Plant Molecular Biology **21**: 1023–1033, 1993.

Sequence analysis of pre-ferredoxin-NADP$^+$-reductase cDNA from *Cyanophora paradoxa* specifying a precursor for a nucleus-encoded cyanelle polypeptide

Johannes Jakowitsch[1], Manfred G. Bayer[2], Thomas L. Maier[2], Angela Lüttke[3], Ulrike B. Gebhart[2], Martin Brandtner[1], Barbara Hamilton[1], Christoph Neumann-Spallart[1], Christine B. Michalowski[4], Hans J. Bohnert[4], Hainfried E. A. Schenk[2] and Wolfgang Löffelhardt[1]*

[1] *Institut für Biochemie und Molekulare Zellbiologie der Universität Wien und Ludwig Boltzmann-Forschungsstelle für Biochemie, A-1030 Vienna, Austria (*author for correspondence);* [2] *Botanisches Institut, Lehrstuhl für Allgemeine Botanik und Pflanzenphysiologie, Universität Tübingen, D-7400 Tübingen, Germany;* [3] *Arnimstr. 2, D-5000 Köln, Germany;* [4] *Department of Biochemistry, University of Arizona, Tucson, AZ 85721, USA*

Received 20 August 1992; accepted in revised form 23 December 1992

Key words: cyanelles, *Cyanophora paradoxa*, peptidoglycan, *petH*, pre-ferredoxin-NADP$^+$ reductase, protein import

Abstract

A cDNA clone for pre-ferredoxin-NADP$^+$ reductase (FNR) was obtained by screening a *Cyanophora paradoxa* expression library with antibodies specific for cyanelle FNR. The 1.4 kb transcript was derived from a single-copy gene. The precursor (41 kDa) and mature forms (34 kDa) of FNR were identified by western blotting of *in vitro* translation products and cyanelle extracts, respectively. The derived amino acid sequence of the mature form was corroborated by data from N-terminal protein sequencing and yielded identity scores from 58% to 62% upon comparison with cyanobacterial FNRs. Sequence conservation seemed to be even more pronounced in comparison with enzymes from higher plants, but using the neighbor joining method the *C. paradoxa* sequence was clearly positioned between the prokaryotic and eukaryotic sequences. The transit peptide of 65 or 66 amino acids appeared to be totally unrelated to those from spinach, pea and ice plant but showed overall characteristics of stroma-targeting peptides.

Introduction

Cyanelles from the protist *Cyanophora paradoxa*, recently also termed cyanoplasts [30], are distinguished among plastids by the retention of the envelope structure of their cyanobacterial ancestors [1, 31]. The 130 kb cyanelle genome, though harboring a large number of genes whose protein

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X66372.

products are nucleus-encoded in higher plants [6], nevertheless cannot account for all necessary polypeptides. Thus, like chloroplasts, cyanelles have to import a multitude of cytoplasmically synthesized proteins [4, 5], most or all of them probably as precursors [35]. The respective apparatus has to achieve preprotein translocation across two membranes and an intermediate peptidoglycan layer. *C. paradoxa* can be considered as a model system for protein trafficking in the early stages of plastid evolution: the cyanobacterial endosymbiont, rendered semiautonomous through partial transfer of its genes to the host cell nucleus still retained its cell wall. A preprotein import machinery evolved, possibly from phage receptors and/or bacteriocin transport systems [14] whereas the pre-existing prokaryotic protein translocation apparatus [40] was maintained, possibly for subtargeting within the organelle [14, 35]. This apparatus might originally have included translocation of periplasmic and outer membrane proteins whereas the recent chloroplasts likely use the machinery for the sole purpose of translocating proteins into the thylakoid lumen.

Transit sequences of between 30 and more than 80 amino acids [38] have been found which are both necessary and sufficient to direct plastid polypeptides as well as chimeric passenger proteins into the stroma. There is no consensus sequence among the relatively hydrophilic transit peptides; hydroxylated and basic amino acids prevail and the postulated random coil structure [37] would imply the participation of chaperones [9] in the recognition and binding process.

While our knowledge on cyanelle gene organization renders the cyanelle genome the best characterized one among chlorophyll *b*-less plastids [6, D. Bryant and V. Stirewalt, personal communication], nothing is known about nuclear genes for cyanelle polypeptides. Many studies have indicated a bridge role for cyanelles between cyanobacteria and the plastids from higher plants based on the homology of genes [24]. We were interested to study whether *Cyanophora* also occupied an intermediary position with respect to protein translocation mechanisms and protein targeting signals as far as they can be deduced from studies of genes.

Especially interesting is the question about the nature and structure of presumptive targeting signals: will they resemble those found to achieve transport into plastids or will these signals be similar to the hydrophobic bacterial leader sequences [40]. Bacterial leader sequences direct preproteins through the inner membrane though in the opposite direction as compared to cyanelle import. There is no evidence that a membrane potential, shown to be essential for bacterial protein export *in vivo* [40] does exist across the inner envelope membrane of cyanelles or of plastids. To begin to answer questions on the nature of protein import mechanisms in cyanelles and on the evolution of our organelle protein translocase, we wanted to investigate a nucleus encoded cyanelle protein [4]. We report here on the initial characterization of a pre-ferredoxin-$NADP^+$ reductase cDNA from *C. paradoxa*. The cDNA specifies a precursor protein that *in vivo* crosses the two envelope membranes and the cyanelle peptidoglycan wall, a structure that is unique among eukaryotic organelles.

Materials and methods

C. paradoxa LB555UTEX was grown and poly(A)$^+$ and cyanelle RNAs were isolated as described [3, 18]. A cDNA library in the vector λ ZAP [34] was established using a cDNA synthesis kit and a cDNA cloning kit obtained from Stratagene. The library was amplified prior to the screening procedure. Purification of cyanelle FNR [13] and raising of specific antibodies has previously been reported [4].

