Amino Acid Sequence of Spinach Ferredoxin:NADP$^+$ Oxidoreductase†

P. Andrew Karplus, Kenneth A. Walsh, and Jon R. Herriott*

ABSTRACT: The amino acid sequence of spinach ferredoxin:NADP$^+$ oxidoreductase was determined by using overlapping sets of peptides derived by cleavage at arginyl or methionyl residues. The protein from different preparations varied in its length at the amino terminus. In the longest form the amino terminus is blocked with a pyroglutamyl residue, as determined by NMR. A single disulfide bond was placed between cysteine residues 132 and 137. The 314-residue sequence corresponds to a molecular weight of 35 317. The carboxyl-terminal half of the sequence has been fit to the electron density map of the NADP binding domain, revealing that this portion of the chain forms a typical nucleotide binding fold.

Ferredoxin:NADP$^+$ oxidoreductase (EC 1.18.1.2) has been isolated from a wide variety of plants (Shin et al., 1963; Sluiters-Scholten et al., 1977; Firenzuoli et al., 1968; Riov & Brown, 1976) as well as cyanobacteria (Hutber et al., 1978) and blue-green algae (Susor & Krogmann, 1966; Manzano et al., 1978; Wada et al., 1983). Since the work of Shin et al. (1963), it has been generally accepted that the biologically relevant activity of ferredoxin reductase (FNR)[1] is to catalyze the transfer of reducing equivalents from the one-electron donor, reduced ferredoxin, to the two-electron acceptor, NADP$^+$. This reaction is the last step in the linear photosynthetic pathway. Some workers have suggested that FNR is involved in the pathways of cyclic photosynthesis also (Forti & Rosa, 1971; Shahak et al., 1981), although this is not universally accepted (Curtis et al., 1973; Böhme, 1977). In any case, FNR may be a key site for the regulation of the relative amounts of cyclic and noncyclic photosynthesis carried out and thus of the relative amounts of NADPH and ATP produced from light energy absorbed.

FNR is coded for by nuclear DNA and synthesized as a higher molecular weight precursor that is processed and transported into the chloroplast posttranslationally (Grossman et al., 1982). The mechanisms involved in the regulation of FNR synthesis and in the transport across the chloroplast membranes are not well understood. The mature form of FNR is associated with the intergrana regions of the thylakoid membrane as an extrinsic membrane protein (Berzborn, 1969; Böhme, 1978). Recent reports indicate that this membrane association has important effects on the structure and activity of the enzyme (Carrillo et al., 1981; Wagner et al., 1982).

Structural studies on FNR have been somewhat complicated by the existence of multiple forms of monomeric FNR as was first reported by Keirns & Wang (1972). Their purified FNR, which gave a single band on SDS–PAGE, could be separated into three active bands by isoelectric focusing. Since then various laboratories have reported up to eight isoelectric species of FNR (Gozzer et al., 1977; Ellefson & Krogmann, 1979; Hasumi et al., 1983). The nature of these species is not understood, but it is known that many of them are interconvertible. In addition to the multiple isoelectric forms of FNR, some preparations yield two bands on SDS–PAGE of approximate molecular weights 35 000–37 000 and 32 000–34 000 (Gozzer et al., 1977; Hasumi et al., 1983).

Our laboratory has been investigating the three-dimensional structure of FNR by X-ray crystallography. Sheriff & Herriott (1981) reported the crystal structure of FNR at 3.7-Å resolution. This study revealed that FNR is a kidney-shaped protein of rough dimension 30 × 50 × 50 Å. One of the two lobes of the enzyme is involved in the binding of NADP$^+$ as shown by soaking NADPH into the crystal; the other is presumed to bind FAD, although the bound FAD had not yet been located with certainty. In order to carry this work further, we deemed it necessary to know the amino acid sequence of the protein so that it could be used to interpret the electron density map and eventually to allow refinement at higher resolution. We report here the complete amino acid sequence of FNR.

Materials and Methods

Preparation of FNR. FNR was prepared from local spinach as reported previously (Sheriff et al., 1980). In some preparations the order of the acetone and ammonium sulfate precipitations was switched, and if at the end of the preparation the ratio of A_{456}/A_{276} was not greater than 0.111, an additional Sephacryl S-200 separation (2.0 × 85 cm column) was performed in 0.2 M K_2HPO_4, pH 7.0. The specific activity of the purified protein, as determined by the transhydrogenase assay (Keister et al., 1960), ranged between 3000 and 4000 units/mg. Some preparations yielded protein that ran as a doublet on SDS–PAGE. As discussed above, the protein showed multiple forms on isoelectric focusing (p*I* of 4.8–6.0), but since these isozymes are interconvertible (Ellefson & Krogmann, 1979), they were assumed to lack covalent differences and a mixture of the forms was used for sequencing. The purified FNR was stored for up to 6 months at −20 °C in 0.2 M K_2HPO_4, pH 7, without the addition of protease inhibitors.

For reduction and carboxymethylation, FNR was transferred into 0.1 M NH_4HCO_3 via gel filtration over Sephadex G-25 and lyophilized. The dry protein was taken up in 6 M Gdn·HCl (Heico), 1 M Tris, and 0.01 M EDTA, pH 8.6, and treated with a 100-fold molar excess of dithiothreitol at 50 °C for 2 h, and then with the sodium salt of iodoacetic acid (or its ^{14}C- or ^3H-radiolabeled counterpart from New England Nuclear) at 1.1 times the theoretical amount of dithiothreitol.

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received June 11, 1984.* This work was supported by grants to J.R.H. (National Science Foundation Grant 81-09700) and to K.A.W. (National Institutes of Health Grant GM 15731). P.A.K. was supported in part by a National Institutes of Health predoctoral traineeship (GM 07270).

[1] Abbreviations: FNR, spinach ferredoxin:NADP$^+$ oxidoreductase; CM, *S*-carboxymethyl; HPLC, high-performance liquid chromatography; TFA, trifluoroacetic acid; SDS–PAGE, polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate; Gdn·HCl, guanidine hydrochloride; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; TPCK, N^{α}-tosylphenylalanine chloromethyl ketone.

0006-2960/84/0423-6576$01.50/0

Dithiothreitol was added once more to stop the reaction, and the mixture was desalted into 0.1 M NH_4HCO_3 by use of Sephadex G-25 and then lyophilized.

For determination of the single disulfide reported to be in FNR (Zenetti & Forti, 1969), the reduction and alkylation procedure was slightly modified. First, the cysteine residues were alkylated by adding a 10-fold excess of unlabeled iodoacetic acid to the denatured protein before reduction. Then a 50-fold excess (over the iodoacetate) of dithiothreitol was added to reduce the disulfide. In order to ensure maximal incorporation of radiolabel into the protein, the excess dithiothreitol was removed by gel filtration over Sephadex G-25 before [^{14}C]iodoacetic acid was added. This procedure introduced radiolabeled carboxymethyl groups into the specific cysteine residues originally involved in the disulfide of the native protein.

