1123

Université de Neuchâtel

Faculté des Sciences

Structural and functional analysis of nuclear genes coding for the elongation factor eEF-1α (Glycine max) and the chloroplast specific thioredoxin *f* (Spinacia oleracea)

Cette-ci est une forme réduite de la thèse pour l'obtention du grade de docteur ès sciences de l'Université de Neuchâtel

Thèse de doctorat

Fernando Aguilar Morales, 1990

IMPRIMATUR POUR LA THÈSE

Structural and functional analysis of nuclear genes coding for the elongation factor, eEF-1α (Glycine max) and the chloroplast specific thioredoxin f (Spinacia oleracea)

de M.onsieur Fernando Aguilar Morales

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES

La Faculté des sciences de l'Université de Neuchâtel sur le rapport des membres du jury,

Messieurs E. Stutz, P. Schürmann et B. Lescure (Toulouse)

autorise l'impression de la présente thèse.

Neuchâtel, le 18 novembre 1991

Le doyen :

A. Robert

360
Plant Molecular Biology 17: 351-361, 1991.

Two genes encoding the soybean translation elongation factor eEF-1α are transcribed in seedling leaves

Fernando Aguilar [1], Paul-Etienne Montandon [2] and Erhard Stutz*
Laboratoire de Biochimie Végétale, Université de Neuchatel, Chantemerle 18, CH-2000 Neuchatel, Switzerland (author for correspondence); present addresses: [1] ISREC, Chemin des Boveresses, CH-1066 Epalinges, Switzerland; [2] Service de l'Hygiène et de l'Environnement, Av. Léopold-Robert 36, CH-2300 La Chaux-de-Fonds, Switzerland*

Received 17 January 1991; accepted in revised form 23 April 1991

Key words: soybean seedlings, eEF-1α, *tef* genes, cDNA, genomic DNA, dark/light transcription, intron

Abstract

A cDNA and a genomic DNA library from soybean (*Glycine max* L.) were used to identify and sequence two genes coding for the α-subunit of the translation elongation factor eEF-1. Within the coding part, the two genes (*tefS1* and *tefS2*) diverge in 80 wobble positions thus yielding an identical protein composed of 447 amino acids. The soybean protein has about 95% similarity with eEF-1α proteins of *Arabidopsis thaliana* and tomato. Both genes *S1* and *S2* contain, within the coding part at a site seemingly unique to higher plants, a single short intron of 86 and 116 nucleotides, respectively. The untranslated leader part of both genes is interrupted by a large intron (partially sequenced). Genes *S1* and *S2* are transcribed in young leaves. cDNA and gene-specific oligonucleotide probes interact with a unique transcript of close to 1.9 kb. Northern hybridization studies using RNAs from dark- and light-grown seedlings show that light sharply increases the level of stable transcripts (1.9 kb). A peak value is measured after about 3 h of illumination, afterwards the transcript concentration drops to about 10% of the peak value. Genes *S1* and *S2* follow a similar transcription pattern in developing seedling leaves, which is distinct from that of the *rbcS* genes measured in parallel experiments. According to northern results, *S1* transcripts are more abundant in leaves at all measured stages of development than *S2* transcripts.

Introduction

The α-subunit of the complex translation elongation factor eEF-1 [12] catalyses the binding of aminoacyl-tRNAs to the translating ribosome in a GTP-requiring step. eEF-1α is an abundant protein and the corresponding genes (*tef*) have been cloned and sequenced in several species in-

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X56856.

cluding the photosynthetically active eukaryotes *Euglena gracilis* [9], *Arabidopsis thaliana* [1] and *Lycopersicon esculentum* [15]. It is generally accepted that eEF-1α activity is higher in young developing tissues (organs) than in older ones and the question arises how and at what level the expression of eEF-1α is controlled. eEF-1α is coded for by one or more genes per haploid genome and it has been reported that members of the gene family can be differentially expressed in function of developmental stages [3].

Higher plants are suitable to address regulatory questions and first results have been reported for tomato plants [11] where it has been shown that young leaves contain higher steady-state levels of eEF-1α mRNA than older tissues without, however, specifying whether one, some or all of the four to eight members of this gene family are actively transcribed. On the other hand, Liboz *et al.* [8] showed by cDNA sequencing that all four *tef* genes of *thaliana* are transcribed in cell suspension cultures.

We have cloned and sequenced two *tef* genes of soybean (*S1* and *S2*) and followed their transcription during light-induced leaf development in seedlings. It turns out that both genes are transcribed according to a similar transcription pattern but at somewhat different intensities. The transcription programme is clearly distinct from that of the *rbcS* genes which code for the small subunit protein of ribulose-1, 5-bisphosphate carboxylase. We also display and discuss some structural features of the two *tefS* genes.

Materials and methods

Growth of soybean plants and seedlings

Soybean (*Glycine max* cv. Maple Arrow) was grown in a phytotron as reported [18] for approximately 3 weeks under the following conditions: light 250 W m^{-2}, 16 h, 25 °C, r.h. 70%. Soybean leaves were hand-picked, immediately frozen in liquid N_2, and stored at − 70 °C until used. Soybean seedlings were grown under the same conditions for approximately 4 weeks in the dark. Seedling leaves were picked after 3, 16, 24 and 48 h of illumination (250 W/cm^2), frozen in liquid N_2 and kept in black bags at − 70 °C. Etiolated seedlings were grown in the dark under the same conditions and leaves were collected under a green safety light.

DNA isolation and Southern analysis

DNA was isolated from soybean leaves according to published protocols [17] with some modifications. Soybean DNA was spooled out by adding 1 vol of cold isopropanol. High-molecular-weight DNA was collected with an inverted Pasteur pipette and washed twice in ice-cold 75% ethanol. Genomic DNA was resuspended in 100 μl of TE buffer (10 mM Tris-Cl pH 8, 1 mM EDTA), treated with 50 μg/ml of RNase (DNase-free, Sigma) for 1 h at 37 °C and then with 100 μg/ml of Proteinase K for 2 h. The phenol/chloroform (1:1) extraction was repeated twice, the salt concentration was increased to 0.1 M NaCl and 2 vol of ethanol were added. DNA was recovered by spooling, washed as before, dissolved in 1–2 ml of TE buffer and stored at 4 °C.

Batches (10 μg) of DNA were restricted under conditions recommended by the supplier (Boehringer, Mannheim), electrophoresed on 0.8% agarose gels, transferred to nitrocellulose membranes (Schleicher & Schuell, BA 83) and the DNA fixed under UV light (Stratalinker 1800, Stratagene). Hybridizations of filter imprints with ^{32}P-DNA probes and 100 μg/ml denatured calf thymus DNA were done in 5 × SSPE (1 × SSPE is 0.18 M NaCl, 0.01 M sodium phosphate pH 7.4, 0.001 M EDTA) at 65 °C. Filters were twice washed in 0.2 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate), 0.1% SDS, for 20 min at 65 °C.

Screening of cDNA library

A soybean (*Glycine max* L. cv. Williams) cDNA library (λgt11, RNA from 6-day-old seedlings) purchased from Clontech Laboratories was

screened according to standard protocols [13] with a 1.2 kb cDNA internal fragment coding for *Euglena gracilis* eEF-1α [9]. Approximately 10^4 recombinant phages grown in *Escherichia coli* strain Y1090 were transferred to nitrocellulose filters. Hybridization conditions were: 5 × SSPE, 30% formamide, 2.5 × Denhardt's solution, 0.1% SDS, 100 μg/ml denatured calf thymus DNA at 42 °C. Filters were washed twice in 5 × SSC at 30 °C for 20 min. Two individual plaques were picked and plaque-purified. The inserts were *Eco* RI-digested and sized on 0.9% agarose gel. Both clones carried an approximately 2.0 kb insert as deduced from Southern hybridization results. One of these (gt11-EFα.1) was subcloned into M13mp8 and into the Bluescript II vector. This clone was used for sequencing and as specific DNA probe (pBSK + EF1α.1).

Screening of genomic library

Soybean DNA (200 μg) was partially restricted with *Sau*3A (1 and 2 U) and *Bam* HI (5 and 10 U) for 30 min and 1 h, respectively, under conditions specified by the supplier. Restricted DNA was pooled and sized on linear (5 to 25%) NaCl gradients. DNA (14–20 kb long) was ligated to λEMBL4 arms (1:1 ratio) overnight at 14 °C with 0.3 U/μl of T4 DNA ligase following the supplier's conditions (Boehringer, Mannheim). Packaging extracts and *in vitro* packaging were as described [5]. The recombinant phages were plated on *E. coli* strain K803, cloning efficiency was about 5×10^5 pfu/μg of soybean DNA. Recombinant phages were screened by plaque hybridization (for conditions see under Screening of cDNA library) except that 50% formamide was used. Filters were washed twice for 30 min in 1 × SSC, 0.1% SDS at 42 °C.