Screening of the library was performed using peroxidase-conjugated anti-rabbit IgG as a secondary antibody and 4-chloro-1-naphthol as substrate, essentially as described [15] except that blocking and washing procedures followed the protocol of Bayer [2]. Immunopositive plaques were picked and rescreened twice after which all plaques gave a positive signal. Recombinant phage containing the FNR cDNA were converted

into Bluescript SK$^-$ plasmids by the Lambda ZAP automated excision process using the helper phage R408 [34]. Plasmid pF10, assumed to contain a full-size insert was chosen for sequence analysis. The supercoil sequencing method [39] using a T7 polymerase sequencing kit (Pharmacia) was adopted on subclones generated by cleavage of the original clones with selected restriction enzymes (*Sac* I, *Sal* I, *Sma* I, *Xho* I) followed by religation and on exonuclease III deletion subclones [16] obtained with the 'Erase-a-Base' kit from Promega. Computer analysis of DNA sequences was conducted using the Dave Mount program 5.07 from the University of Arizona. Alignment of FNR amino acid sequences was performed with the program package MacPROT [23]. Hydropathy plots using the Kyte and Doolittle scale [22] and amino acid frequency calculations were done with programs as described [23]. Gene trees were constructed according to the neighbor-joining method [28] using a distance matrix based on a log-normal model and the Fitch and Margoliash method [12] with a distance matrix based on the minimal number of mutations required for amino acid substitutions. In distance calculations, gaps were either excluded or treated as individual insertion/deletion events. Robustness of trees inferred by the neighbor-joining method was tested with bootstrapped samples [10].

Northern and Southern hybridizations using an internal 870 bp *Sst* II fragment ranging from position 115 to position 984 (Fig. 3) as a probe were performed as described [18].

In vitro translation was performed using poly(A)$^+$ RNA from *C. paradoxa* and a commercially available wheat germ extract (Boehringer Mannheim). After incubation at 30 °C for 60 min aprotinin (0.3 μM), EDTA (1 mM), PMSF (0.1 mM), ε-aminocaproic acid (1 mM) and trypsin inhibitor (10 μM) were added (final concentrations). The mixture was centrifuged at 140000 × *g* for 45 min in a Beckman airfuge. The supernatant (100 μl) was desalted [3] and the water removed in the Speed Vac concentrator. SDS gel electrophoresis and electroblotting onto a nitrocellulose membrane were performed as already published [4]. Immunodetection followed the instructions of the manufacturer (Auroprobe BLplus, Amersham), the anti-FNR antibody was used in a dilution of 1: 500.

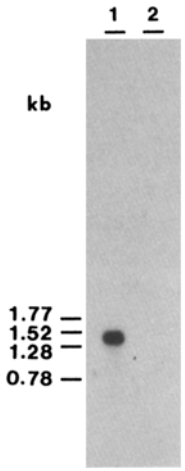

Fig. 1. Northern blot analysis. 5 μg each of *C. paradoxa* poly(A)$^+$ RNA (lane 1) and of cyanelle RNA (lane 2) were fractionated by electrophoresis under denaturing conditions. The RNA was blotted onto a Hybond-N membrane and hybridized with a ^{32}P-labelled 870 bp internal *Sst* II fragment of FNR cDNA.

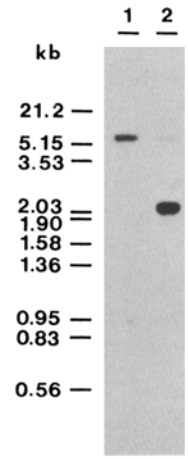

Fig. 2. Southern blot analysis. *C. paradoxa* genomic DNA (10 μg per lane) was digested with *Bgl* II (lane 1) and *Hin*d III + *Bgl* II (lane 2) and separated on a 0.7% agarose gel. The DNA was transferred to a Hybond-N membrane and probed with the ^{32}P-labelled 870 bp internal *Sst* II fragment of FNR cDNA. Lambda *Hin*d III + *Eco* RI DNA fragments were used as molecular size markers.

Log-phase cultures of *Escherichia coli* XL1-Blue containing the plasmid pF10 in LB-amp medium and untransformed control cultures were induced by the addition of IPTG to a final concentration of 1 mM and incubated at 37 °C for another three hours. The pellet obtained after centrifugation at 14000 × *g* for 30 s was solubilized in Laemmli sample buffer, boiled for 5 min, cooled to room temperature and recentrifuged at 14000 × *g* for 10 min. Immunoblotting was performed as described. Western blotting of cyanelle stroma proteins followed published methods [3].