Selective Cleavage. *S*-Carboxymethyl-FNR (CM-FNR, 600 nmol) was citraconylated in 4 mL of 6 M Gdn·HCl and 40 mM Na_2HPO_4, pH 8.8, by the addition of 100 μL of citraconic anhydride (Eastman Kodak) over 30 min, maintaining the pH with NaOH. After removal of the guanidine by gel filtration in 0.1 M NH_4HCO_3, pH 8.8, the protein was treated with 200 μg of TPCK–trypsin (Worthington) for 2 h at 37 °C and directly fractionated by gel filtration in the same buffer. Decitraconylation was carried out by incubation of the citraconyl peptides in 30% HCOOH for 3 h at 37 °C.

For cleavage at methionine, CM-FNR was incubated at room temperature with 1% by weight cyanogen bromide in 70% HCOOH for 15 h. The reaction mixture was lyophilized before gel filtration to remove the excess reagent.

Digests with trypsin or subtilisin (Nagarse) were done in 0.1 M NH_4HCO_3, pH 8, at 37 °C for 2 h with 2% mole fraction of protease. Digests with *Staphylococcus aureus* V8 protease (Miles Laboratories) were done in 0.1 M K_2HPO_4, pH 7.8, at 37 °C for 15 h with 2% mole fraction of protease unless otherwise specified.

Specific chemical cleavage at asparaginyl–glycine bonds was attempted by incubation of the whole protein or peptide R3 for 8 h at 45 °C in 6 M Gdn·HCl and 2 M hydroxylamine hydrochloride adjusted to pH 9.0 with LiOH (Bornstein & Balian, 1977).

HPLC Chromatography. Reversed-phase HPLC utilized a Varian 5000 liquid chromatograph and μBondapak C_{18} columns (Waters). Unless otherwise specified, chromatography was done at 2 mL/min at pH 2 by using 0.1% trifluoroacetic acid (TFA) as the aqueous phase and by eluting with an increasing concentration of acetonitrile (Mahoney & Hermodson, 1980). A flat base line at 206 nm was maintained by adding an empirically determined amount of TFA to the acetonitrile. In some cases chromatography took place in 5 mM NH_4HCO_3 adjusted to pH 6.5 with HCOOH and by eluting with a gradient of acetonitrile.

High-pressure ion-exchange chromatography on a sulfonylpropyl resin (TSK SP-5PW from Toyo Soda) was used for the isolation of a blocked amino-terminal peptide from a tryptic digest of citraconylated CM-FNR. The column was equilibrated and loaded in 25 mM NaH_2PO_4, pH 2.1, and eluted with a linear gradient to 25 mM Na_2HPO_4 and 2.0 M NaCl, pH 6.0.

Analytical Methods. Amino acid compositions were determined on a Dionex amino acid analyzer (Model D-500), after hydrolysis in 5.7 N HCl at 108 °C for 16–24 h unless otherwise noted. Automated sequence analyses were performed on a Beckman sequencer (Model 890C) in the presence of Polybrene (Pierce) by using the method of Edman & Begg (1967) with the Quadrol program of Brauer et al. (1975). Peptides E1, E2, and M4 were sequenced on a gas-phase instrument (Applied Biosystems, Inc.) (Hunkapiller et al., 1983). Phenylthiohydantoin derivatives of amino acids were identified by reversed-phase HPLC using two complementary systems (Ericsson et al., 1977; Bridgen et al., 1976).

NMR spectra were obtained on a Bruker WM-500 instrument. Salt-free samples were dissolved in 99.996% D_2O (Stohler) for analysis. A 2 mM solution of pyroglutamylalanine (Bachem) was scanned 256 times and a 0.14 mM solution of R1a-N1 was scanned 1024 times, both with a spectral width of 5200 Hz. The data were then Fourier transformed without resolution enhancement.

Results

The sequence data that comprise the simplest proof of the structure are summarized in Figure 1. Most of the sequence (270 out of 314 residues) was obtained by direct sequence analysis of 10 fragments resulting from cleavage at arginyl residues by trypsin. The majority of these fragments were ordered by sequence analyses of peptides generated by cleavage at methionyl residues. The few remaining overlaps were established with peptides derived from digests with *S. aureus* V8 protease or trypsin.

Sequence Analysis of Whole Protein. Amino-terminal analyses of various preparations of FNR, both native and carboxymethylated, gave one of three results: (1) low yields (0–20%) of a sequence starting at Ile-2 (FNRα in Figure 1), (2) amino-terminal serine, aspartate, valine, and alanine residues in approximate yields of 30, 10, 15, and 10%, respectively, and (3) a sequence starting at residue 16 (VEKHSK...). There was some correlation of the sequence results with the pattern obtained upon SDS–PAGE. In the cases analyzed, those preparations giving a single 35000-dalton band yielded sequence results of the first kind. Results of the second kind (a mixture) were obtained twice for preparations that had at least 50% of the 33000-dalton band. In these cases the sequence that followed could not be interpreted with confidence, but was consistent with a mixture of sequences beginning at positions 4, 5, 6, and 8 of the whole protein. Only a single preparation, corresponding to that used for the original cleavage at arginyl residues, gave the third kind of sequence results, indicating the lack of 15 residues at the amino terminus. Unfortunately, this preparation was not analyzed by SDS–PAGE.

Arginine Fragments. Two tryptic digestions were prepared of CM-FNR that had been citraconylated in order to direct cleavage to arginyl residues. Because the first digest was done on protein that was missing the first 15 residues, a second digest was prepared to isolate the blocked amino-terminal peptide. In the first digest, primary separation of the arginine fragments was carried out by gel filtration chromatography, and each of the eight resulting pools was rechromatographed on reversed-phase HPLC at pH 2 (Figure 2). This yielded 10 unique and nonoverlapping peptides encompassing all but the first 15 residues of the protein and beginning with the peptide R1′ (Table I). The yield of each of these peptides was greater than 40%. The larger peptides R5, R6, R9, and R10 were not soluble in acid until after decitraconylation. After reversed-phase chromatography only R2, R4, and R6 were not completely pure, as noted in Table I.

Each of the peptides R1′, R2, R4, R6, R7, R8, and R10 was sequenced through its carboxyl terminus. R5 (50 nmol) was subdigested with trypsin, and the sequence of R5-T1, isolated by reversed-phase HPLC at pH 2, completed the sequence analysis of R5. Two attempts were made to sequence