RNA isolation, northern analysis and tracing of autoradiographs

RNA was isolated according to the guanidinium isothiocyanate method [13] with some modifications. After the CsCl ultracentrifugation step, the high-molecular-weight RNA was precipitated by adding $\frac{1}{4}$ vol of 10 M LiCl, quickly frozen in dry ice and stored overnight at 4 °C. RNA was pelleted and dissolved in 500 μl TE and adjusted to 0.1 M NaCl, 1 mM $MnCl_2$, and 100 μg/ml of RNase-free DNase (Worthington). After 30 min at 0 °C, the reaction was stopped with 5 mM EDTA, 0.1% SDS and extracted once with phenol/chloroform (1:1). The RNA was precipitated with 2.5 vol of ethanol, dissolved in distilled H_2O (0.2% diethyl pyrocarbonate) and stored at −20 °C.

Denaturation and sizing of total RNA (20 μg) were done according to Thomas [19]. Northern hybridizations followed standard protocols. Filters were washed twice for 10 min in 2 × SSC, 0.1% SDS at room temperature, twice for 10 min in 0.1 × SSC, 0.1% SDS at room temperature, and once in 0.1 × SSC, 0.1% SDS at 50 °C for 1 h. Filters were exposed overnight at −70 °C with an intensifying screen.

Band intensities of autoradiographs of northern imprints were measured in a Zeiss optical scanning device at about 700 nm.

DNA sequencing

The inserts cloned into Bluescript II vector were sequenced using the *Exo* III/mung bean nuclease nested deletions method as described by the supplier (Stratagene) and DNA sequencing was done according to the deoxy/dideoxy method of Sanger *et al.* [14].

Melting profiles (T_m) of gene S1- *and* S2-*specific DNA hybrids*

Oligonucleotides (18-mer) specific, respectively, for gene *S1* (5′-GTCCCTAACAGCAAACCT, complementary to sequence positions 1267–1285, Fig. 1) and gene *S2* (5′-ATCCCTGACAGCAAAGCG) were synthesized in order to monitor the transcript level of the corresponding genes. The specificity of the probes was tested in

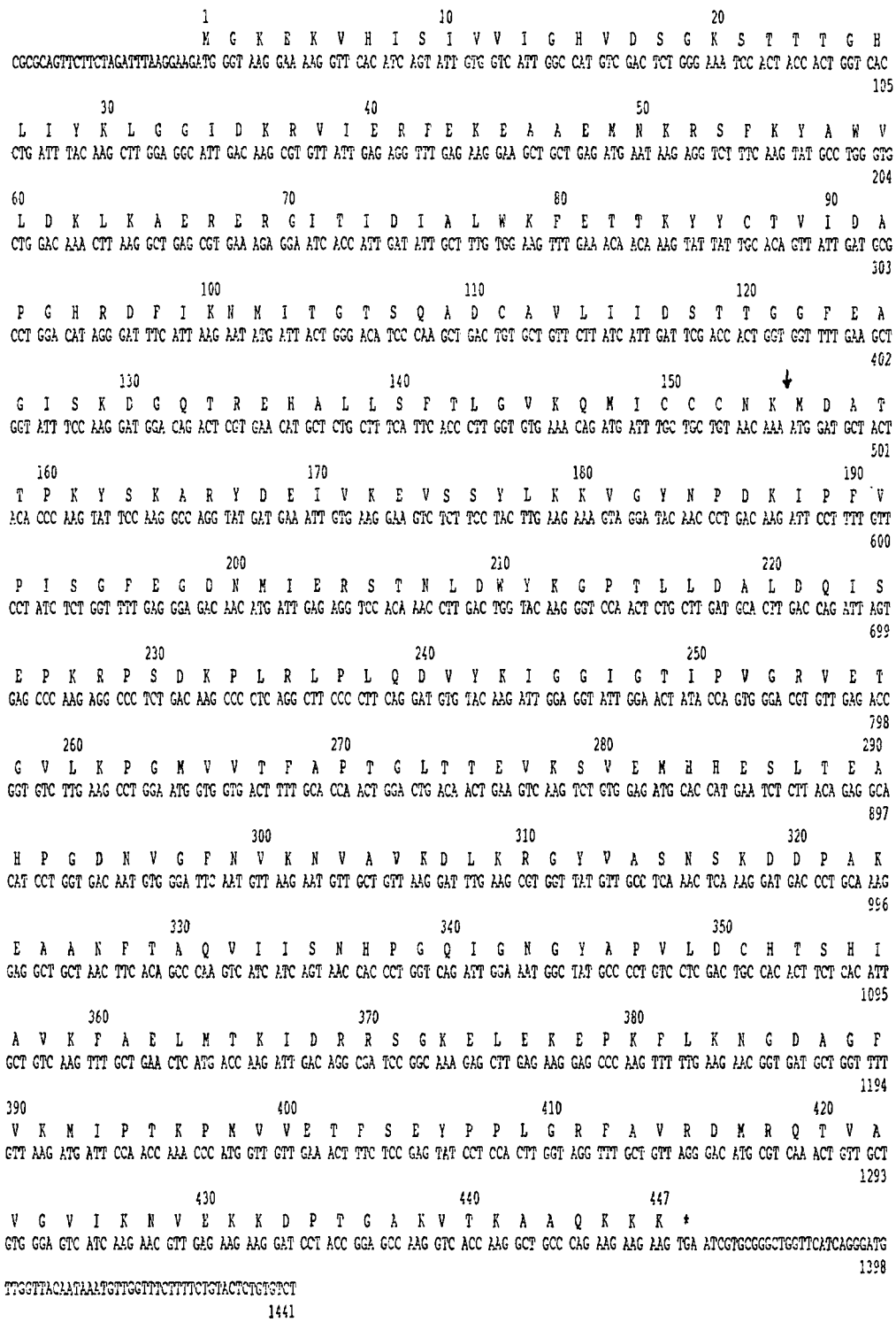

Fig. 1. Nucleotide sequence of eEF-1α cDNA (pBSK + EF1α.1) and decoded amino acid sequence (one-letter code). The arrow marks the position of the intron. Bold letters (5′ end) mark parts of the leader.

establishing the T_m of homologous and heterologous DNA hybrids essentially following published protocols [13] with some modifications. Samples (2.5 μg) of cloned *S1* and *S2* gene sequences and of Bluescript II vector DNA (control), respectively, were adjusted to 50 μl with buffer (20 mM Tris-Cl pH 8, 1 mM EDTA) and kept for 10 min in a boiling water bath, denatured with 50 μl of 1 M NaOH at room temperature for 20 min and neutralized with 400 μl of buffer (1.5 M NaCl, 1 M Tris-Cl pH 8). The samples were immediately loaded onto nitrocellulose filters using a dot blot apparatus. After washing each well with 500 μl of 20 × SSC buffer, DNA was fixed with UV light. Prehybridization and hybridization were done at 40 °C for 2 h and overnight, respectively, in 6 × SSPE, 0.1% SDS, 5 × Denhardt's solution and 100 μg/ml denatured calf thymus DNA. The filters were first washed in 2 × SSC at room temperature, at 40 °C and subsequently in temperature increments of 5 °C. Remaining ^{32}P label was monitored by Cerenkov counting (Kontron, BETAmatic). The respective melting curves (not shown) revealed that the oligonucleotides are suitable probes to discriminate between *S1* and *S2* transcripts under the hybridization and filter washing conditions used in the northern experiments (see above).

Results

Nucleotide sequence of tefS1 *cDNA and decoded amino acid sequence of the eEF-1α protein*

A soybean cDNA library was screened using an *Euglena gracilis tef* gene DNA fragment. In Fig. 1 the sequence of a cDNA clone is presented including the coding part, 27 nucleotides of the untranslated leader and parts of the downstream region which precedes the poly(A) tail. We arbitrarily call this gene *S1*. The coding part corresponds to the protein eEF-1α, being composed of 447 amino acid residues and matching higher-plant counterparts for more than 95% (comparison not shown). We subsequently used a homologous DNA probe to screen a soybean genomic library and to monitor Southern imprints of restricted total soybean DNA as shown in Fig. 2. We noticed in all lanes obtained with several restriction enzymes several bands two of which seem to be more pronounced than the others. Upon double digestion some of the higher and fainter bands disappear (not shown). The soybean genome contains at least two loci which strongly interact with the *tef* cDNA probe but additional *tef* gene loci and/or *tef* gene-related sequences of unknown structure exist. A restriction site map for the *S1* and *S2* gene locus is given in Fig. 3. Preliminary additional data (A. Spiel-

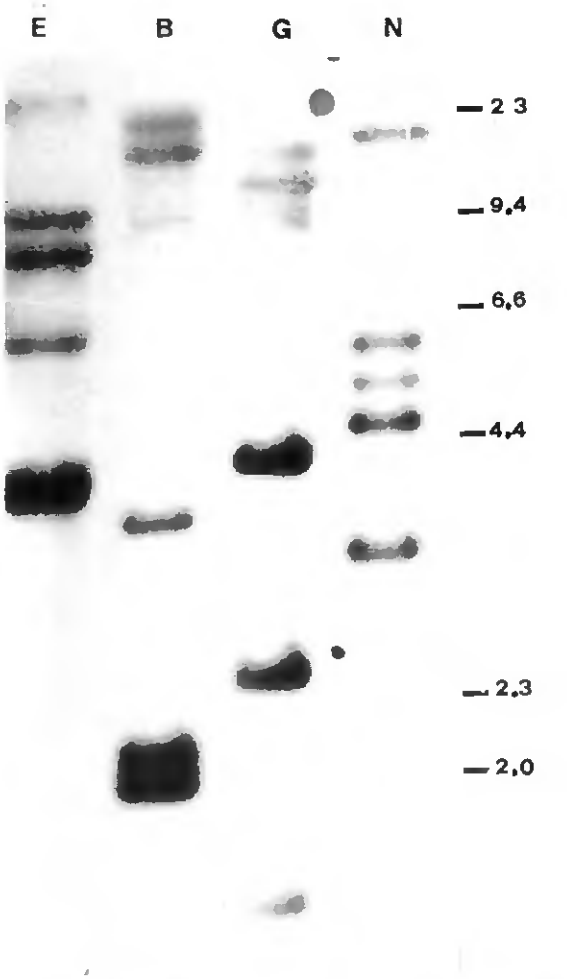

Fig. 2. Autoradiograph of Southern imprints from restricted total DNA. E, *Eco* RI; B, *Bam* HI; G, *Bgl* II; N, *Nco* I. Size marker in kb.