Results

Screening of a *C. paradoxa* cDNA library in λ ZAP (ca. 50000 plaques) with antibodies directed against cyanelle FNR yielded 12 positive clones. Analysis of phage DNA preparations showed that 9 clones contained identical inserts of 1.3 kb whereas the other clones contained truncated versions of this transcript. Northern analysis using a 870 bp *Sst* II fragment resulted in a signal at approximately 1.35 kb with poly(A)$^+$ RNA only (Fig. 1). The respective gene locus was confined to a 2.2 kb genomic DNA fragment by Southern hybridization using the same probe (Fig. 2). This leads us to assume that *petH* is a single-copy gene in *C. paradoxa*. Sequence analysis of the insert in plasmid pF10 obtained by excision from recombinant λZAP revealed the coding region for a polypeptide of 363 amino acids flanked by a 5′-untranslated region of 4 bp (plus 8 bp originating from the *Eco* RI linker) and a 3′-untranslated region of 186 bp (Fig. 3). Compared with the transcript size about 50 bp appear to be missing from the 5′-untranslated region. Western blotting of *E. coli* proteins after transformation with Bluescript SK$^-$ containing the FNR insert resulted in a band corresponding to the fusion protein (predicted molecular mass 44.5 kDa) and a more intense band at 34 kDa (Fig. 4A) indicating a processing activity in *E. coli*. *In vitro* translation products from *C. paradoxa* poly(A)$^+$ RNA yielded (Fig. 4B) a signal upon immunostaining as expected for the FNR precursor (40.5 kDa). No band was obtained with the wheat germ extract when poly(A)$^+$ RNA was omitted (not shown). Western blotting of cyanelle extracts showed the 34 kDa mature polypeptide (Fig. 4C). Sequence comparison of the cyanelle protein with FNRs from three higher plants [17, 25, 27] and three cyanobacteria [11, 32, 41] revealed a high degree of conservation (Fig. 5) particularly in domains crucial for enzyme activity which are known from the three-dimensional structure of the spinach FNR [19]. Identity scores were slightly higher when the mature cyanelle polypeptide was aligned to its higher-plant counterparts (Table 1). With the exception of the two N-terminal amino acids and the three C-terminal residues, there was almost no similarity between the transit peptides of the *C. paradoxa* and the higher plant enzymes which – among themselves – are quite well conserved (Table 1). Amino terminal protein sequencing of cyanelle FNR [13] resulted in Ala-Val-Asp-Ala-Lys-Lys-Lys. This would

Table 1. Amino acid identity (%) in FNR transit peptides[1] and mature proteins[2].

	C.p.	M.c.	S.o.	P.s.	Sp.s.	An.v.
Cyanophora paradoxa (C.p.)	–	(25)	(21)	(22)	–	–
Mesembryanthemum crystallinum (M.c.)	66/64	–	(71)	(62)	–	–
Spinacia oleracea (S.o.)	65/62	91/90	–	(52)	–	–
Pisum sativum (P.s.)	65/64	87/86	89/87	–	–	–
Spirulina sp. (Sp.s.)	62/64	58/55	56/52	57/55	–	–
Anabaena variabilis (An.v.)	58/59	53/51	52/48	55/52	70/70	–
Synechococcus sp. (Sy.s.)	59/57	53/52	53/52	55/53	72/69	64/61

[1] Identity scores for transit peptides are in parentheses (gaps excluded).
[2] Identity scores based on 321 amino acids (gaps excluded/gaps considered).

```
   AATTCCGTCGCA ATG GCC TTC GTC GCG TCT GTC CCC GTC TTC GCC    45
                 M   A   F   V   A   S   V   P   V   F   A
AAC GCC TCT GGC CTT AAG ACC GAG GCT AAG GTC TGC CAG AAG CCC    90
 N   A   S   G   L   K   T   E   A   K   V   C   Q   K   P
GCG CTG AAG AAC AGC TTC TTC CGC GGC GAG GAG GTT ACC TCT CGC   135
 A   L   K   N   S   F   F   R   G   E   E   V   T   S   R
TCG TTC TTC GCC AGC CAG GCT GTG TCG GCG AAG CCG GCG ACC ACC   180
 S   F   F   A   S   Q   A   V   S   A   K   P   A   T   T
TTC GAG GTC GAC ACC ACC ATC CGC GCG CAG GCG GTT GAC GCC AAG   225
 F   E   V   D   T   T   I   R   A   Q   A   V   D   A   K
AAG AAG GGC GAC ATC CCG CTC AAC CTC TTC CGC CCG GCC AAC CCT   270
 K   K   G   D   I   P   L   N   L   F   R   P   A   N   P
TAC ATT GGC AAG TGC ATC TAC AAC GAG CGC ATC GTC GGC GAG GGC   315
 Y   I   G   K   C   I   Y   N   E   R   I   V   G   E   G
GCT CCG GGT GAG ACC AAG CAC ATC ATC TTC ACC CAC GAG GGC AAG   360
 A   P   G   E   T   K   H   I   I   F   T   H   E   G   K
GTC CCG TAC CTT GAG GGC CAG TCG ATC GGC ATC ATC CCC CCG GGC   405
 V   P   Y   L   E   G   Q   S   I   G   I   I   P   P   G
ACC GAC AAG GAT GGC AAG CCC CAC AAG CTC CGC CTC TAC TCC ATT   450
 T   D   K   D   G   K   P   H   K   L   R   L   Y   S   I
GCC TCG ACC CGC CAC GGT GAC TTC GGC GAT GAC AAG ACC GTC TCG   495
 A   S   T   R   H   G   D   F   G   D   D   K   T   V   S
CTC TCG GTG AAG CGC CTC GAG TAC ACT GAT GCC AAT GGC AAC CTC   540
 L   S   V   K   R   L   E   Y   T   D   A   N   G   N   L
GTC AAG GGT GTC TGC TCG AAC TAC CTC TGC GAC CTG AAG CCC GGC   585
 V   K   G   V   C   S   N   Y   L   C   D   L   K   P   G
GAT GAG GTT ATG ATC ACC GGC CCG GTC GGC ACC ACC ATG CTC ATG   630
 D   E   V   M   I   T   G   P   V   G   T   T   M   L   M
CCC GAG GAC CAG AGC GCG ACC ATT ATC ATG CTC GCC ACC GGC ACT   675
 P   E   D   Q   S   A   T   I   I   M   L   A   T   G   T
GGC ATT GCG CCC TTC CGC TCT TTC CTC CGC CGC ATG TTC GAG GAG   720
 G   I   A   P   F   R   S   F   L   R   R   M   F   E   E
ACC CAC GCC GAC TAC AAG TTC AAC GGC CTC GCG TGG CTC TTC CTT   765
 T   H   A   D   Y   K   F   N   G   L   A   W   L   F   L
GGT GTC CCC ACC TCC TCC ACC CTC CTC TAC CGC GAG GAG CTC GAG   810
 G   V   P   T   S   S   T   L   L   Y   R   E   E   L   E
AAG ATG CAG AAG GCC AAC CCC AAC AAC TTC CGC CTC GAC TAC GCG   855
 K   M   Q   K   A   N   P   N   N   F   R   L   D   Y   A
ATC TCC CGC GAG CAG ACC GAC TCG AAG GGC GAG AAG ATG TAC ATC   900
 I   S   R   E   Q   T   D   S   K   G   E   K   M   Y   I
CAG AAC CGC ATT GCG GAG TAC GCC AAC GAG TTC TGG AAC ATG ATC   945
 Q   N   R   I   A   E   Y   A   N   E   F   W   N   M   I
CAG AAG CCG AAC ACC TTC GTC TAC ATG TGC GGT CTC CGC GGC ATG   990
 Q   K   P   N   T   F   V   Y   M   C   G   L   R   G   M
GAG GAC GGC ATC CAG CAG TGC ATG GAG GAC ATC GCC AAG GCG AAC  1035
 E   D   G   I   Q   Q   C   M   E   D   I   A   K   A   N
GGC ACC ACC TGG GAC GCC GTC GTC AAG GGC CTG AAG AAG GAG AAG  1080
 G   T   T   W   D   A   V   V   K   G   L   K   K   E   K
CGC TGG CAC GTT GAG ACC TAC TAA GCGGGGAATTAGCCCGTTTAGTATGCA  1131
 R   W   H   V   E   T   Y
CAGCTTCGACGCAGGCCGCTCCGCGGAGCTAGGCTGCGCTCTTAGGCGCGGGCTACGAA  1190
CCGGAGTTCAGTTTTCGCCCTTGCTCTCCTGTAGGGGCCCCCATGTACCATTGCATAAT  1249
TTGGCTTTTTCAAATGCAGGAAGGTGATTGTTCGCAAAACGG                  1291
```

