The amino acid sequence of five ferredoxins of the plant–algal type have been sequenced to date and include the ferredoxins from spinach (Matsubara et al., 1967), alfalfa (Keresztes-Nagy et al., 1969), *Leucaena glauca* (Benson and Yasunobu, 1969), taro (Rao and Matsubara, 1970), and *Scenedesmus* (Sugeno and Matsubara, 1969). Thus far, the protein sequencer has not been used to complete the sequence of a ferredoxin of this type. The sequence and chemical properties of the *S. maxima* ferredoxin were especially suited for the sequence analysis in the protein sequencer and at least 50 residues (out of a total of 98 residues) from the NH_2-terminal end could be clearly established in the automated instrument. Wada et al. (1974) have reported the partial sequence of the *Aphanothece sacrum* ferredoxin and have determined about 37 residues from the NH_2-terminal residue of this algal ferredoxin in the protein sequencer which is 13 residues less than was determined for the *S. maxima* ferredoxin. In order to complete the sequence of the *S. maxima* ferredoxin it was necessary to resort to trypsin digestion of the protein. The tryptic peptides yielded sufficient sequence data so that the sequence of the protein could be completed. A technique which has not been previously used was found to be useful for determining the

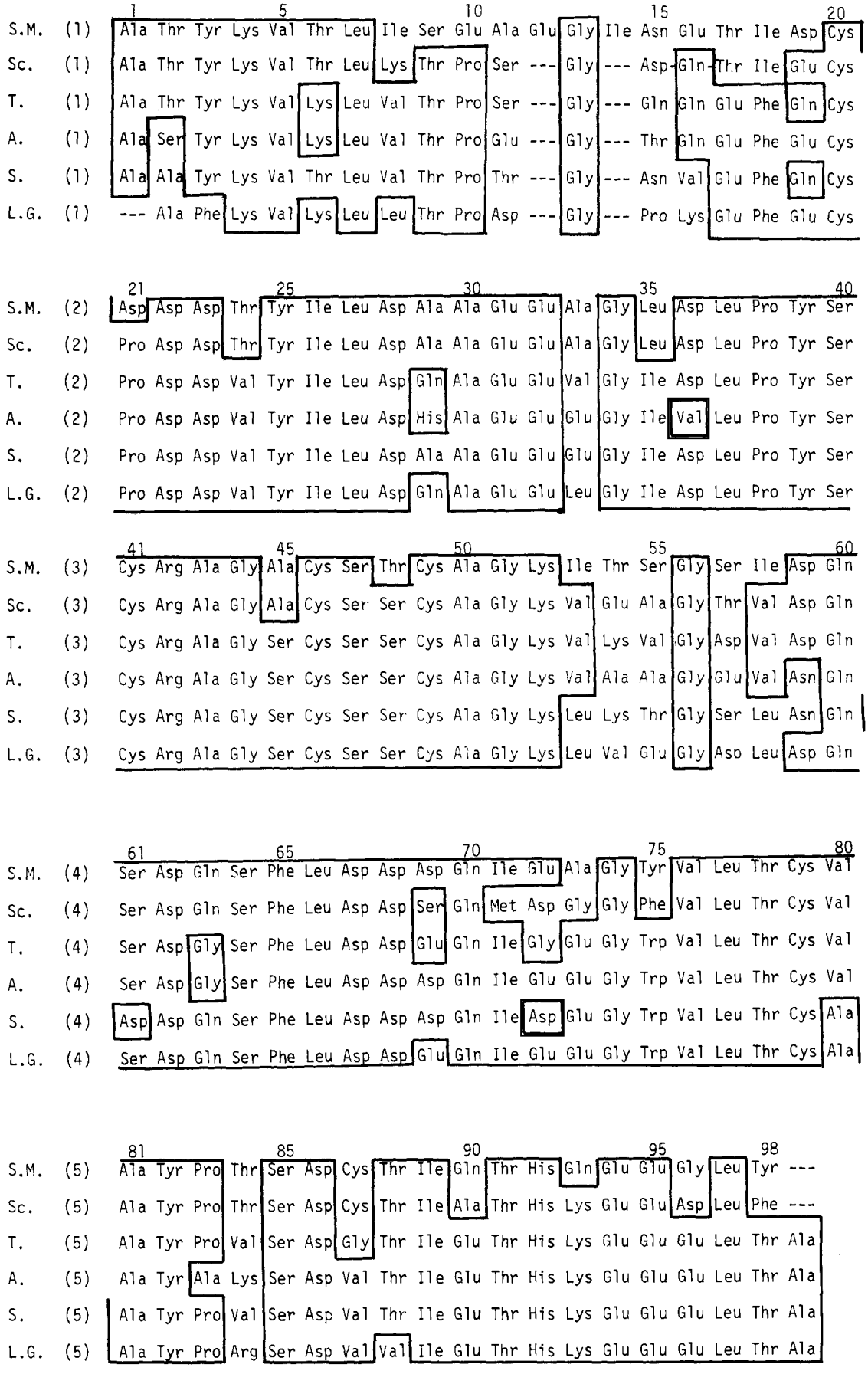

FIGURE 2: Comparison of the amino acid sequences of various plant and algal ferredoxins. Abbreviations used are: S.M., *Spirulina maxima;* Sc., *Scenedesmus;* T., taro; A., alfalfa; S., spinach; and L.G., *Leucaena glauca.* Identical residues have been blocked off. An insertion of a glutamic acid and an isoleucine residue at positions 12 and 14 in the *S. maxima* ferredoxin or a deletion of these residues in the eucaryote ferredoxins was assumed.