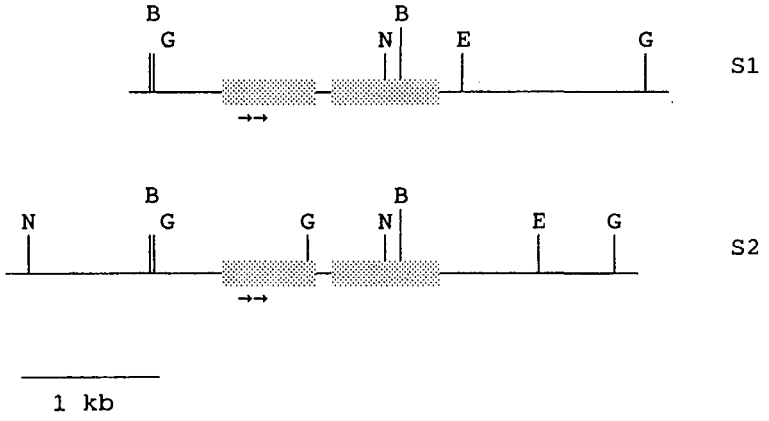

Fig. 3. Tentative restriction sites map of gene loci *S1* and *S2*. Interrupted stippled bar: transcribed region with small intron (line). →→ translation polarity. R-enzyme symbols as in Fig. 2.

mann, unpublished) suggest that the *S1* locus is duplicated.

We isolated and sequenced two genomic clones represented by the two *Bam* HI fragments of about 1.9 kb (Fig. 2, lane B). One of these clones had an insert with the identical sequence shown in Fig. 1 (coding part) which we call, by analogy, gene *S1* (*tefS1*). The insert of the other clone (gene *S2*) was also sequenced. It diverges within the coding part in 80 wobble positions (not shown) thus yielding an identical decoded eEF-1α protein.

Length of stable transcript, untranslated leader and introns

Sequencing of the two genomic clones revealed that both genes *S1* and *S2* carry a single short intron after codon 154 (see Fig. 1). The sequence of the two introns is given in Fig. 4. For *S1* and *S2* the respective intron length is 86 and 116 nucleotides with about 70% similarity. The short introns are AT-rich and carry structural features considered typical of plant introns, such as the increased content of pyrimidines in the 3′ terminal part which seems to be a requirement for

```
   939
AAA gtaaattatg tctcgtgttt taaggcccct ttcattttat ta........  S1
    ||||||||||  | | |  ||| |||||||| |  |   |   | ||
AAG gtaaattatg tgttgaattt taaggccctt atgtctcctt tacatcacac  S2

    .......... .......... ..gattattt tattttaatt aatgttcttg  S1
                            | | || |  |  |||||| | |||| |
    caatatcttc attccttttg tggttcatat gagattaatt agtgttgtct  S2

                   1024
    tcccattttg caacag ATG                                  S1
    ||  |||||||  |||||
    tctcattttg aaacag ATG                                  S2
```

Fig. 4. Nucleotide sequence of introns (lower-case letters) in coding part of *tef* genes *S1* and *S2*, respectively ... gap in *S1*.

```
TTTTTAGTTT CTTTGCAGAT TTAAGGAAGATG  tefS2

TTTTCTGGAT TTTTGCAGAT TTAAGGAAGATG  tefS1
    |    | | ||  |||| |||||||||
  CGCGCAGT TCTTCTAGAT TTAAGGAAGATG  cDNA
```

Fig. 5. Comparison of the 5′ terminal part of cDNA sequence with the genomic sequences of *tef S1* and *S2*. A possible acceptor splice site is underlined.

correct splicing [4]. *Arabidopsis* [8] and tomato [15] *tef* genes have a similar intron at exactly the same position, unlike *Euglena gracilis* [9] *tef* genes which are not split within the coding part.

The genomic clones *S1* and *S2* have been sequenced so far up to positions −445 and −479, respectively. In Fig. 5 we compare the 5′ terminal part of the cDNA sequence with the two genomic sequences. We noticed sequence identity for the nearby leader part (13 nt) and considerable sequence divergence for the rest. We did not find the remaining 13 nucleotides within the sequenced upstream region which indicates that the leader part is split as reported for the *tef* genes of *Arabidopsis* and tomato.

Transcription of tefS1 *and* tefS2 *in greening seedlings*

The cDNA sequence given in Fig. 1 proves that gene *S1* is actively transcribed in greening seedlings. We used *S1*- and *S2*-specific DNA probes (see Materials and methods) to monitor the transcript level of both genes. Figure 6 shows northern experiments with RNA from seedlings kept in the dark or exposed to light for various periods of time. We included hybridizations with rDNA probes in order to monitor the amount of loaded RNA (not shown). The results can be resumed as follows.

1. A single stable transcript of close to 1.9 kb interacts with the 1.2 kb internal cDNA probe (*S1*) independent of developmental stage.
2. Exposure to light for three hours (we did not measure shorter intervals) increases the steady-state mRNA level by a factor of about 50 as estimated from optical density scanning profiles of autoradiographs shown in the first and second lane (Fig. 6A).
3. In more developed leaves (16 to 48 light) the mRNA concentration drops to a definitely lower level in the range of about 10% of the peak level.

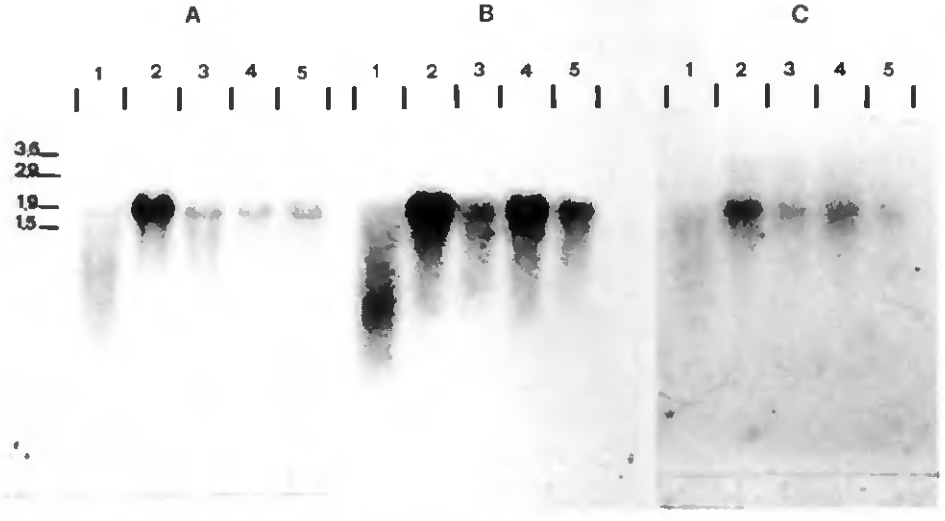

Fig. 6. Autoradiograph of northern experiments. RNA samples are from seedlings grown in the dark (1), or in light for 3 h (2), 16 h (3), 24 h (4) or 48 h (5). A. cDNA probe of eEF-1α coding part. B. *S1*-specific oligonucleotide probe. C. *S2*-specific oligonucleotide probe. Size marker in kb.

4. Essentially the same transcription pattern is seen using *S1*-specific (panel B) and *S2*-specific (panel C) DNA probes. We notice, however, in lane 4 of panel B and C a relative increase in band intensities not seen in panel A. We cannot explain yet this discrepeney other than in technical terms. Under the prevailing analytical conditions the *S2*-specific probe routinely gave a fainter signal than the *S1*-specific probe suggesting that the two genes as differentially transcribed and/or that *S1* and *S2* transcripts have a different turn-over.