Fig. 3. DNA sequence and deduced amino acid sequence of *C. paradoxa* FNR cDNA. An arrow indicates the proposed site of precursor processing . Residues obtained from amino terminal peptide sequencing of the mature protein are in bold face.

point to a transit peptide of 66 amino acids. From the known protease sensitivity of spinach FNR, where the majority of N-terminal peptides starts with the amino acid isoleucine [20, 33], and the processing site consensus (Fig. 5) we consider it possible that the amino terminal amino acid of the mature polypeptide is glutamine as it was observed for the spinach and pea enzymes. Thus, a transit sequence of 65 amino acids is most likely.

Despite the low degree of sequence conservation between the transit peptides of *Cyanophora* and higher-plant (Table 1) hydropathy plots

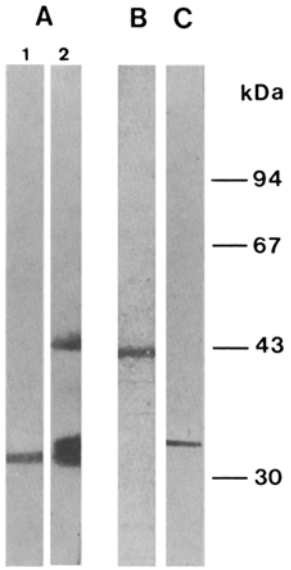

Fig. 4. A. Expression of a *C. paradoxa* FNR precursor fusion protein in *E. coli*. Protein extracts of *E. coli* transformed with pF10 after induction with IPTG (lane 2) and untransformed control cells (lane 1) were separated with SDS-PAGE and transferred to a nitrocellulose membrane and stained with the anti-FNR antibody. B. Immunodetection of the FNR precursor. *In vitro* translation products from *C. paradoxa* poly(A)$^+$ RNA were resolved on SDS-PAGE, electroblotted onto a nitrocellulose membrane and probed with the anti-FNR antibody. C. Western blot of soluble cyanelle proteins showing the 34 kDa mature FNR polypeptide. Markers in kDa are indicated.

(Fig. 6) clearly depict characteristic domains of transit peptides in general [38]. The prominent N-terminal hydrophobic region of 14–16 residues which lacks charged amino acids (Fig. 6, stippled) is followed by a more hydrophilic stretch. Fourier transform analyses [23] indicated an amphipathic beta sheet structure at or close to the putative cleavage sites (data not shown). Significant differences in mean frequencies of individual amino acids exist for alanine, serine, threonine and valine which occur more frequently in transit peptides than in mature FNRs, and for glutamic acid, aspartic acid and glycine occurring less frequently (not shown).