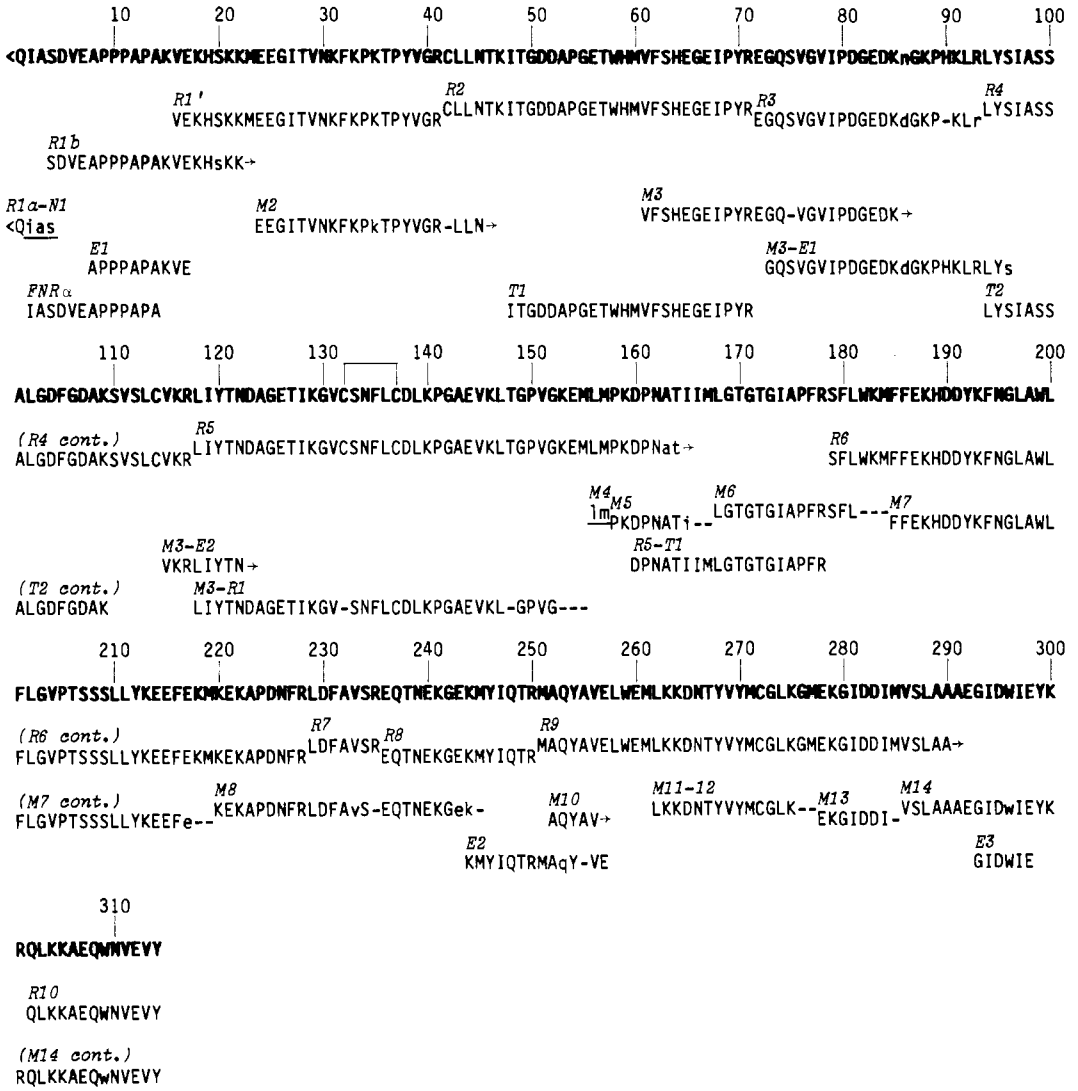

FIGURE 1: Summary of proof of the amino acid sequence of FNR from spinach. The final proven sequence of FNR appears in bold type in single-letter code (see Table I). Sequences of specific peptides, as determined by Edman degradation, are given in capital letters below the summary sequence. A lower-case letter indicates that the identification of the residue in a given Edman degradation is only tentative, a hyphen indicates an unidentified residue, an underline indicates placement by composition only, and an arrow indicates that the carboxyl-terminal residues of a particular peptide were not sequenced. Each peptide name appears in italics just above the first residues of the peptide. The prefixes R and M refer to peptides derived by cleavage at arginyl and methionyl bonds, respectively, and the prefixes E and T refer to peptides derived from digests with *S. aureus* V8 protease and trypsin, respectively. FNRα indicates the sequence obtained from undigested protein. The symbol <Q denotes an amino-terminal pyroglutamyl group identified by NMR of peptide R1a-N1. The asparagine indicated by a lower-case n at residue 86 is only tentatively identified (see text). The disulfide bond proven to be between Cys-132 and Cys-137 is indicated.

R3, and during both sequenator runs there was a sharp drop in yield (to <5%) at turn 15 that showed a small amount of aspartic acid. Thereafter, the sequence continued at good stepwise yield, but at a very low level, through the carboxyl-terminal arginine. This behavior is what one would expect if residue 15 of this peptide were an asparagine and the asparaginyl–glycine bond had rearranged to form a β-peptide bond (Hermodson et al., 1972). It is known that this rearrangement can occur during the course of a sequencing run, but in this case it appears to have been preformed, as the peptide was resistant to digestion by hydroxylamine at pH 9. Attempts to obtain better Edman degradations through this region after subdigestion of R3 were unsuccessful.

The blocked amino-terminal peptide was isolated from a separate digest of 100 nmol of citraconyl CM-FNR. Immediately after digestion, the solution was lyophilized, decitraconylated, and fractionated by reversed-phase HPLC to yield two peaks containing amino-terminal peptides R1a and R1b at 25 and 23% acetonitrile, respectively. These were further purified by ion-exchange HPLC, resulting in final yields of 20 and 30%, respectively. Their amino acid compositions revealed that R1a had an additional alanine, isoleucine, and glutamic acid (Table I). R1a was completely blocked at the amino terminus, but sequence analysis of R1b showed it to be a 30:70 mixture of the peptides beginning at Ser-4 and Asp-5. The sequence of R1b provided the overlap of FNRα to R1'. For analysis of the blocking group, peptide R1a was subdigested with trypsin to yield a peptide comprising residues 1–15. This peptide was then further subdigested with subtilisin to give the blocked tetrapeptide R1a-N1. Twenty nanomoles of this peptide was analyzed by NMR to identify the blocking group as pyroglutamic acid (Figure 3).

Methionine Fragments. Digestion of [^{3}H]CM-FNR with cyanogen bromide yielded a complete set of peptides, save the blocked amino-terminal peptide. After digestion and lyophilization the digest was initially fractionated by gel filtration chromatography (Figure 4). As with the R fragments, each pool was further fractionated by reversed-phase HPLC at pH 2. This yielded all M fragments in pure form except M9 and M11. These were finally purified by reversed-phase chro-