The transcription patterns of soybean *tef* genes *S1* and *S2* are different from those reported for soybean *rbcS* genes [16]. We analysed *rbcS* transcription levels in parallel experiments (Fig. 7A) and found that signal intensities increase with increasing length of illumination up to a high plateau. Transcription control of *rbcS* genes was extensively studied [16] and shown to be very complex. On the one hand, light stimulates the transcription rate; on the other hand, it was argued that light also favours degradation of the transcript accelerating the turn-over. Our data only show that soybean *tef* genes and *rbcS* genes follow a different transcription pattern under strictly comparable conditions in developing seedlings. A first hint towards understanding some of the regulatory steps may be obtained from the result given in panel B. We overexposed an autoradiograph obtained with the *S1* DNA probe and we definitely see that the 'dark' RNA sample has a much higher ratio of degraded RNA interacting with the probe than the 'light' RNA samples suggesting that light increases the relative stability of the *tef* transcripts.

Discussion

Genome loci and structural aspects

Our sequencing data show that soybean nuclear DNA contains at least two genes coding for the α-subunit of EF-1 and the Southern experiments indicate that additional sequences exist which interact with the *S1* cDNA probe. Whether the fainter bands carry functional *tef* genes or *tef*-related sequences such as pseudo-genes or other functional genes with *tef*-related domains remains to be shown.

Both genes, *S1* and *S2*, are interrupted by a single short intron within the coding part separating codons 154 from 155. This is one of three domains known to interact with GTP [2]. The *tef* genes of *Arabidopsis thaliana* [8] and tomato [15] are interrupted at exactly the same position what seems to be typical for higher plant *tef* genes since in all other cases reported either introns do not exist [10] or introns are located at other positions [7, 20].

Our sequencing data also indicate that the un-

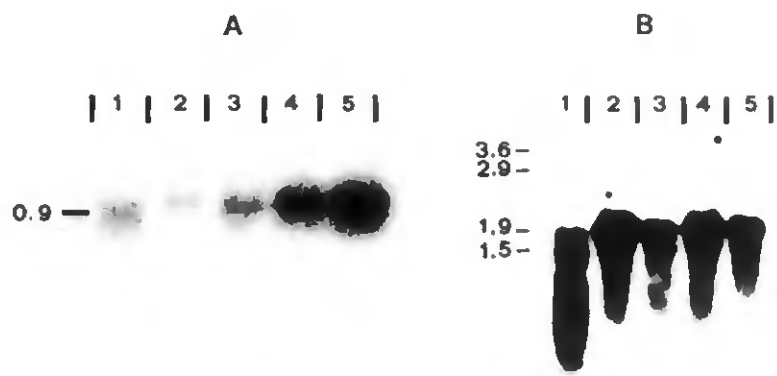

Fig. 7. Autoradiograph of northern experiments. RNA samples are as shown in Fig. 7. A. *rbcS*-specific DNA probe, 27-mer. B. *S1*-specific oligonucleotide probe (exposure time 5 days). Size markers in kb.

translated leader of genes *S1* and *S2* are interrupted by a large intron as was reported for *Arabidopsis* and tomato.

Transcriptional aspects

It was shown in many instances that eEF-1α activity positively correlates with the level of ongoing protein synthesis and it was shown that young tomato leaves or fruits contain higher levels of eEF-1α mRNA than older tissues [11]. According to our data *tef* mRNA concentration drastically increases in a specific way in developing seedling leaves. Though we have not measured eEF-1α activity *per se* we may reasonably assume that the increased *tef* mRNA concentration results in a proportionate increase in eEF-1α protein concentration in greening leaves. More surprising is the pronounced decrease in *tef* transcripts during the greening phase, i.e. after a peak around three hours after onset of light. The sharp rise and fall of *tef* mRNA is seen with both the *S1* and the *S2* gene probe, however, not with a nuclear rDNA probe (not shown) nor a nuclear DNA probe which codes for the chloroplast-specific EF-Tu protein (C. Bonny, unpublished results). The *tef* genes transcription pattern is also distinct from that of the *rbcS* genes as shown here under strictly comparable conditions. We therefore postulate that soybean genes *S1* and *S2* follow a specific transcription pattern in developing seedling leaves, which must be due to variation of transcription rates and/or transcript turnover. Furthermore, RNA samples from etiolated seedlings have a proportionate higher concentration of degraded *tef* DNA-interacting transcripts than samples from greening seedlings.

Another question is whether genes *S1* and *S2* are differentially expressed in developing leaves and possibly other organs of soybean. Our results suggest that gene *S2* mRNA concentration is inferior to that of gene *S1* in developing seedlings. It was shown in other cases that one or the other member of a *tef* gene family gets differentially expressed depending on the developmental stage [3].

Acknowledgements

We thank M. Keller, CNRS Strassbourg, for the *rbcS* oligonucleotide probe; T. Seebeck, Bern University, for providing the EMBL4 vector system; S. Marc-Martin and A. Spielmann for technical assistance. F.A. is recipient of a Bourse Fédérale pour Etudiants étrangers. This research is funded by Fonds Nationale Suisse de la Recherce scientifique to E.S. This work is part of a doctoral programme carried out by F.A.

References

1. Axelos M, Bardet C, Liboz T, Le Van Thai A, Curie C, Lescure B: The gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1α. Mol Gen Genet 219: 106–112 (1990).
2. Dever TE, Glynias MJ, Merrick WC: GTP-binding domain: Three consensus sequence elements with distinct spacing. Proc Natl Acad Sci USA 84: 1814–1818 (1987).
3. Djé MK, Mazabraud A, Viel A, le Maire M, Denis H, Crawford E, Brown D: Three genes under different developmental control encode elongation factor 1-α in *Xenopus laevis*. Nucl Acids Res 18: 3489–3493 (1990).
4. Goodall GJ, Filipowicz W: The minimum functional length of pre-mRNA introns in monocots and dicots. Plant Mol Biol 14: 727–733 (1990).
5. Hohn B: *In vitro* packaging of lambda and cosmid DNA. Meth Enzymol 68: 299–309 (1979).
6. Krämer A, Keller W, Appel B, Lührmann R: The 5′ terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. Cell 38: 299–307 (1984).
7. Lenstra JA, van Vliet AV, Arnberg AC, van Hemert FJ, Möller W: Genes coding for the elongation factor EF-1α in *Artemia*. Eur J Biol 155: 475–483 (1986).
8. Liboz T, Bardet C, Le Van Thai A, Axelos M, Lescure B: The four members of the gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1α are actively transcribed. Plant Mol Biol 14: 107–110 (1989).
9. Montandon P-E, Stutz E: Structure and expression of the *Euglena gracilis* nuclear gene coding for the translation elongation factor EF-1α. Nucl Acids Res 18: 75–82 (1990).
10. Nagashima K, Kasai M, Nagata S, Kaziro Y: Structure of two genes coding for polypeptide chain elongation factor 1α (EF-1α) from *Saccharomyces cerevisiae*. Gene 45: 265–273 (1986).
11. Pokalsky AR, Hiatt WR, Ridge N, Rasmussen R, Houck CM, Shewmaker CK: Structure and expression of elon-

gation factor 1α in tomato. Nucl Acids Res 17: 4661–4673 (1989).
12. Safer B: Nomenclature of initiation, elongation and termination factors for translation in eukaryotes. Recommendations 1988. Eur J Biochem 186: 1–3 (1989).
13. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
14. Sanger F, Coulson AR, Barell BG, Smith AJH, Roe BA: Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J Mol Biol 143: 161–178 (1980).
15. Shewmaker CK, Ridge NP, Pokalsky AR, Rose RE, Hiatt WR: Nucleotide sequence of an EF-1α genomic clone from tomato. Nucl Acids Res 18: 4276 (1990).
16. Shirley BW, Meagher RB: A potential role for RNA turnover in the light regulation of plant gene expression: ribulose-1,5-bisphosphate carboxylase small subunit in soybean. Nucl Acids Res 18: 3377–3385 (1990).
17. Shure M, Wessler S, Federoff N: Molecular identification and isolation of the *Waxy* locus in maize. Cell 35: 225–233 (1983).
18. Spielmann A, Ortiz W, Stutz E: The soybean chloroplast genome: Construction of a circular restriction site map and location of DNA regions encoding the genes for rRNAs, the large subunit of the ribulose-1,5-bisphosphate carboxylase and the 32 KD protein of the photosystem II reaction center. Mol Gen Genet 190: 5–12 (1983).
19. Thomas PS: Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Meth Enzymol 100: 255–266 (1983).
20. Uetsuki T, Naito A, Nagata S, Kaziro Y: Isolation and characterization of the Human chromosomal gene for polypeptide chain elongation factor-1α. J Biol Chem 264: 5791–5798 (1989).

Eur. J. Biochem. 182, 315–322 (1989)

Primary structure of spinach-chloroplast thioredoxin f

Protein sequencing and analysis of complete cDNA clones for spinach-chloroplast thioredoxin f

Masaharu KAMO[1], Akira TSUGITA[1], Christoph WIESSNER[2], Norbert WEDEL[2], Dieter BARTLING[2], Reinhold G. HERRMANN[2], Fernando AGUILAR[3], Laura GARDET-SALVI[3] and Peter SCHÜRMANN[3]

[1] Research Institute for Biosciences, Science University of Tokyo
[2] Botanisches Institut der Ludwig-Maximilians-Universität, München
[3] Laboratoire de Biochimie végétale, Université de Neuchâtel

(Received February 6, 1989) – EJB 89 0141

The primary structure of thioredoxin f from spinach chloroplasts was determined by standard amino acid sequencing and furthermore by sequencing the corresponding nuclear genome region. The protein, with a calculated molecular mass of 12564 Da and a molar absorption coefficient at 280 nm of 17700 M^{-1} cm^{-1}, consists of 113 residues and exhibits 24% residue identities with spinach chloroplast thioredoxin mb or *Escherichia coli* thioredoxin.