Discussion

The first nuclear gene for a cyanelle-located polypeptide, *petH*, represents also the first FNR-coding gene from an eukaryotic microorganism. The gene and protein sequences of FNR are particularly suited for tracing the ancestry of cyanelles because a number of both cyanobacterial and higher plant plastid FNR sequences are available. Based on other comparisons between cyanelle and plastid DNA encoded genes a bridge position has been ascribed to cyanelles that places these organelles between cyanobacteria and plastids. Arguments for such a connecting bridge include analysis of both amino acid and nucleotide sequence comparisons. For example, the deduced amino acid sequence of the *psbA* product, D1, indicates the presence of a carboxy terminal segment of the protein comprising seven amino acids, that is specific for cyanobacterial D1 proteins and missing in plastid D1 proteins from higher plants and green algae. However, the nucleotide sequence and overall deduced amino acid sequence of cyanelle *psbA* indicates closer similarity with higher plant counterparts [18]. Similar features are characteristic of cyanelle *rbcL*, encoding Rubisco LSU [24]. In stark contrast, the sequences of *petFI* (ferredoxin) [27], *tufA* (elongation factor Tu) [21], *rps12* (ribosomal protein S12) [21] and *rbcS* (Rubisco SSU) [36] exhibit a higher sequence similarity and more pronounced topological relatedness to the cyanobacterial counterparts [24]. Within the FNR tree (Fig. 7), the nucleus-encoded cyanelle FNR protein occupies a position that is also intermediary between higher plants and cyanobacteria, but the cyanelle FNR is placed slightly closer to the prokaryotic sequences [29]. The length of the *Cyanophora* branch corrobates earlier notions indicating that the lineage extending towards cyanelles evolved separately and into a different direction compared with the lineage that gave rise to chloroplasts [18]. The first evolutionary tree based on a *Cyanophora* nuclear gene sequence of a cyanelle protein – previously such comparisons were exclusively based on cyanelle DNA sequences – results in an evolutionary assignment that is very similar to previous assignments.

Biochemical investigations and the elucidation of the three-dimensional X-ray structure of FNR have mainly focused on the spinach enzyme. All

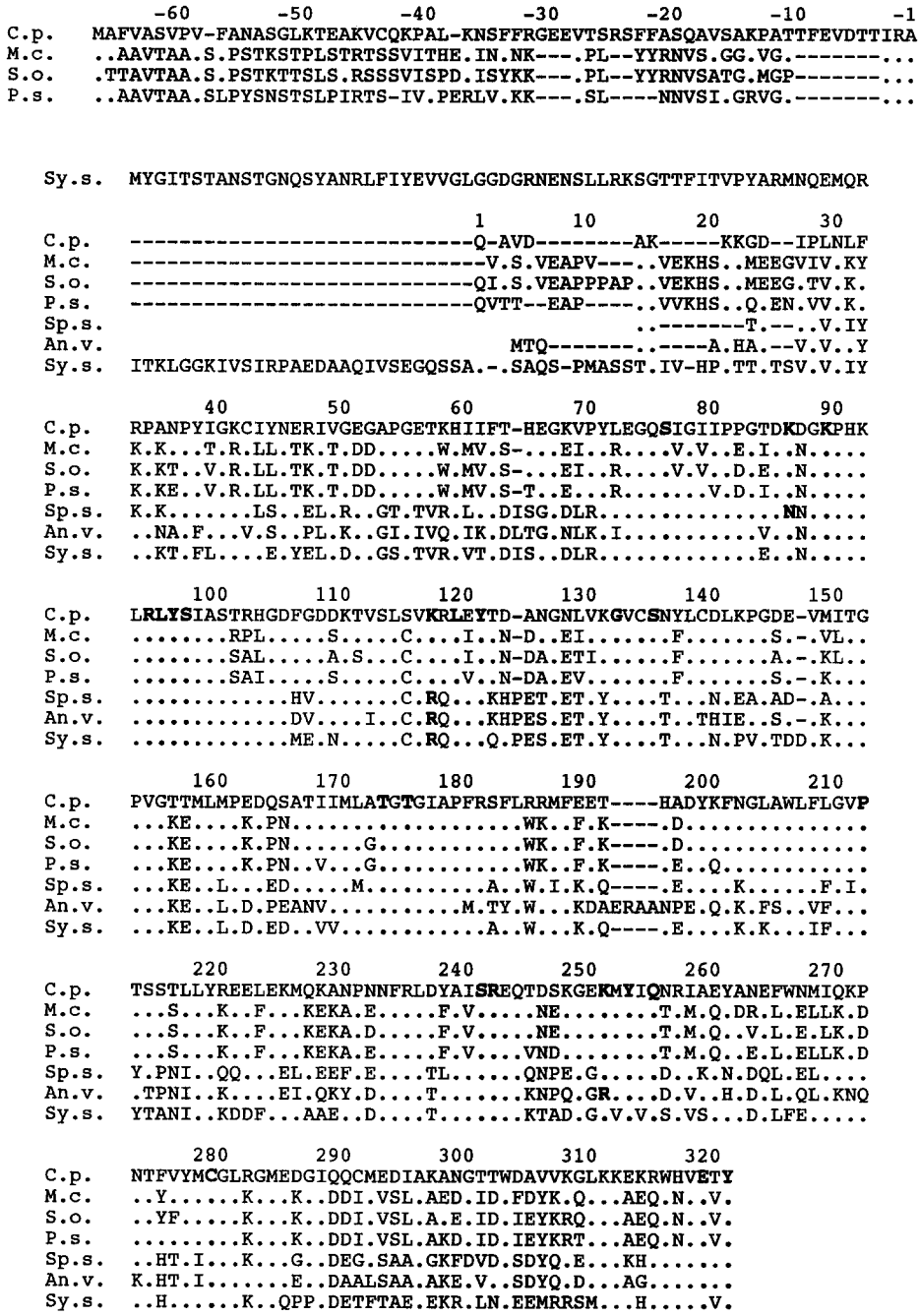

Fig. 5. Alignment of deduced cyanelle transit peptide and mature FNR sequences with those from higher plant and cyanobacterial species: *Cyanophora paradoxa* (C.p.), *Mesembryanthemum crystallinum* (M.c.) [25], *Spinacia oleracea* (S.o.) [17], *Pisum sativum* (P.s.) [27], *Spirulina* sp. (Sp.s.) [41], *Anabaena variabilis* (An.v.) [11], *Synechococcus* sp. (Sy.s.) [32]. Periods indicate identical amino acids, dashes indicate gaps inserted for optimal alignment. Sequences denoted with negative numbers represent the respective transit peptides. Residues of the mature proteins in bold face are involved in ligand binding and are referred to in the discussion.