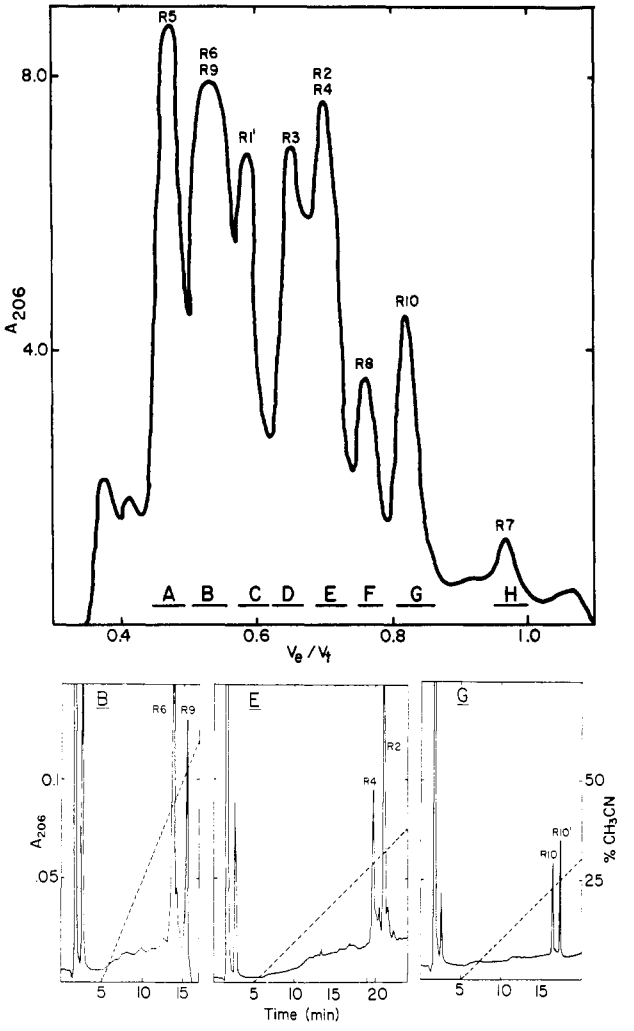

FIGURE 2: Isolation of peptides produced by cleavage at arginyl residues. Upper panel: Five milliliters of digest on a Sephadex G-50 superfine column (2 × 190 cm) in 0.1 M NH_4HCO_3, pH 8.8. The flow rate was 11 mL/h. All pooled fractions (A–H) were rechromatographed by reversed-phase HPLC at pH 2. The arginine peptides found in each pool are indicated by using the peptide nomenclature defined in Figure 1. Panels B, E, and G: Elution profiles for reversed-phase HPLC separations of 0.25% of fractions B, E, and G. These were the only pools that showed more than one major peak. The elution gradients were 5%/min modifier for pool B and 2%/min for pools E and G. Peptide R10′ in panel G has the same sequence as R10, but the amino-terminal glutaminyl residue has formed a pyroglutamyl residue.

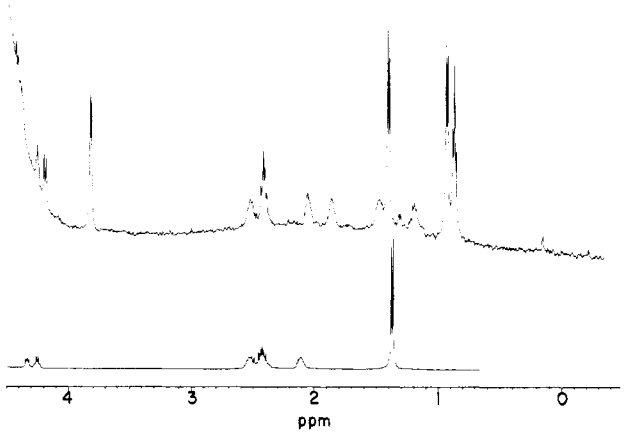

FIGURE 3: Identification of pyroglutamyl residue by NMR spectroscopy of R1a-N1 (upper spectrum) and pyroglutamylalanine (lower spectrum). Chemical shift assignments are as follows: 0.87 (Ile δCH_3), 0.93 (Ile γCH_3), 1.20 and 1.49 (Ile γCH_2), 1.40 (Ala βCH_3), 1.87 (Ile βCH), 2.05, 2.42, and 2.52 (pyroglutamyl βCH_2 and γCH_2), 3.83 (Ser βCH_2), 4.20 (Ile αCH), 4.26 (Ser αCH), 4.38 (pyroglutamyl αCH), and 4.43 ppm (Ala αCH). The shifts assigned to all side-chain hydrogens are within 0.055 ppm of those reported by Bundi & Wüthrich (1979) or in the case of the pyroglutamyl residue to those presented here. In addition, shifts obtained for the pyroglutamyl side-chain hydrogens are characteristically different from the expected shifts for glutamyl (1.969, 2.092, 2.28, 2.31 ppm) or glutaminyl (2.01, 2.13, 2.38 ppm) residues, making identification conclusive. All shifts are given in parts per million relative to an external standard of 4,4-dimethyl-4-silapentane-1-sulfonate.

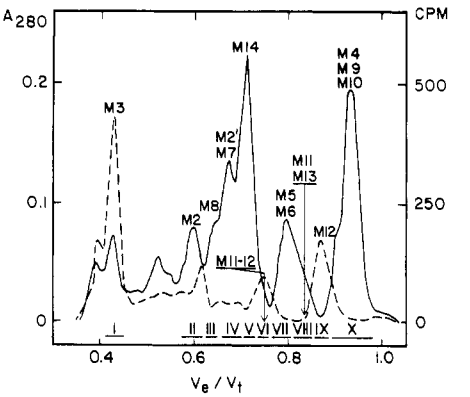

FIGURE 4: Primary separation of peptides after cleavage at methionyl residues. After treatment with CNBr, 1200 nmol of [^{3}H]CM-FNR was taken up in 3 mL of 9% HCOOH and loaded onto a Sephadex G-50 superfine column (2 × 190 cm) equilibrated with the same solvent. The flow rate was 11 mL/h. The solid line indicates absorbance at 280 nm and the dashed line traces the radioactivity of 0.1% of each sample. Pools I–X were further purified by reversed-phase HPLC.

matography at pH 6.5. Most of the M peptides were recovered at yields near 30%. In addition to the unique peptides M2 through M14, an overlap peptide M11–12 was recovered at a yield of 10%, as was a 28-residue peptide (M2′ in Figure 4) resulting from an anomalous (proteolytic?) cleavage after Phe-32.

Direct sequence analysis of M2 and M3 provided the overlap for R1′ through R3. Peptides M3-E1 and M3-E2 (produced by anomalous cleavages after Ser-96 and Cys-132, respectively) were isolated from an *S. aureus* V8 subdigest of M3 (Figure 5) to provide overlaps for R3 through R5. The extreme hydrophobicity of the carboxyl-terminal six residues of M6 made its sequence analysis very difficult. Only after covalent attachment of the peptide to glass through the carboxyl-terminal homoserine (Horn & Laursen, 1973) were we able to obtain sequence beyond Arg-178. This sequence together with that of R5-T1 provided the overlap for R5 to R6. The peptide M8 overlapped R6 through R8 and M14 overlapped R9 and R10. Peptides M14 and R10 proved the sequence of the carboxyl terminus as they contained no homoserine and arginine, respectively, and had the same carboxyl-terminal sequence.

Analysis of the M fragments did not reveal any corresponding to the amino terminus of the intact protein (residues 1–23). Furthermore, whereas digestion with *S. aureus* V8 protease of either the whole protein or a CNBr digest thereof yielded peptide E1 (residues 8–17), no fraction from a Sephadex G-50 column of the same CNBr digest yielded E1 on subdigestion. It was finally shown that the amino-terminal M fragment binds to glass in either 9% formic acid or 0.1 M NH_4HCO_3 and that it had been lost during purification of the other M fragments.

Other Peptides. An *S. aureus* V8 digest of 100 nmol of CM-FNR yielded a precipitate containing a hydrophobic fragment starting at Met-155 and probably ending at Glu-187.