A monospecific antibody elicited against thioredoxin f has been used to select recombinant phage from spinach cDNA libraries in λgt11. The inserts of positive clones were sequenced. They code for a polypeptide of 190 amino acids, composed of the thioredoxin f sequence (113 residues) and an upstream element (77 residues) which most probably consitutes the N-terminal transit peptide that directs the polypeptide into chloroplasts. *In vitro* transcription and translation of this construct generates a polypeptide of approximately 21 kDa, which is imported by isolated spinach chloroplasts and processed to the mature 12.5-kDa protein.

Thioredoxins are ubiquitous, low-molecular-mass proteins that are characterized by the presence of an exposed, active center with the amino acid sequence: -Cys-Gly-Pro-Cys-. The two cysteine residues which are in close proximity can form an intramolecular disulfide bridge. Thioredoxins participate in numerous redox reactions during which the disulfide is reversibly reduced to a dithiol and they can serve as subunit for a virus DNA polymerase [1]. In photosynthetic cells thioredoxins play an important role in a light-dependent enzyme regulatory system [2–4].

Whereas most non-photosynthetic cells contain one type of thioredoxin which is closely related to the well studied and characterized thioredoxin from *Escherichia coli* [1], photosynthetic cells contain several types of thioredoxins which are distinguished by their specific functions. Thioredoxins capable of activating chloroplast fructose-1,6-bisphosphatase and certain other key enzymes of CO_2 assimilation are called f-type thioredoxins, those activating NADP-dependent malate dehydrogenase (the enzyme responsible for malate synthesis in chloroplasts) are called m-type thioredoxins [2, 3]. In higher plant cells, both of these thioredoxins are confined to the chloroplast and are reduced by light through the ferredoxin/thioredoxin system [4]. Recently the presence of a third type of thioredoxin, a cytoplasmic thioredoxin h (heterotroph) has been reported [5, 6] which is specifically reduced by an NADP/thioredoxin system.

In order to understand structure/function relationships in mechanisms involving thioredoxins it is necessary to study the structures of these proteins. At the protein level, complete primary structures of thioredoxins from different sources have been determined. The amino acid sequences of several bacterial thioredoxins [7–10] have been reported. Isolated from oxygenic photosynthetic cells, one cyanobacterial thioredoxin, from *Anabaena* Sp.7119, has been sequenced completely [11] and another, from *Anabaena cylindrica*, partially [12]. Both of these are m-type thioredoxins. From higher plants the sequence of thioredoxin m from spinach chloroplasts is known [13].

At the gene level, full-length cDNA clones of thioredoxins have been reported from two bacteria, *E. coli* [14–17] and *Corynebacterium nephridii* [18], three cyanobacterial strains, *Anabaena* 7119 [19] and *Anabaena* 7120 [20], *Anacystis nidulans* [21] and for chicken [22] and human thioredoxin [23]. Higher plant thioredoxin genes have not yet been cloned.

So far no information on the primary structure of f-type thioredoxins has been available except for the peptide sequence around the active site [24] which established that this regulatory protein is a thioredoxin. In this paper we present the complete primary structure of thioredoxin f from spinach chloroplasts obtained by protein sequencing and the isolation and analysis of cDNA clones for thioredoxin f. In addition, we present evidence that thioredoxin f is synthesized as a

Correspondence to P. Schürmann, Laboratoire de Biochimie, Université de Neuchâtel, Chemin de Chantemerle 18, CH-2000 Neuchâtel, Switzerland

Abbreviation. Hse, homoserine.

Enzymes. Trypsin (EC 3.4.21.4); carboxypeptidase A (EC 3.4.17.1.); carboxypeptidase B (EC 3.4.17.2.); carboxypeptidase P (EC 3.4.16.1.); fructose-1,6-bisphosphatase (EC 3.1.3.11); $NADP^+$-dependent malate dehydrogenase (EC 1.1.1.82); restriction endonucleases (EC 3.1.21.4); DNA polymerase I (EC 2.7.7.7); T4 DNA ligase (EC 6.5.1.1).

Note. These sequence data will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession number X14959 thioredoxin f.

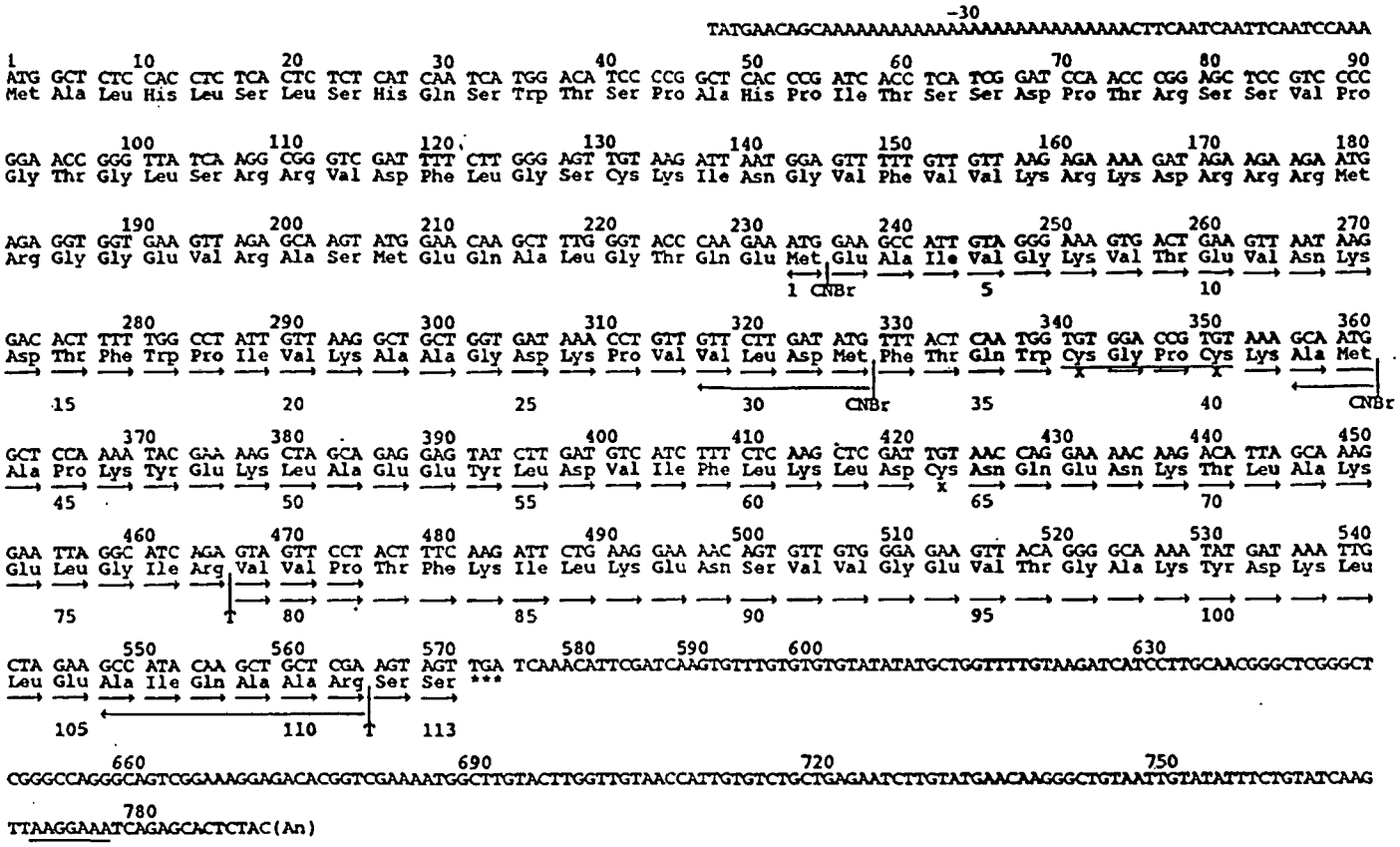

Fig. 1. *Amino acid sequence of native thioredoxin f and nucleotide sequence with deduced amino acid sequence of the p6SocTf-4 insert.* Details are given in the text. (→) Edman degradation; (←) carboxypeptidase digestion. Cleavage sites by trypsin (T) and CNBr are indicated. The active site and a potential polyadenylation signal are underlined

precursor protein, imported into the intact chloroplast and processed to its mature size.