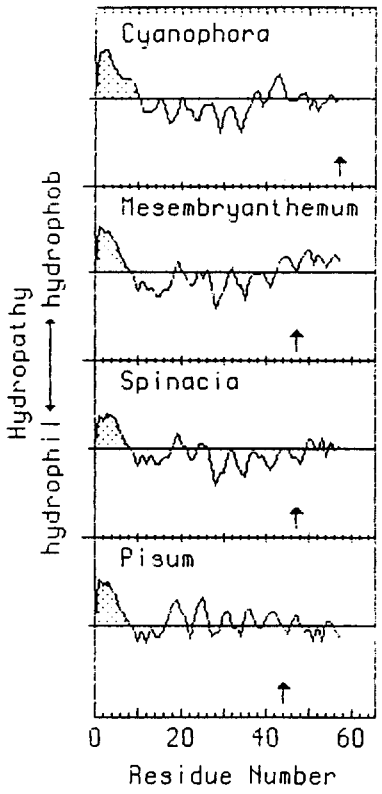

Fig. 6. Hydropathy profiles for FNR transit sequences according to the Kyte-Doolittle scale using a window size of 9 residues. The prominent hydrophobic N-terminal region (stippled) characteristic for higher-plant transit peptides [38] comprises 14–16 residues. Arrows denote the first residue in the last running window before the putative cleaving site.

amino acid residues of the spinach protein to which a specific function has been assigned [19] are conserved in the cyanelle protein. Likewise, residues which have been shown to be structurally important are generally invariant, in the cyanelle as well as in the other enzyme sequences. In the following compilation, the numbering system based on the spinach FNR structure is adopted: FAD-binding domain: serine-75, leucine-94, tyrosine-95, serine-96, glutamic acid-312 and tyrosine-314 are involved in binding of the flavin moiety; leucine-118 and tyrosine-120 play a role in binding of adenine and ribose; arginine-93, glycine-130 and serine-133 are important for binding of the pyrophosphate group.

NADP-binding domain: cysteine-272 and threonine-172 are responsible for binding of the nicotinamide moiety; through X-ray studies using 2′-phospho-AMP it became apparent that proline-205, threonine-170 and glutamine-248 interact with ribose and adenine, respectively, whereas serine-234, arginine-235, and threonine-246 are crucial for hydrogen bonds to the 2′-phosphate enabling the enzyme to discriminate against NAD^+. Two further basic residues considered important, lysine-116 and lysine-244, are in some cases replaced by arginine: the former in all cyanobacterial proteins, the latter in that from *A. variabilis*.

Ferredoxin binding is thought to occur in a pocket between the two domains mentioned above, lysine-85 (replaced by asparagine with *S. subsalsa*) and/or lysine-88 being involved in ion pair formation with acidic residues from ferredoxin.

Based on the results of Southern hybridization (Fig. 2) we estimate that *C. paradoxa* FNR is the product of a single nuclear gene, in contrast to the situation in higher plants where *petH* genes constitute a small gene family. The hybridization conditions were such as to allow the detection of conserved genes with the aid of heterologous probes [18]. Even after prolonged exposure no additional weak signals came up. Due to the stringent structure-function correlation in FNR the existence of additional genes with more divergent sequences is rather improbable. Isolation of genomic counterparts of FNR cDNA will have to be done to investigate the genomic organization of *petH*. With a genome size of approximately 150000 kb, which was estimated by staining DNA in nuclei and quantitation of the signal by flow cytometry (H.J. Bohnert, unpublished), the screening process of nuclear DNA libraries is approximately as complex as it is in *Arabidopsis thaliana*.

Expression of the FNR preprotein from the full-length cDNA clone in *E. coli* (Fig. 4A, lane 2) yielded two immunoreactive bands of molecular masses equivalent to 43 kDa and 34 kDa, respectively. The band of higher molecular mass is expected for the entire preprotein (which is extended at its amino terminal end due to the presence of a short sequence resulting from the *lacZ* fusion).

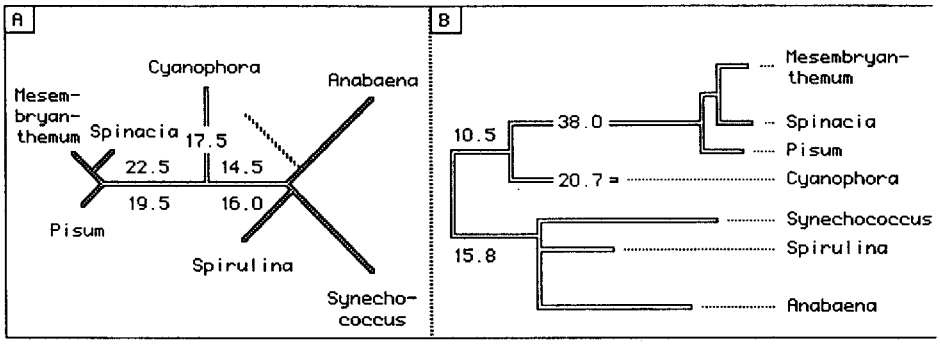

Fig. 7. Topologies of gene trees for mature FNRs inferred by two distance matrix methods with branch lengths approximately to scale. Numbers below/above (excluding/considering gaps) or on branches refer to fixed mutations per 100 residues (note that different branch lengths arise due to different modes in matrix calculations). A. Topology inferred by the neighbor joining method [28] excluding gaps. The broken branch denotes the position of *Spirulina* when gaps are considered as individual insertions/deletions. For either case the bridging position of *Cyanophora* remains unchanged in bootstrapped samples ($n = 100$). B. Tree inferred by the method of Fitch and Margoliash [12]; identical trees result by omitting or taking into account insertions/deletions.