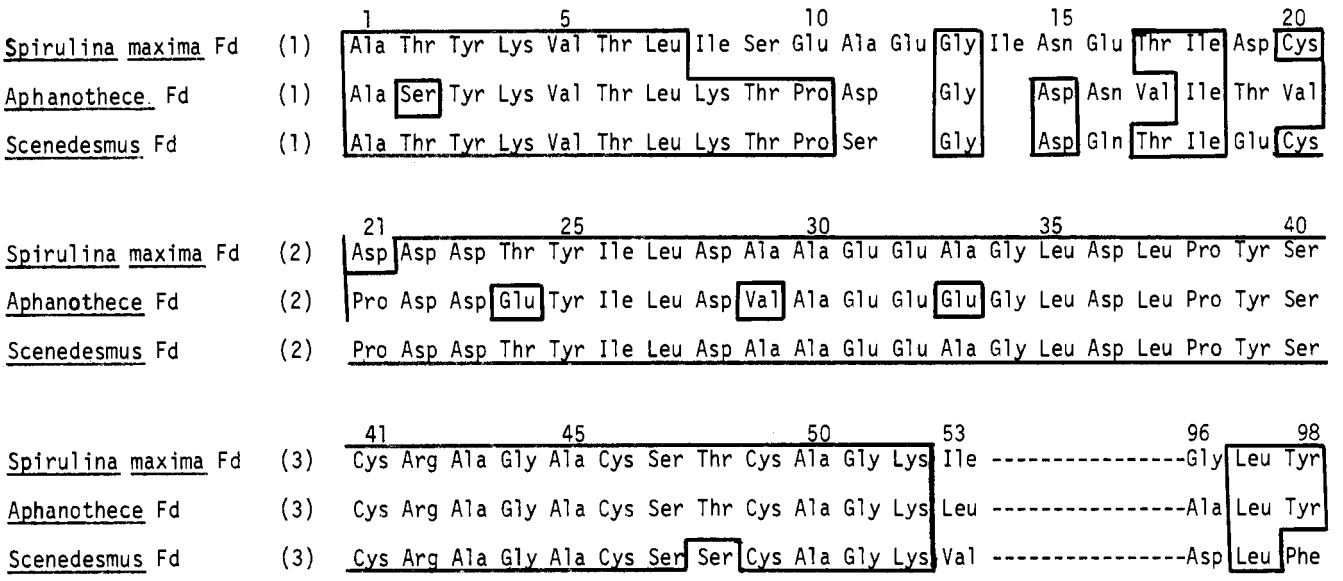

FIGURE 3: NH_2- and COOH-terminal amino acid sequences of the various algal ferredoxins which have been sequenced.

sequence of peptide T-4 which contained 46 amino acid residues. The peptide was analyzed in the protein sequencer and it was noted that after step 29, the yield of the Pth-tyrosine dropped greatly. In the second run in the protein sequencer, the residual peptide after step 28 was removed from the cup and the remainder of the steps of the Edman degradation were performed manually. By this modification, it was possible to completely sequence the peptide (18 additional steps of Edman degradation). The use of sequanal quality reagents and benzene or ethyl acetate for the extraction steps of Edman degradation of peptides has made it possible in our experience to go up about 20 steps in favorable cases. Although not studied here, it would appear that in the case of peptide T-4, the washing out of the residual peptide by organic solvents used to extract residual reagents and the thiazolinone derivatives of the amino acids resulted in a washing out of the residual peptide. The technique of removing residual peptide during the automated Edman degradation when extraction losses occur is an obvious technique but has not been used previously. Current modifications have included the use of 4-sulfophenyl isothiocyanate to make the peptide more hydrophilic (Braunitzer et al., 1970, 1972) or have involved the coupling of COOH-terminal lysine peptides to the glass of the sequencer reaction vessel (Laursen et al., 1972; Wachter et al., 1973). The only drawback with the proposed procedure is that the sequence must be determined manually but in the case of peptide T-4, the complete sequence of the peptide was determined by the modified procedure. Therefore, it was not necessary to fragment the peptide into smaller fragments which also meant a saving of time. Although the sequence was determined on about 40 mg of protein, the sequence could easily be determined on about a half of this amount of protein.

The sequences of all the plant–algal type of ferredoxins sequenced to date are shown in Figure 2. All of the ferredoxins of this type contain 96–98 amino acids and the sequence similarities vary from about 60 to 81 residues (Table III) which indicates great sequence homologies of these proteins.