MATERIALS AND METHODS

Restriction endonucleases, T7 RNA polymerase, *E. coli* DNA polymerase I large fragment and T4 DNA ligase were purchased from Boehringer (Mannheim) and Biolabs Inc. (Bad Schwalbach, Taunus), [α-^{32}P]dNTP (specific activity 400 Ci/mmol) and [^{35}S]methionine (specific activity 800 Ci/mmol) were from Amersham (Braunschweig). Spinach (*Spinacia oleracea* var. Monatol) was grown for 3–4 weeks in Einheitserde with 14 h illumination daily. Intact chloroplasts were purified in Percoll (Pharmacia) gradients, washed and used for import experiments essentially as described in [25, 26]

Thioredoxin f was purified from spinach leaves essentially as described earlier [27] except that for the last chromatographic step the protein was eluted by FPLC from a Mono Q column with a 0–350-mM NaCl gradient in 20 ml 20 mM Tris/Cl, pH 7.9.

The ultraviolet spectrum of thioredoxin f in 20 mM Tris/Cl, pH 7.9 was obtained with a Perkin-Elmer model 555 spectrophotometer. It shows a single peak at 280 nm and a shoulder at 288 nm. Aliquots of this solution were used for the determination of the absorption coefficient which was calculated to be 17700 M^{-1} cm^{-1}. Fluorescence emission spectra were determined as in [13].

Antibodies against thioredoxin f were raised in rabbits by injecting the pure protein emulsified with Freund's complete adjuvant. The IgG fraction of the serum was purified by ammonium sulfate precipitation, DEAE-cellulose and Sephacryl S-300 chromatography. The purified antibody fraction was concentrated by ultrafiltration and stored frozen at −20°C.

Thioredoxin f was carboxymethylated with iodo[^{3}H]-acetate (100 Ci/mol, Amersham) as described in [28]. The carboxymethylated protein was either cleaved by CNBr or succinylated with [^{14}C]succinic anhydride (100 Ci/mol Amersham) according to [29] and digested with trypsin. The peptides were separated by molecular sieving through Biogel P10 and P2 columns (0.9 × 120 cm) and further purified by HPLC using a TSK IEX 540K column (Toyosada, Japan) with a linear gradient of 0.1–1 M NH_4HCO_3 or by reverse phase chromatography on a Microsorb C_{18} column (Rainin Instruments, USA) with a gradient of 0–100% acetonitrile in 0.1% trifluoroacetic acid. Amino acid sequences were determined by Edman degradation with an Applied Biosystems 477A/120A sequencer [30]. Phenylthiohydantoin amino acid derivatives from sequencer degradations were identified on line. Total amino acid and peptides compositions were determined by an Irica (A-5500, Japan) amino acid analyzer. C-terminal sequences were determined using carboxypeptidases A, B (Sigma Biochemicals) and P (Takara, Japan) [31].

The construction of the recombinant λgt11 cDNA libraries using polyadenylated RNA from etiolated spinach seedlings illuminated for 4 h and 16 h has been described [32, 33]. The phage were digested with *Eco*RI, and the cDNA inserts were recloned into the *Eco*RI sites of vectors M13mp18 and Bluescript M13$^+$ for sequence analysis. The clones were sequenced by the dideoxy nucleotide chain termination method

Table 1. *Amino acid compositions of thioredoxin f and its peptides obtained by CNBr cleavage and tryptic digestion*
Values of amino acids in bold-face type were taken as the bases for the calculations. Values in parentheses are integer values of the number of amino acids. The data is taken from the sequence

Residue	Cyanogen bromide peptides						Tryptic peptides				Thioredoxin f
	C(2–32)		C(33–43)		C(44–113)		T(1–78)		T(79–111)		
	mol/mol										
Asx	4.8	(4)	0.4	(0)	6.1	(6)	8.5	(8)	2.1	(2)	10
Thr	1.8	(2)	0.8	(1)	2.4	(3)	3.2	(4)	1.5	(2)	6
Ser	0.3	(0)	0.6	(0)	2.1	(3)	0.8	(0)	1.3	(1)	3
Glx	2.4[a]	(2)	1.6[a]	(1)	10.4	(10)	9.4	(9)	4.1	(4)	13
Pro	1.8	(2)	1.1	(1)	2.1	(2)	4.5	(4)	1.0	(1)	5
Gly	2.4	(2)	1.5	(1)	3.5	(3)	4.6	(4)	2.5	(2)	6
Ala	3.0	(3)	1.0	(1)	7.0	(7)	7.0	(7)	4.0	(4)	11
Val	5.1	(6)	0	(0)	5.1	(6)	6.0	(7)	4.1	(5)	12
Met	0	(1)	0	(1)	0	(0)	2.3	(3)	0	(0)	3
Ile	1.7	(2)	0	(0)	3.2	(4)	3.3	(4)	1.5	(2)	6
Leu	1.1	(1)	0	(0)	9.1	(9)	6.6	(7)	2.7	(3)	10
Tyr	0.2	(0)	0	(0)	2.4	(3)	1.7	(2)	0.7	(1)	3
Phe	1.1	(1)	1.2	(1)	2.0	(2)	3.1	(3)	1.0	(1)	4
His	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0
Lys	3.6	(4)	1.0	(1)	8.2	(9)	10.4	(10)	4.2	(4)	14
Arg	0	(0)	0	(0)	2.1	(2)	1.1	(1)	1.0	(1)	2
Cys	0	(0)	1.2[b]	(2)	0.7[b]	(1)	2.4[b]	(3)	0	(0)	3
Trp	–	(1)	–	(1)	–	(0)	–	(2)	–	(0)	2
Total		(31)		(11)		(70)		(78)		(33)	113

[a] The peak included homoserine.
[b] Cysteine was analyzed as carboxymethyl cysteine.

[34] of single stranded (M13) or double-stranded (Bluescript M13$^+$) templates.

The 5'-terminal 136-bp *Eco*RI/*Bst*NI fragment of p6SocTf-2 used for plaque screening was cloned into appropriately digested Bluescript vector. DNA (2 µg) was digested with 4 units *Hind*III, and transcribed with 30 units T7 polymerase in the presence of 0.5 mM ribonucleotides in a final volume of 30 µl. RNA transcribed from 0.5 µg p6SocTf-4 DNA linearized with *Sph*I was translated in a rabbit reticulocyte assay (Amersham, Braunschweig) following the supplier's instructions, in the presence of 500 µCi [^{35}S]methionine. The translation products were used for *in organello* import experiments as described in [33]

RESULTS AND DISCUSSION

Sequence analysis

The complete amino acid sequence of mature thioredoxin f from spinach chloroplasts is presented in Fig. 1. It has been obtained by sequence analysis of peptides produced by cyanogen bromide and trypsin cleavage. In different experiments, including Edman degradation, the amino terminus of thioredoxin f was found to be blocked. Since the amino acid composition [27] showed two or three methionine and two arginine residues, two fragmentation methods were used: cyanogen bromide cleavage and trypsin digestion of succinylated protein. Cyanogen bromide cleavage followed by molecular sieving chromatography (Bio-Gel P2) resulted in three fragments, C(2–32), C(33–43) and C(44–113). Table 1 lists the compositions of these three peptides. Sequencing of C(2–32) was carried out by Edman degradation through 31 steps; carboxypeptidase P digestion showed the carboxy-terminal sequence, Val-Leu-Asp-Hse. The peptide C(33–43) was completely sequenced by 11-step Edman degradation and carboxypeptidase A digestion which gave the carboxy-terminal dipeptide sequence to be Ala-Hse. The peptide C(44–113) was partially sequenced by Edman degradation (38 steps). Carboxypeptidase digestion did not liberate Hse, suggesting that this peptide constitutes the carboxyl end of the protein. The order of C(2–32) and C(33–43) had been established in the previous report [24]. Thus these three CNBr peptides were linked as C(2–32) – C(33–43) – C (44–113).

Succinylated thioredoxin f was digested with trypsin which yielded, after Bio-Gel P10 column chromatography the two peptide fragments T(1–78) and T(79–111). Their composition is shown in Table 1. The amino terminus of the peptide T(1–78) was also found to be blocked, suggesting this peptide to be the amino end of the protein. The other tryptic peptide T(79–111) was sequenced completely by Edman degradation and carboxypeptidase digestion as shown in Fig. 1. The sequence of C(44–113) overlapped with those of T(1–78) and T(79–111), thus connecting these two peptides. The comparison of the peptide chains indicates that the tryptic chain has one methionine residue at the amino terminus and the cyanogen bromide chain has two serine residues at the carboxy terminus. In fact, seryl-serine was recovered from the tryptic digest, but an amino-terminal-blocked methionine derivative was not observed in the cyanogen bromide cleavage mixture even by fast atom bombardment or field desorption mass spectrometry. We provisionally conclude that a blocked methionine residue is at the amino terminus of the protein and that the two serine residues at the carboxy end are at positions 112 and 113. Further work to identify the exact amino terminus is in progress.