Table I. Amino Acid Compositions[a] of FNR and Its Tryptic Peptides[b]

		R1a	R1b	R1'	R2[c]	R3	R4[d]	R5	R6[e]	R7	R8	R9	R10	Whole Protein[f]
Residues		1-41	4-41	16-41	42-71	72-93	94-117	118-178	179-228	229-235	236-250	251-301	302-314	
Asp/Asn	(D/N)	1.9(2)	1.7(2)	1.1(1)	3.0(3)	2.8(3)	2.0(2)	5.9(6)	4.8(5)	1.0(1)	1.0(1)	4.7(5)	1.0(1)	29.5(29)
Thr	(T)	1.7(2)	1.8(2)	2.0(2)	2.7(3)			5.7(6)	1.1(1)		2.2(2)	1.1(1)		14.5(15)
Ser	(S)	1.9(2)	1.4(2)[g]	0.9(1)	0.7(1)	0.9(1)	4.1(5)	1.0(1)	3.2(4)	1.1(1)		1.0(1)		16.4(16)
Glu/Gln	(E/Q)	4.8(5)	3.7(4)	3.0(3)	2.9(3)	3.0(3)		3.1(3)	4.8(5)		4.9(5)	6.2(6)	3.5(4)	36.4(34)
Pro	(P)	5.8(6)	5.7(6)	2.1(2)	2.1(2)	2.0(2)		5.3(5)	2.4(2)					16.5(17)
Gly	(G)	2.4(2)	2.0(2)	2.0(2)	2.9(3)	3.8(4)	2.0(2)	7.8(8)	2.0(2)	0.4	1.1(1)	4.1(4)		26.0(26)
Ala	(A)	4.0(4)	3.2(3)		1.0(1)		2.9(3)	4.0(4)	1.9(2)	1.0(1)		5.0(5)	0.9(1)	21.2(21)
Cys[h]	(C)				0.9(1)		1.0(1)	2.0(2)				1.1(1)		5.5(5)
Val	(V)	3.7(4)	3.5(4)	2.8(3)	0.9(1)	1.7(2)	2.1(2)	2.7(3)	0.8(1)	0.9(1)		2.7(3)	1.9(2)	16.6(19)
Met	(M)	0.8(1)	0.8(1)	0.9(1)	0.9(1)			2.7(3)	1.5(2)		0.9(1)	4.7(5)		10.9(13)
Ile	(I)	1.7(2)	1.2(1)	1.0(1)	1.8(2)	0.7(1)	1.0(1)	4.3(5)			1.0(1)	3.7(4)		14.1(16)
Leu	(L)				2.0(2)	1.0(1)	3.0(3)	6.0(6)	5.9(6)	1.0(1)		4.3(4)	1.0(1)	24. (24)
Tyr	(Y)	0.8(1)	0.9(1)	1.0(1)	0.9(1)		0.9(1)	0.9(1)	1.9(2)		1.0(1)	4.2(4)	1.0(1)	9.0(12)
Phe	(F)	1.0(1)	0.9(1)	0.9(1)	0.7(1)		0.6(1)	2.0(2)	7.0(7)	1.0(1)		0.3		10.1(13)
His	(H)	1.0(1)	1.0(1)	0.9(1)	1.8(2)	0.9(1)			1.0(1)					5.3(5)
Lys	(K)	6.3(7)	6.3(7)	5.5(6)	1.2(1)	2.8(3)	2.2(2)	4.7(5)	6.5(7)		2.1(2)	4.9(5)	1.8(2)	32.5(34)
Arg	(R)	1.1(1)	1.0(1)	1.0(1)	1.0(1)	1.0(1)	1.0(1)	1.0(1)	1.0(1)	1.0(1)	1.0(1)	1.0(1)		9.5(9)
Trp[i]	(W)				(1)				(2)			(2)	(1)	(6)
Yield (%)		20	30	63	92	75	90	51	72	60	69	46	51	

[a]Residues per molecule by amino acid analysis (values < 0.3 not reported) or in parentheses from the sequence.

[b]Cleavage was made specific for arginyl residues by citraconylation of the lysyl residues.

[c]Values are corrected for contamination with 0.16 mole fraction R4.

[d]Values are corrected for contamination with 0.13 mole fraction R2.

[e]Values are corrected for contamination with 0.17 mole fraction R9.

[f]Hydrolysis of FNR was for 24, 40 and 96 h. Thr and Ser values were extrapolated to t = 0. Val and Ile values for 96 h are reported. Carboxymethyl-cysteine was determined from a separate 24 h hydrolysis.

[g]Theoretical value for serine is 1.7 based on a 30:70 ratio of peptides starting at residues 3 and 4 respectively.

[h]Cysteine identified as carboxymethyl-cysteine.

[i]Tryptophan was not quantitated.

The soluble material was acidified and fractionated by reversed-phase HPLC. Selected peaks eluting near 12, 22, and 21% CH_3CN were rechromatographed at pH 6.5 to give pure peptides E1, E2, and E3, respectively. E1 was useful in defining the amino-terminal region, E2 overlapped R8 to R9, and E3 confirmed the tryptophan at position 296. A tryptic digest of 100 nmol of CM-FNR yielded peptides T1 and T2, eluting near 23 and 25% CH_3CN, respectively, which confirmed portions of the sequence obtained for R2 and R4.

Location of Disulfide Bridge. CM-FNR, radiolabeled selectively on the half-cystine residues with [^{14}C]carboxymethyl groups (see Materials and Methods), was treated with cyanogen bromide, and the products were separated by gel filtration (Figure 5, inset). A peak corresponding to M3 had over 90% of the radioactivity. Extensive subdigestion of the pooled radioactive fractions with *S. aureus* V8 protease caused some anomalous cleavage after serine, and reversed-phase HPLC yielded two radioactive peptides (Figure 5). Composition and sequence analyses clearly identified these peptides as TIKGVCS (residues 127–133) and NFLCD (residues 134–138). The specific radioactivity of the two CM-cysteines were within 5% of each other. It is not clear why the unusual cleavage after serine took place, but under these conditions cleavage at other serine residues was also observed, such as in the generation of peptide M3-E1 (residues 73–96).

Discussion

Summary of Proof. The proof of the primary structure of spinach FNR is derived mostly from peptides generated by cleavage after arginyl and methionyl residues. The remaining overlap peptides come from digests selected on the basis of partial sequences. Over 95% of the residues have been sequenced in more than one peptide.

All of the overlaps in the proof presented here are at least three residues long. The identity of residue 86 as presented is unconfirmed. The composition data and the behavior of the peptide during sequencing suggest that this residue is an asparagine, but a direct chemical identification has not been made. Pending identification, the possibility remains that this is an aspartic acid that has undergone some other unusual posttranslational modification. It is even possible that this aspartyl residue is in a β-peptide bond in the native protein.