The band at lower molecular mass is close to what is expected of a processed preprotein after removal of the transit peptide. This result is not surprising. Evidence has been provided for the processing of foreign FNR in *E. coli*. When the spinach pre-FNR protein was expressed in *E. coli* the bacterial La protease recognized a sequence close to the chloroplast processing site and cleaved the protein [8]. In this region of the preprotein, although the correct cleavage site is not known, cyanelle and spinach pre-FNR show considerable sequence conservation.

The sequence of FNR from *Synechococcus* sp. PCC 7002 which became available recently [32] contains a domain not present in other FNR sequences. This cyanobacterial FNR protein contains an amino terminal extension of approximately 90 amino acids which is remarkable as the additional sequence shows similarity to CpcD-type linker polypeptides that function in binding the terminal α- and β-phycocyanin trimers in the outer rod structures of phycobilisomes [7]. This finding gave rise to the hypothesis [32] of the extension being responsible for the attachment of this cyanobacterial FNR enzyme to phycobilisomes. Experimental results with *Synechococcus* do indeed suggest that this FNR can be extracted from phycobilisome preparations. The cyanelle protein which lacks the amino terminal extension and thus may not bind to phycobilisomes that are present in the cyanelles behaves in this respect like the higher-plant plastid FNR.

There is no amino acid sequence homology between the cyanelle FNR transit peptide and chloroplast (FNR) transit peptides. Irrespective of that the cyanelle FNR transit peptide resembles other plastid stroma-targeting peptides (STP) in the relative hydrophobicity of the amino terminus and in the high frequency of hydroxylated amino acids. Hydroxylated amino acids comprise 20% of the FNR transit peptide as compared to 24 to 28% serine and threonine for plastid STP. In contrast to plastid sequences, the hydroxy amino acids are more frequent at the carboxy terminus of the peptide. In accordance to what is typically found with plastid targeting sequences, tryptophane and tyrosine residues are missing, and several basic amino acids are found. Commonly not found in plastid transit peptides, the cyanelle sequence shows five acidic residues, although the overall positive charge (+ 3) is conserved. There are several reports indicating that structure, probably a random coil, rather than sequence is important for organelle targeting of nucleus-coded proteins [37]. From this we can infer a significant overall structural similarity between the preprotein domains of cyanelle and chloroplast transit peptides. Similarity can, perhaps, also be expected for the protein import apparatus. The peptidoglycan sacculus does not represent a physical

barrier, but rather a network stabilizing cyanelle structure, with a mesh size sufficiently large to permit the passage of preproteins. In this scenario, very similar, evolutionarily conserved protein import mechanisms might exist.

Acknowledgements

This work was supported by grants from the Fonds zur Förderung der wissenschaftlichen Forschung (S 6008-BIO) to W.L., from the Deutsche Forschungsgemeinschaft (SPP Intrazelluläre Symbiose, Sche 98/10-5) to H.E.A.S. and from the Arizona Agricultural Experimental Station to H.J.B.

References

1. Aitken A, Stanier RY: Characterization of peptidoglycan from the cyanelles of *Cyanophora paradoxa*. J Gen Microbiol 212: 218–229 (1979).
2. Bayer MG: Detektion von kern- und cyanellenkodierten Proteinen bei *Cyanophora paradoxa* im Mikromaßstab sowie biochemische und molekularbiologische Charakterisierung des Ferredoxins. Thesis, University of Tübingen (1991).
3. Bayer MG, Gebhard UB, Maier TL, Schenk HEA: Two-step purification of *Cyanophora* ferredoxin and its identification in soluble protein preparations by isoelectric focusing. Protein Express Purif 2: 240–247 (1991).
4. Bayer MG, Maier TL, Gebhart UB, Schenk HEA: Cyanellar ferredoxin-NADP$^+$-oxidoreductase is encoded by the nuclear genome and synthesized on cytoplasmic 80S ribosomes. Curr Genet 17: 265–267 (1990).
5. Bayer MG, Schenk HEA: Biosynthesis of proteins in *Cyanophora paradoxa*. I. Protein import into the endocyanelle analyzed by micro two-dimensional gel electrophoresis. Endocyt Cell Res 3: 197–202 (1986).
6. Bohnert HJ, Löffelhardt W: Molecular genetics of cyanelles from *Cyanophora paradoxa* In: Reisser W (ed) Algal Symbioses, pp. 379–397. Biopress, Bristol (1992).
7. Bryant, DA: Genetic analysis of phycobilisome biosynthesis, assembly, structure, and function in the cyanobacterium *Synechococcus* sp. PCC 7002. In: Stevens SE, Bryant DA (eds) Light-Energy Transduction in Photosynthesis: Higher Plant and Bacterial Models, pp. 62–90. American Society of Plant Physiologists, Rockville (1988).
8. Ceccarelli EA, Viale AM, Krapp AR, Carrillo N: Expression, assembly, and processing of an active plant ferredoxin-NADP$^+$ oxidoreductase and its precursor protein in *Escherichia coli*. J Biol Chem 266: 14283–14287 (1991).
9. Ellis RJ, van der Vies SM: Molecular chaperones. Annu Rev Biochem 60: 321–347 (1991).
10. Felsenstein J: Phylogenies from molecular sequences: Inference and reliability. Annu Rev Genet 22: 521–565 (1988).
11. Fillat MF, Bakker HAC, Weisbeek P: Sequence of the ferredoxin-NADP$^+$ reductase from *Anabaena* PCC 7119. Nucl Acids Res 18: 7161 (1990).
12. Fitch WM, Margoliash E: Construction of phylogenetic trees. Science 155: 279–284 (1967).
13. Gebhart UB, Stefanovic S, Bayer MG, Maier TL, Schenk HEA: Ferredoxin-NADP$^+$-oxidoreductase of *Cyanophora paradoxa*: purification, partial characterization, N-terminal amino acid sequence. Prot Express Purif 3: 228–235 (1992).
14. Hartl F-U, Neupert W: Protein sorting to mitochondria: evolutionary conservation of folding and assembly. Science 247: 930–938 (1990).
15. Harlow E, Lane D (eds) Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1988).
16. Henikoff S: Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351–359 (1984).
17. Jansen T, Reiländer H, Stepphuhn J, Herrmann RG: Analysis of cDNA clones encoding the entire precursor-polypeptide for ferredoxin: NADP$^+$ oxidoreductase from spinach. Curr Genet 13: 517–522 (1988).
18. Janssen I, Jakowitsch J, Michalowski C, Bohnert HJ, Löffelhardt W: Evolutionary relationship of *psbA* genes from cyanobacteria, cyanelles and plastids. Curr Genet 15: 335–340 (1989).
19. Karplus PA, Daniels MJ, Herriott JR: Atomic structure of ferredoxin-NADP$^+$ reductase: prototype for a structurally novel flavoenzyme family. Science 251: 60–66 (1991).
20. Karplus PA, Walsh KA, Herriott JR: Amino acid sequence of spinach ferredoxin: NADP$^+$ oxidoreductase. Biochemistry 23: 6576–6583 (1984).
21. Kraus M, Götz M, Löffelhardt W: The cyanelle *str* operon from *Cyanophora paradoxa*: Sequence analysis and phylogenetic implications. Plant Mol Biol 15: 561–573 (1990).
22. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132 (1982).
23. Lüttke A: MacPROT: a set of basic programs for protein structure analysis. Comp Meth Progr Biomed 31: 105–112 (1990).
24. Lüttke A: On the origin of chloroplasts and rhodoplasts: Protein sequence comparison. Endocyt Cell Res 8: 75–82 (1991).