The protein sequencing results are in complete agreement with the amino acid sequence deduced from the nucleotide

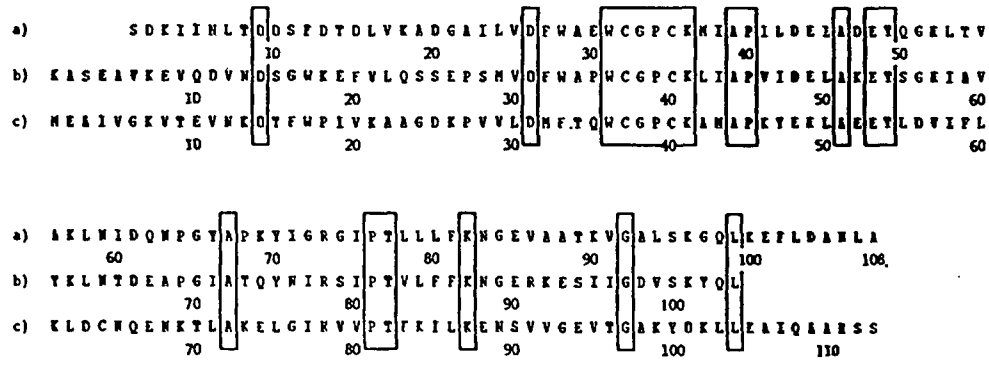

Fig. 2. *Comparison of the amino acid sequences of thioredoxins.* (a) Thioredoxin from *E. coli* [7, 14], (b) thioredoxin mb from spinach chloroplasts [13], (c) thioredoxin f from spinach chloroplasts. Strictly conserved residues are enclosed in frames

Table 2. *Thiol contents of spinach chloroplast thioredoxins*
Thiol contents were determined according to Habeeb [59] and are expressed as number of sulphydryl groups/mol protein. Thioredoxins were reduced by incubation with 10 mM dithiothreitol under argon for 20 min in 100 mM phosphate buffer, pH 8.0, containing 1 mM EDTA. Reductant was removed by FPLC through a G-25SF column equilibrated with the incubation buffer kept under nitrogen. The protein-containing fractions were immediately reacted with 5,5'-dithiobis(2-nitrobenzoate)

Protein	Oxidized		Reduced	Reduced minus oxidized
	− SDS	+ SDS		
	SH groups/mol protein			
Thioredoxin m	0	0	1.7	1.7
Thioredoxin f	0.8	0.8	2.9	2.1

sequence (Fig. 1) and confirm the most important feature, the strictly conserved active-site sequence -Cys-Gly-Pro-Cys- of thioredoxins we had reported earlier [24]. Thioredoxin f contains 113 amino acids (Table 1) resulting in a calculated molecular mass of 12564 Da. Its size is comparable to all the other thioredoxins whose primary structures are known. However there have been f-type thioredoxins isolated from *Anabaena* [35], *Scenedesmus* [36] and spinach [37,38] which are larger in size with apparent molecular masses of 25.5 kDa, 28 kDa and 18 kDa, respectively. Apart from the fact that these thioredoxins activate fructose-bisphosphatase and show affinity towards blue Sepharose, they seem to have little in common with the thioredoxin f described here.

The amino acid composition deduced from the sequence agrees quite well with the composition reported earlier [39], another fact which we consider is in favor of the proposed N-terminus. Thioredoxin f has no histidine, like thioredoxin m from spinach [13], but in contrast contains three half-cystines; two of them make up the active-site disulfide [24], whereas the third, revealed also by the thiol titrations summarized in Table 2, is present as Cys64 in the C-terminal half of the protein. The results in Table 2 indicate that in oxidized as well as reduced thioredoxin f Cys64 is easily accessible to thiol reagents.

The presence of additional Cys is known from other thioredoxins. All mammalian thioredoxins contain two structural half-cystine residues in their C-terminal half. There, they may play a role in regulation of thioredoxin activity, since their oxidation leads to the formation of an intramolecular disulfide and to the loss of activity [1]. A third Cys was recently found in the C-terminal half of a second thioredoxin from *Corynebacterium nephridii* [18]. The function of this protein however is not known. It shows 25% similarity with thioredoxin f, and its third Cys is positioned four residues closer to the active site than in thioredoxin f. A third Cys was also reported for the f-type thioredoxin from *Scenedesmus* [36] but no sequence information is available.

In the active-site sequence which we reported earlier [24], Met43 was erroneously positioned 16 residues further away from the active site towards the C-terminus, due to contamination of the peptide analyzed. This error is corrected by the present analysis. In contrast to thioredoxin m from spinach chloroplasts we observed no N-terminal redundancy. There is only one f-type thioredoxin in spinach chloroplasts [27], the one described here. We have never been able to find evidence for the second, dimeric thioredoxin f reported in [37, 38].

Due to the lower Trp contents of thioredoxin f, its absorption coefficient (17700 M^{-1} cm^{-1}) is smaller than that for thioredoxin m (19300 M^{-1} cm^{-1} [13]) which contains one more Trp residue. Oxidized thioredoxin f, when excited at 280 nm, shows a weak fluorescence maximum around 330 nm which does not increase upon reduction of the protein, as was also observed with thioredoxin f from *Anabaena* sp 7119 [35] and *Scenedesmus* [36]. This contrasts with the large fluorescence increases observed with thioredoxin from *E. coli* [40] and several other thioredoxins [8, 41, 42]. This fluorescence increase is attributed to Trp28 in *E. coli* thioredoxin (compare Fig. 2) and signals a localized conformational change in the active site [43, 44]. The replacement of the corresponding Trp by Phe33 in thioredoxin f explains the absence of any fluorescence changes upon reduction of this protein.

Sequence comparison

In order to compare thioredoxin f with its companion thioredoxin mb from spinach chloroplasts which is related to the *E. coli* thioredoxin, the primary structures of these three thioredoxins have been aligned, starting from their active sites, in Fig. 2. With its 113 amino acids thioredoxin f is the longest polypeptide of the three. Only 27 out of its 113 residues (24%) are found at identical positions in thioredoxin mb or thioredoxin from *E. coli*, whereas 49 out of 104 residues (47%) are identical between thioredoxin mb and *E. coli* thioredoxin. When admitting conservative changes the percentage similarity increases to 31% with thioredoxin mb, 33% with *E. coli* thioredoxin, but to 60% between thioredoxin mb and that of *E. coli*. Similarities between all the bacterial thioredoxins are

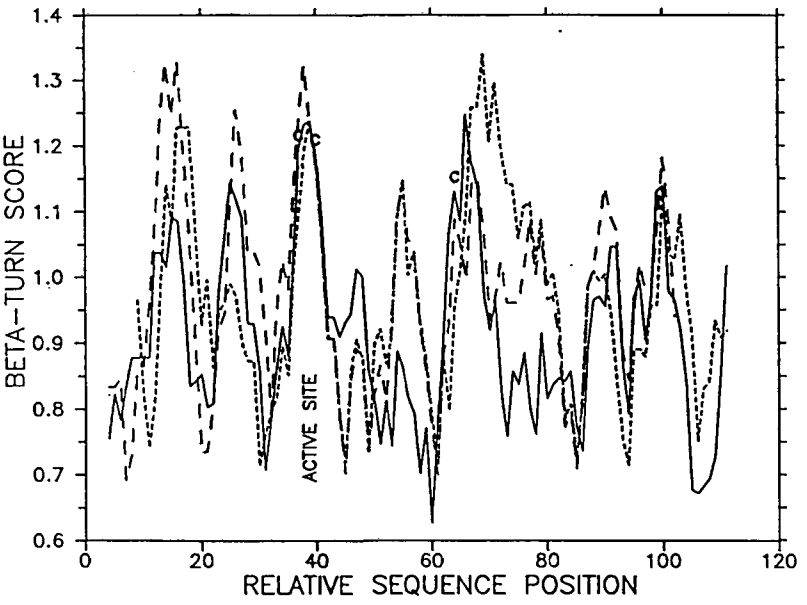

Fig. 3. *Comparison of the β turn scores, calculated with a six-residue window according to [45] for thioredoxin from* E. coli *(------), thioredoxin mb (– – –) and thioredoxin f (———) from spinach chloroplasts.* The plots are aligned at the active site. C marks the positions of Cys in thioredoxin f

in the range 50–60% [9(< 53%), 11(49%), 18(> 50%)]. These results clearly show that the primary structure of thioredoxin f is quite different from prokaryotic and other plant thioredoxins and suggests that it may have diverted from a common ancestral protein early in evolution. It is interesting to note that thioredoxin f shows some striking similarity with the mammalian thioredoxins [22, 23]

Despite differences in primary structure the secondary and tertiary structures of thioredoxin f might be quite similar to those of the well characterized thioredoxin from *E. coli*. We have used the method of Chou and Fasman [45] for predicting secondary structure based on sequence information, to compare the three thioredoxins. The analyzes, shown only for the β-turns in Fig. 3, suggest that the secondary structures of all three thioredoxins are quite similar and that, therefore, also the folding of the proteins might be alike. This will expose similar residues to the surface of the protein molecule where they can influence the interaction of thioredoxins with target proteins. Conformational studies on *E. coli* thioredoxin have indicated that certain residues (Gly33, Pro34, Ile75, Pro76, Val91, Gly92, Ala93) make up a hydrophobic patch on the protein surface close to the active site and that this area might be responsible for the protein-protein interactions [46]. Recent studies with mutant *E. coli* thioredoxin [47] and hybrid *E. coli/Anabaena* thioredoxin [48] support the idea that certain residues in the C-terminal half of the protein are important for the thioredoxin-protein interactions. In thioredoxin f all the residues of the hydrophobic patch are conserved except for Val91 which is replaced by Thr96, a change that should not drastically alter its properties. Apart from these residues around the active center, there are still more strictly conserved residues, especially charged residues, as can be seen in Fig. 2.