Another uncertainty in the sequence of FNR relates to the identification of the amino terminus as pyroglutamic acid. It

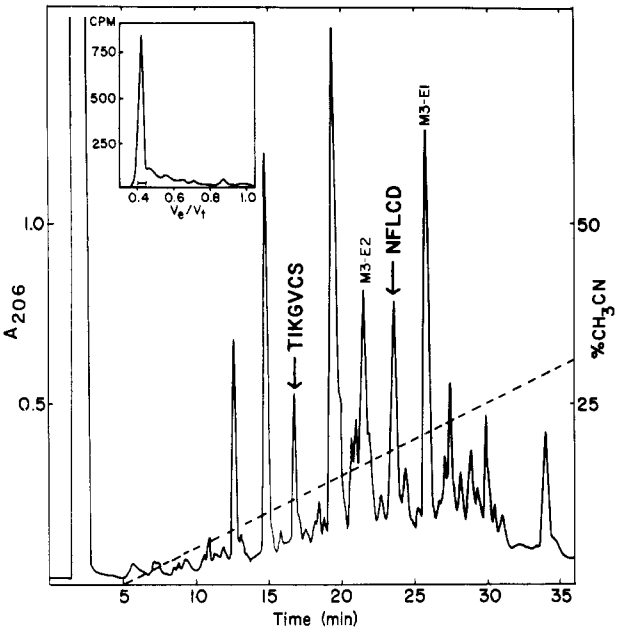

FIGURE 5: Separation of CNBr peptides specifically labeled at a cystine residue. FNR (90 nmol) was first treated with unlabeled iodoacetate to block free sulfhydryls and then with dithiothreitol and [^{14}C]-iodoacetate to specifically radiolabel cysteine residues derived from a cystine residue (see Materials and Methods). Inset: After CNBr treatment, the material was dissolved in 1 mL of 0.1 M NH_4HCO_3 and chromatographed on a Bio-Rad P-10 column (0.9 × 185 cm) equilibrated in the same buffer. Fractions (2 mL) were collected at a flow rate of 2 mL/h, and 1% of each was used for radiolabel detection. The pooled radioactive fractions were first digested with 2% mole fraction of *S. aureus* V8 protease in NH_4HCO_3 for 15 h at 37 °C. An analytical reversed-phase HPLC run at pH 2 indicated that the digest was not complete, so the solution was lyophilized, redissolved in 0.1 M K_2HPO_4, pH 7.8, and incubated 18 h at 37 °C with an additional 5% mole fraction of enzyme. The digest was acidified and separated by reversed-phase HPLC at pH 2. Each radioactive peak, containing approximately 1.5×10^5 cpm, is indicated with an arrow and with the derived sequence of the radioactive peptide. Peaks corresponding to two other peptides (M3-E1 and M3-E2) used in the proof of structure (Figure 1) are also identified. Each of these pooled fractions was further purified by reversed-phase HPLC at pH 6.5 before composition and sequence analyses.

is conceivable that the purported amino terminus is not the true one, but was produced by the tryptic digestion. In this case there would be at least one other tryptic peptide of undetermined length on the amino side of the sequence presented here. We believe this is unlikely due to the agreement of the calculated and observed amino acid compositions (Table I) and of the molecular weight of the protein (see below). In addition, one would need to explain why the precursor glutamine at position 1 was found to be 100% cyclized in R1a whereas that at the amino terminus of R10 was only ~50% cyclized (Figure 2). Even if the pyroglutamyl residue is the amino terminus of FNR after isolation, it is still possible that FNR exists in the chloroplast in a longer form that undergoes amino-terminal degradation and subsequent cyclization of an exposed glutamine.

The sequence of FNR presented here contains only five cysteines. Amino acid compositions in this and other laboratories consistently gave values closer to six than five. Also amperometric titration of the enzyme (Zanetti & Forti, 1969) and covalent modification studies (Valle et al., 1982) suggested that native FNR has four cysteines and one cystine. The explanation for this discrepancy is not clear. The firmest evidence supporting the existence of only five cysteinyl residues in FNR comes from the compositions of the tryptic fragments (R1a through R10 in Table I) that span the entire sequence. Also the radioactivity profiles during gel filtration chromatography of tryptic digests of [^{14}C]CM-FNR (with and without citraconylation) and of CNBr digests are consistent with the proposed cysteine placements.

The amino acid composition of the whole protein is in reasonable agreement with that obtained by sequence analysis (Table I). The molecular weight calculated from the sequence is 35 317. This agrees well with the results of SDS–PAGE and ultracentrifugation analyses (Keirns & Wang, 1972; Sheriff et al., 1980), which both indicate a molecular weight range of 33 000–37 000.

Heterogeneity of FNR. The sequence analysis revealed no evidence of allotypic heterogeneity. With the exception of asparagine/aspartate and glutamine/glutamate substitutions, we believe this would have been observable at a level of 5–10%. Thus, as was suggested by the reported interconversion of isoelectric species, allotypic sequence differences do not account for the multiple species observed during isoelectric focusing. Other possibilities are differential ion binding of some sort (but this would have to be stable in 6 M urea), varying oxidation states of surface residues (Ellefson & Krogmann, 1979), or possibly various conformational states that are in slow equilibrium with one another. The possibility of differential ion binding is supported by NMR experiments that suggest the existence of a very tightly bound phosphate ion (Ulrich & Markley, 1980). Also, we have observed a correlation between the presence of phosphate and the ease of crystallization of FNR.

Sequence analyses of the intact protein indicated that some of our preparations contained FNR that had been degraded to various extents at the amino terminus. Forms of FNR were obtained that started at residues 1 (blocked), 2, 4, 5, 6, 8, and 16. The observation that some preparations yield only blocked FNR suggests that amino-terminal heterogeneity is not intrinsic to the native protein but is the result of endo- and exopeptidase activity in the harvested plant or during the preparation. Although a quantitative relationship has not been proven, our results suggest that the upper band on SDS–PAGE is mostly intact FNR and the lower band a mixture of degraded species of FNR.

This explanation of the heterogeneity is consistent with the results obtained by Gozzer et al. (1977) and Hasumi et al. (1983), both of whose preparations gave doublets on SDS–PAGE. Gozzer et al. reported that the lower molecular weight band had the more basic p*I*, and the sequence results predict that loss of 3–7 residues would yield a more basic protein due to both the generation of a free α-amino group and the loss of Asp-5 and Glu-7. Hasumi et al. found that two of their purified isoelectric species gave multiple amino termini including serine, aspartate, valine, and alanine.

Sequence Comparisons. Figure 6 shows an alignment of the sequence of FNR presented here with sequences determined for FNR from two other sources. A partial sequence derived from Japanese spinach appears to align with ours beginning at Val-16 but differs in 5 out of 25 residues. The two proteins are clearly related, but the extent of the differences is surprisingly high for what was thought to be the same protein from two strains of the same species. However, there are certain morphological differences between the plants, suggesting that some divergence has occurred. Apparently more discrimination must be used in defining the source of FNR. Hasumi et al. (1983) suggested that the preponderance of basic residues in their sequence may be related to the interaction with ferredoxin. Although this may be true, it should be noted that three of the nine basic residues are not conserved.