25. Michalowski CB, Schmitt JM, Bohnert HJ: Expression during salt stress and nucleotide sequence of cDNA for ferredoxin-NADP$^+$ reductase from *Mesembryanthemum crystallinum*. Plant Physiol 89: 817–822 (1989).
26. Neumann-Spallart C, Brandtner M, Kraus M, Jakowitsch J, Bayer MG, Maier TL, Schenk HEA, Löffelhardt W: The *petFI* gene encoding ferredoxin I is located close to the *str* operon on the cyanelle genome of *Cyanophora paradoxa*. FEBS Lett 268: 55–58 (1990).
27. Newman BJ, Gray JC: Characterisation of a full-length cDNA clone for pea ferredoxin-NADP$^+$ reductase. Plant Mol Biol 10: 511–520 (1988).
28. Saitou N, Nei M: The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425 (1987).
29. Schenk HEA, Bayer MG, Maier TL, Lüttke A, Gebhart UB, Stevanovic S: Ferredoxin-NADP$^+$-oxidoreductase of *Cyanophora paradoxa*, nucleus-encoded but cyanobacterial. Gene transfer from symbiont to host, an evolutionary mechanism originating new species. Z Naturforsch 47c: 347–358 (1992).
30. Schenk HEA: Cyanobacterial symbioses. In: Ballows A, Trüper HG, Dworkin M, Harder W, Schleifer KH (eds) The Prokaryotes, vol. 4 pp. 3819–3854. Springer, New York (1992).
31. Schenk HEA: Nachweis einer lysozymempfindlichen Stütsmembran der Endocyanellen von *Cyanophora paradoxa* Korsch. Z Naturforsch 25b: 656 (1970).
32. Schluchter WM, Bryant DA: Molecular characterization of ferredoxin-NADP$^+$ oxidoreductase in cyanobacteria: cloning and sequence of the *petH* gene of *Synechococcus* sp. PCC 7002 and studies on the gene product. Biochemistry 31: 3092–3102 (1992).
33. Shin M, Tsujita M, Tomizawa H, Sakihama N, Kamei K, Oshino R: Proteolytic degradation of ferredoxin-NADP$^+$ reductase during purification from spinach. Arch Biochem Biophys 279: 97–103 (1990).
34. Short JM, Fernandez JM, Sorge JA, Huse WD: Lambda ZAP: a bacteriophage Lambda expression vector with *in vivo* excision properties. Nucl Acids Res 16: 7583–7600 (1988).
35. Smeekens S, Weisbeek P, Robinson C: Protein transport into and within chloroplasts. Trends Biochem Sci 15: 73–76 (1990).
36. Starnes SM, Lambert DH, Maxwell ES, Stevens SE, Porter RD, Shively JM: Cotranscription of the large and small subunit genes of ribulose-1.5-bisphosphate carboxylase/oxygenase in *Cyanophora paradoxa*. FEMS Microbiol Lett 28: 165–169 (1985).
37. von Heijne G: Chloroplast transit peptides: the perfect random coil? FEBS Lett 278: 1–3 (1991).
38. von Heijne G, Steppuhn J, Herrmann RG: Domain structure of mitochondrial and chloroplast targeting peptides. Eur J Biochem 180: 535–545 (1989).
39. Wang Y: Double-stranded DNA sequencing with T7 polymerase. Biotechniques 6: 843–845 (1988).
40. Wickner W, Driessen AJM, Hartl F-U: The enzymology of protein translocation across the *Escherichia coli* plasma membrane. Annu Rev Biochem 60: 101–124 (1991).
41. Yao Y, Tamura T, Wada K, Matsubara H, Kodo K: *Spirulina* ferredoxin-NADP$^+$ reductase. The complete amino acid sequence. J Biochem 95: 1513–1516 (1984).