Stanier et al. (1971) have emphasized that unicellular blue-green algae which belong to the order *Chroococcales* should be classified as bacteria rather than algae. Although the sequence of a ferredoxin from the unicellular blue-green algae has not been determined, the amino acid composition of the *Microcystis* ferredoxin has been determined (Rao et al., 1972) and it contains about 97 amino acids like the plant ferredoxins. The *S. maxima* is a filamentous blue-green algae and, as shown in Figure 3, the ferredoxin from this algae resembles the plant–algal type. Thus, the photosynthetic machinery of the blue-green algae appears to have evolved to a greater extent than the photosynthetic apparatus of the photosynthetic bacteria. Mutation rates of the ferredoxin containing organisms need to be known and the sequence studies may provide this type of data. However, until these data are available, the ferredoxin sequences cannot be used to create phylogenetic trees. At the present time, the tree obtained from sequence data does not agree with botanists' tree (Fitch and Yasunobu, 1975).

Table III: Intraspecies Similarities of the Amino Acid Sequences of the Algal and Chloroplast Ferredoxins.[b]

Species[a]	S.M.	Sc.	T.	A.	S.	L.G.
S.M.		73	63	61	65	60
Sc.	0.30		71	68	67	64
T.	0.48	0.38		81	79	78
A.	0.47	0.42	0.21		78	75
S.	0.43	0.41	0.26	0.28		78
L.G.	0.52	0.46	0.25	0.25	0.30	

[a] Abbreviations used are: S.M., *Spirulina maxima;* Sc., *Scenedesmus;* T., taro; A., alfalfa; S., spinach; and L.G., *Leucaena glauca.* [b] Numbers above the diagonal refer to the number of residues which are identical in the intraspecies comparisons. Numbers below the diagonal refer to the minimal base difference per codon values.

References

Ambler, R. B. (1967), *Methods Enzymol. 11,* 436.
Benson, A. M., and Yasunobu, K. T. (1969), *J. Biol. Chem. 244,* 955.
Bradbury, J. H. (1958), *Biochem. J. 68,* 475.
Braunitzer, G., Chen, R., Schrank, B., and Stangl, A. (1972), *Hoppe-Seyler's Z. Physiol. Chem. 353,* 832.
Braunitzer, G., Schrank, B., and Ruhfus, A. (1970), *Hoppe-Seyler's Z. Physiol. Chem. 351,* 1589.
Edman, P. (1970), in Protein Sequence Determination,

Needleman, S. B., Ed., West Berlin, Springer-Verlag, p 211.
Edman, P., and Begg, G. (1967), *Eur. J. Biochem. 1,* 80.
Fitch, W. M., and Yasunobu, K. T. (1975), *J. Mol. Evol. 5,* 1-24.
Hall, D. O., Rao, K. K., and Cammack, R. (1972), *Biochem. Biophys. Res. Commun. 47,* 798.
Hirs, C. H. W. (1967), *Methods Enzymol. 11,* 197.
Keresztes-Nagy, S., Perini, F., and Margoliash, E. (1969), *J. Biol. Chem. 244,* 981.
Laursen, R. A., Horn, M. J., and Bonner, A. G. (1972), *FEBS Lett. 21,* 67.
Liu, T.-Y. (1972), *Methods Enzymol. 25B,* 44.
Lovenberg, W., Ed. (1973), in Iron-Sulfur Proteins, Vols. I and II, New York, N.Y., Academic Press.
Matsubara, H., Sasaki, R. M., and Chain, R. K. (1967), *Proc. Natl. Acad. Sci. U.S.A. 57,* 439.
Pisano, J. J., and Bronzert, T. J. (1969), *J. Biol. Chem. 244,* 5597.
Rao, K. K., and Matsubara, H. (1970), *Biochem. Biophys. Res. Commun. 38,* 500.
Rao, K. K., Smith, R. V., Cammack, R., Evans, M. C. W., Hall, D. O., and Johnson, C. E. (1972), *Biochem. J. 129,* 1159.
Spackman, D. H., Moore, S., and Stein, W. H. (1958), *Anal. Chem. 30,* 1190.
Stanier, R. Y., Kunisawa, R., Mandel, M., and Cohen-Bazire, G. (1971), *Bacteriol. Rev. 35,* 171.
Sugeno, K., and Matsubara, H. (1969), *J. Biol. Chem. 244,* 2979.
Tanaka, M., Haniu, M., Zeitlin, S., Yasunobu, K. T., Evans, M. C. W., Rao, K. K., and Hall, D. O. (1975), *Biochem. Biophys. Res. Commun. 64,* 399.
Van Orden, H. O., and Carpenter, F. H. (1964), *Biochem. Biophys. Res. Commun. 14,* 399.
Wachter, E., Machleidt, W., Rofner, H., and Otto, J. (1973), *FEBS Lett. 35,* 97.
Wada, K., Kagamiyama, H., Shin, M., and Matsubara, H. (1974), *J. Biochem. (Tokyo) 76,* 1217.
Yasunobu, K. T., and Tanaka, M. (1973), in Iron-Sulfur Proteins, Vol. II, Lovenberg, W., Ed., New York, N.Y., Academic Press, p 27.