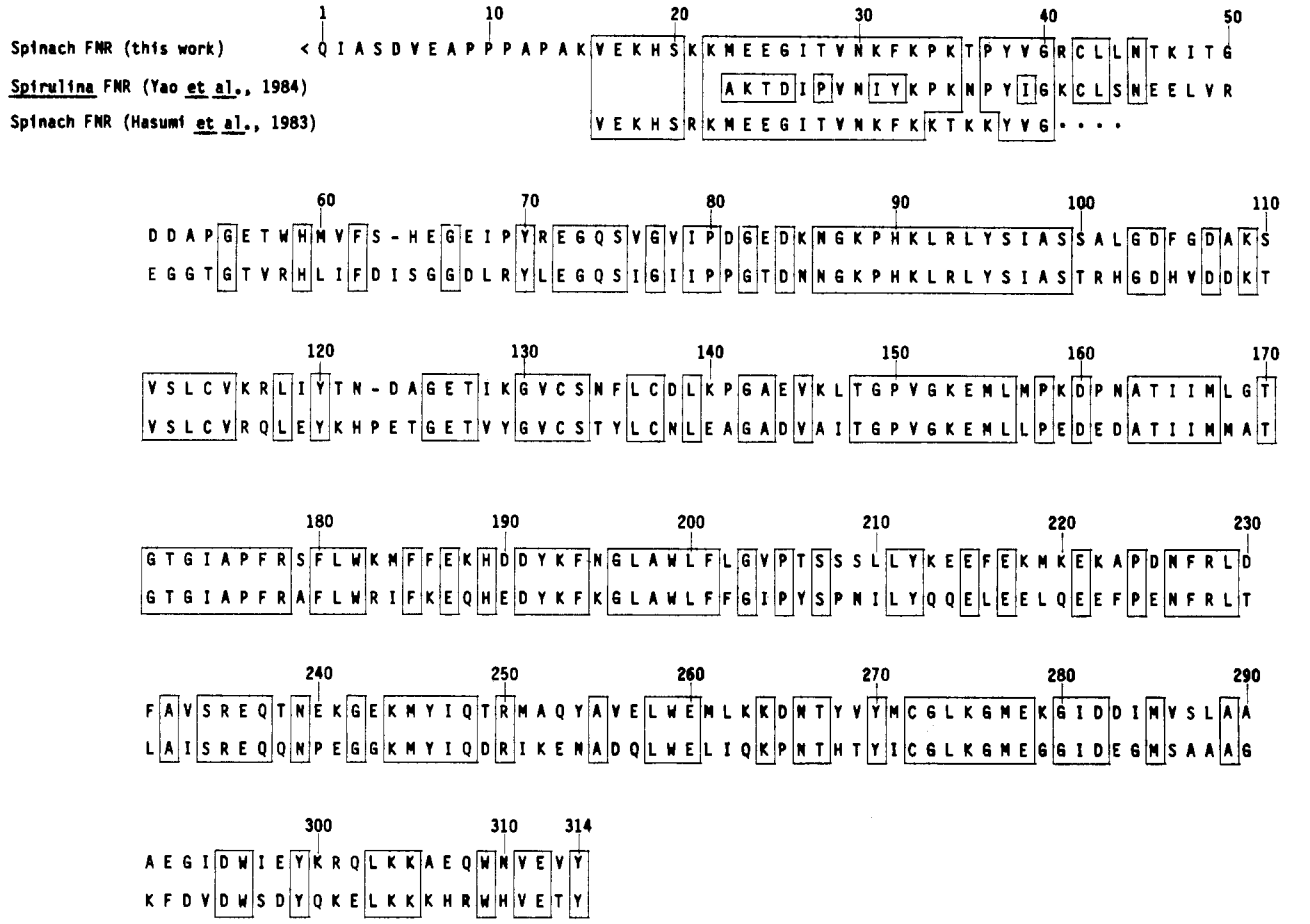

FIGURE 6: Alignments of known FNR sequences. The sequence from spinach FNR presented here (top line) is aligned with the complete sequence reported for the blue-green algae *Spirulina* (Yao et al., 1984) (second line) and with the amino terminus of FNR from Japanese spinach (Hasumi et al., 1983) (third line of upper set). The numbering system corresponds to that in Figure 1. A single residue gap by residue 63 is indicated by a dash and an incomplete sequence by "...". Identical residues are enclosed in boxes.

The alignment presented for *Spirulina* FNR shows identity of 161 residues (55%). Segments as long as 14 residues are identical in sequence. The homologous relationship is found throughout the two molecules, although the sequence reported here has a 22-residue extension on the amino-terminal side of the sequence of FNR from *Spirulina*. It is possible that this extension is related to the mechanics of transport into the chloroplast, which is required in spinach but not in blue-green algae, which lack chloroplasts.

The sequence of FNR was tested for similarity to other proteins of known sequence by searching the protein sequence data base of the National Biomedical Research Foundation (Jan 1984 revision). The search was done in lengths of 25 residues by using the program SEARCH (Dayhoff, 1979) on a VAX 11/780. No sequences were found that gave scores outside of the random distribution expected from statistics of unrelated proteins.

Despite the lack of extended similarity of this sequence with those of other proteins, certain segments of the sequence merit comment. Figure 7 shows an alignment of the sequence surrounding the disulfide of FNR with those surrounding the redox-active disulfides of the flavoproteins lipoamide dehydrogenase and glutathione reductase. In this alignment, there are from four to six identities out of 12 residues. This is not statistically significant, but it is enough to suggest the possibility of a functional and possibly an evolutionary relationship. Early reductive titrations of FNR indicated that there may be a second redox-active group in the protein (Massey, 1968). However, later studies in the presence of catalytic amounts of flavodoxin indicated only a single redox-active group (Massey et al., 1970).

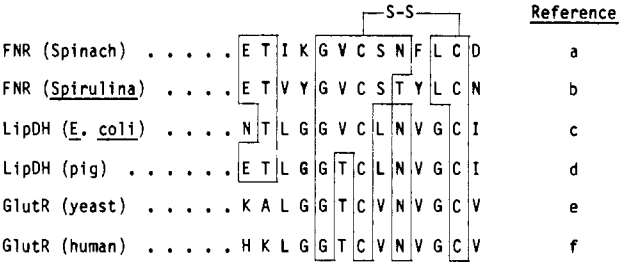

FIGURE 7: Comparison of FNR disulfide region with the FAD-linked, redox-active disulfide regions from lipoamide dehydrogenase (LipDH) and glutathione reductase (GlutR). References are as follows: (a) present work, (b) Yao et al., 1984, (c) Brown & Perham, 1972, (c) Matthews et al., 1974, (e) Jones & Williams, 1975, and (f) Untucht-Grau et al., 1981.

Among the many nucleotide binding proteins of known structure, there is very little sequence identity (Rossmann et al., 1975). However, one invariant residue among all nucleotide binding folds is the glycine at the end of the first strand of the parallel pleated sheet (Rossmann et al., 1975). In addition, it has been pointed out by Untucht-Grau et al. (1981) that there tends to be a high content of glycine among the five residues that follow the invariant glycine and form the so-called pyrophosphate binding loop. This was particularly striking in the FAD domain of glutathione reductase where five of the six residues are glycine (Untucht-Grau et al., 1981). Inspection of the sequence of FNR reveals that residues 169–173 (GTGTG) constitute the most glycine-rich segment of the protein. That this segment may be a pyrophosphate binding

loop is supported by secondary structural predictions (Chou & Fasman, 1974) of a strand of β-sheet (residues 164–168) and a helix (residues 176–187). Together these observations suggest a β–α structure characteristic of nucleotide binding folds. In addition, the electron density at 3.7-Å resolution for the strand of sheet and the helix that form the NADPH binding site (Sheriff & Herriott, 1981) is entirely consistent with the amino acid sequence presented here. When the glycine-rich loop was used as a starting point, the carboxyl-terminal half of the sequence was fit relatively easily to the electron density map (3.25-Å resolution) in the NADP binding domain (unpublished results). The amino-terminal half of the sequence presumably makes up the FAD domain, but an unambiguous fit to the density remains elusive. The sequences of greatest similarity between the enzymes of spinach and blue-green algae (e.g., residues 86–99 in Figure 6) are likely candidates for FAD interaction regions.

Acknowledgments

We thank L. Ericsson, R. Wade, Dr. K. Titani, and Dr. K. Takio for many helpful discussions during the course of this work. We also thank L. Ericsson and R. Wade for their help with Edman degradations and amino acid analyses and Dr. R. E. Klevit, Dr. D. Wemmer, and P. Weber for suggesting, obtaining, and processing the NMR spectra.

References

Berzborn, R. J. (1969) *Z. Naturforsch., B: Anorg. Chem., Org. Chem., Biochem., Biophys., Biol. 24B*, 436.

Böhme, H. (1977) *Eur. J. Biochem. 72*, 283.

Böhme, H. (1978) *Eur. J. Biochem. 84*, 87.

Bornstein, P., & Balian, G. (1977) *Methods Enzymol. 47*, 132.

Brauer, A. W., Margolies, M. N., & Haber, E. (1975) *Biochemistry 14*, 3029.

Bridgen, P. J., Cross, G. A. M., & Bridgen, J. (1976) *Nature (London) 263*, 613.

Brown, J. P., & Perham, R. N. (1972) *FEBS Lett. 26*, 221.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285.

Carrillo, N., Lucero, H. A., & Vallejos, R. H. (1981) *J. Biol. Chem. 256*, 1058.

Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry 13*, 222.

Curtis, V. A., Siedow, J. N., & San Pietro, A. (1973) *Arch. Biochem. Biophys. 158*, 898.

Dayhoff, M. O. (1979) in *Atlas of Protein Sequence Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 3, pp 1–8, National Biomedical Research Foundation, Washington, DC.

Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1*, 80.

Ellefson, W. L., & Krogmann, D. W. (1979) *Arch. Biochem. Biophys. 194*, 593.

Ericsson, L. H., Wade, R. D., Gagnon, J., MacDonald, R. M., Granberg, R. R., & Walsh, K. A. (1977) in *Solid Phase Methods in Protein Sequence Analysis* (Previero, A., & Coletti-Previero, M. A., Eds.) p 137, Elsevier/North-Holland, New York.

Firenzuoli, A. M., Ramponi, G., Vanni, P., & Zanobini, A. (1968) *Life Sci. 7 (Part 2)*, 905.

Forti, G., & Rosa, L. (1971) *FEBS Lett. 18*, 55.

Gozzer, C., Zanetti, G., Galliano, M., Sacchi, G. A., Minchiotti, L., & Curti, B. (1977) *Biochim. Biophys. Acta 485*, 278.

Grossman, A. R., Bartlett, S. G., Schmidt, G. W., Mullet, J. E., & Chua, N.-H. (1982) *J. Biol. Chem. 257*, 1558.

Hasumi, H., Nagata, E., & Nakamura, S. (1983) *Biochem. Biophys. Res. Commun. 110*, 280.

Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., & Walsh, K. A. (1972) *Biochemistry 11*, 4493.

Horn, M. J., & Laursen, R. A. (1973) *FEBS Lett. 36*, 285.

Hunkapiller, M. W., Hewick, R. M., Dreyer, W. J., & Hood, L. E. (1983) *Methods Enzymol. 91*, 399.

Hutber, G. N., Smith, A. J., & Rogers, L. J. (1978) *Biochem. Soc. Trans. 6*, 1214.

Jones, E. T., & Williams, C. H., Jr. (1975) *J. Biol. Chem. 250*, 3779.

Keirns, J. J., & Wang, J. H. (1972) *J. Biol. Chem. 247*, 7374.

Keister, D. L., San Pietro, A., & Stolzenbach, F. E. (1960) *J. Biol. Chem. 235*, 2989.

Mahoney, W. C., & Hermodson, M. A. (1980) *J. Biol. Chem. 255*, 11199.

Manzano, C., Candau, P., & Guerrero, M. G. (1978) *Anal. Biochem. 90*, 408.

Massey, V. (1968) in *Flavins and Flavoproteins* (Yagi, K., Ed.) pp 7–12, University of Tokyo Press, Tokyo.

Massey, V., Matthews, R. G., Foust, G. P., Howell, L. G., Williams, C. H., Zanetti, G., & Ronchi, S. (1970) in *Pyridine Nucleotide-Dependent Dehydrogenases* (Sund, H., Ed.) p 393, Springer-Verlag, West Berlin.

Matthews, R. G., Arscott, L. D., & Williams, C. H., Jr. (1974) *Biochim. Biophys. Acta 370*, 26.

Riov, J., & Brown, G. (1976) *Physiol. Plant. 38*, 147.

Rossmann, M. G., Liljas, A., Brändén, C. I., & Banaszak, L. J. (1975) *Enzymes, 3rd Ed. 11*, 61.

Shahak, Y., Crowther, D., & Hind, G. (1981) *Biochim. Biophys. Acta 636*, 234.

Sheriff, S., & Herriott, J. R. (1981) *J. Mol. Biol. 145*, 441.

Sheriff, S., Teller, D. C., & Herriott, J. R. (1980) *Arch. Biochem. Biophys. 205*, 499.

Shin, M., Tagawa, K., & Arnon, D. I. (1963) *Biochem. Z. 338*, 84.

Sluiters-Scholten, C. M. Th., Moll, W. A. W., & Stegwee, D. (1977) *Planta 133*, 289.

Susor, W. A., & Krogmann, D. W. (1966) *Biochim. Biophys. Acta 120*, 65.

Ulrich, E. L., & Markley, J. L. (1980) Abstracts of the 9th International Conference on Magnetic Resonance in Biological Systems, Bendor, France, p 100.

Untucht-Grau, R., Schirmer, R. H., Shirmer, I., & Krauth-Siegel, R. L. (1981) *Eur. J. Biochem. 120*, 407.

Valle, E. M., Carrillo, N., & Vallejos, R. H. (1982) *Biochim. Biophys. Acta 681*, 412.

Wada, K., Tamura, T., Matsubara, H., & Kodo, K. (1983) *J. Biochem. (Tokyo) 94*, 387.

Wagner, R., Carrillo, N., Junge, W., & Vallejos, R. H. (1982) *Biochim. Biophys. Acta 680*, 317.

Yao, Y., Tamura, T., Wada, K., Matsubara, H., & Kodo, K. (1984) *J. Biochem. (Tokyo) 95*, 1513.

Zanetti, G., & Forti, G. (1969) *J. Biol. Chem. 244*, 4757.