The apparent spatial similarities between the three thioredoxins contrast with the specificity of thioredoxin f in its interaction with chloroplast fructose-1,6-bisphosphatase and other chloroplast enzymes. It is known that fructose-1,6-bisphosphatase forms a complex with thioredoxin f [38, 49, 50] but apparently not with thioredoxin m nor with thioredoxin from *E. coli* since neither of these thioredoxins can activate fructose-1,6-bisphosphatase [24]. Some unique structural property seems to be responsible for the specific interaction between thioredoxin f and its target enzymes. A conspicuous difference in the C-terminal part of thioredoxin f is the presence of Cys64 which is apparently located at the surface of the protein. Experiments to determine whether this Cys is instrumental in thioredoxin f specificity are under way.

Isolation and characterization of cDNA clones

Approximately 10^5 recombinant phage from λgt11 expression libraries made from cotyledon polyadenylated RNA of spinach seedlings illuminated for 4 h or 16 h after etiolation [32, 33] were screened with homologous, monospecific polyclonal antibodies raised against pure thioredoxin f. Positive plaques (3 in 10^5 screened) were rescreened at low plaque density and phage DNAs were prepared from single reacting plaques by the plate lysis method [51]. The inserts of the three phage which were selected in this screen were excised with *Eco*RI, sized on polyacrylamide gels, recloned into the expression vector Bluescript $M13^+$ (Stratagene, San Diego) and sequenced. A 5′-terminal 136-bp *Eco*RI/*Bst*NI fragment of the largest insert (550 nucleotide pairs) was subsequently cloned into the same vector and used as a hybridization probe to select phage that carried inserts encoding the entire protein-coding region of a thioredoxin f mRNA. Eleven phage with inserts of 700–900 nucleotide pairs were obtained which approached the size of the message, approximately 1100 nucleotides, as determined by Northern blot analysis with poly(A)-rich RNA from spinach seedlings. Restriction maps of these inserts showed a number of common restriction sites, suggesting that the clones were very similar. This was substantiated by nucleotide sequence analysis.

The complete nucleotide sequence and deduced amino acid sequence for the largest cDNA clone (p6SocTf-4) selected are shown in Fig. 1. The insert is 852 nucleotide pairs long excluding a 3′ poly(A) tail of 41 nucleotides. The sequence has one long open reading frame extending for 570 nucleotides. The first methionine occurs at the 15th codon from the

5′ end of the open reading frame which results in a polypeptide of 190 amino acid residues (21 kDa). Alignment of the spinach cDNA and protein sequences to give maximum sequence match indicates that the derived amino acid sequences of all the inserts sequenced correspond to the amino acid sequence determined by protein sequencing. We conclude that p6SocTf-1 to p6SocTf-14 are very similar to thioredoxin f mRNAs. Moreover, the sequences of all inserts were identical, with the exception of one, p6SocTf-5, which is 45 nucleotides shorter at the 3′ end and lacks a poly(A) track, suggesting that all inserts represent a single mRNA species. Preliminary gene-copy-number determinations by Southern blotting indicate that there are two genes for thioredoxin f/haploid genome (A. Schneiderbauer, personal communication).

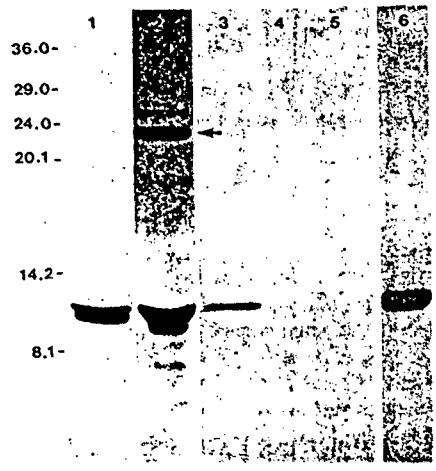

Fig. 4. In vitro *transcription/translation of p6SocTf-4 RNA and uptake of the* in vitro *thioredoxin f precursor product into isolated spinach chloroplasts.* Fluorograph of a 15% denaturing SDS/8 M urea polyacrylamide gel [58]. Lane 1, control translation without added RNA; lane 2, products of translation of p6SocTf-4 RNA in a rabbit reticulocyte lysate; lane 3, product found in the stromal fraction from unbroken chloroplasts after import of the thioredoxin precursor and thermolysin treatment of the organelles; lane 4, equivalent data from the thylakoid fraction; lane 5, stromal fraction after treatment with thermolysin, and lane 6, authentic thioredoxin f (Coomassie brilliant blue stain). Molecular mass markers: glyceraldehyde-phosphate dehydrogenase (36.0 kDa), carbonic anhydrase (29.0 kDa), trypsinogen (24.0 kDa), soybean trypsin inhibitor (10.0 kDa) and two CNBr fragments of myoblobin (14.2 and 8.1 kDa)

Import in organello

Biochemical evidence suggests that thioredoxin f is located exclusively in chloroplasts [27]. This implies that the protein, synthesized in the cytosol, has to be imported into the organelle. Since precursors that are processed to mature forms have been demonstrated for all nuclear-encoded, plastid-located proteins studied, e.g. [52, 53], it is possible that thioredoxin f is synthesized as a precursor polypeptide. Transient N-terminal sequences found on nuclear-encoded plastid proteins serve to initiate transport across the chloroplast envelope. Several structurally and, possibly, functionally distinct regions have been postulated to represent basic domains of such sequences, cf. [52, 54, 55]. The structural elements for recognition include positively charged and polar amino acid residues as well as residues with small side-chains at positions −1 and −3, i.e. preceding the endonucleolytic cleavage site and, if translocation through thylakoid membranes is required, a terminal hydrophobic region that is preceded by an amphipathic β sheet [55, 56]. Following the assumed N-terminal methionine residue of thioredoxin f precursor is, in fact, a region of approximately 70 amino acid residues with clustered positive and hydroxylated amino acids but also a conspicuous presence of four or six negatively charged residues, depending on the assumed processing site. Based on predictions for signal sequence cleavage sites [54], a potential processing site for the putative transit sequence could be the serine residue, 68 amino acids from the assumed initial methionine. However, our protein sequencing data suggest that the cleavage site is after the Glu, 77 amino acids from the assumed initial methionine. Since the N-terminus of the native protein is blocked, the question concerning the exact cleavage site remains to be clarified.

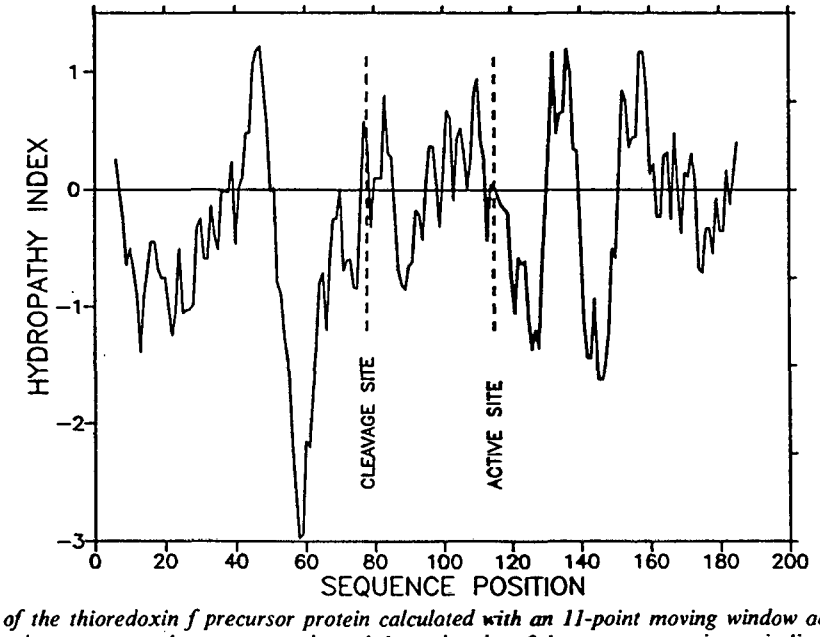

Fig. 5. *Hydropathy plot of the thioredoxin f precursor protein calculated with an 11-point moving window according to [57].* The proposed cleavage site between transit sequence and mature protein and the active site of the mature protein are indicated

In order to test that thioredoxin f is produced as a precursor and that the isolated cDNA clones carried the entire transit sequence with the basic attributes for intracellular sorting and import into the chloroplast, p6SocTf-4 DNA linearized with *Sph*I was transcribed with T7 polymerase and precursor protein was synthesized from the resulting transcript in a rabbit reticulocyte cell-free translation assay, in the presence of [^{35}S]methionine. This polypeptide was then offered to isolated intact spinach chloroplasts. The result of this experiment is shown in Fig. 4. The major immunologically reacting translation product of approximately 21 kDa could be imported and processed to its mature size in a protease-insensitive form. The predicted molecular mass of 12.5 kDa, excluding the putative N-terminal transit sequence of 77 residues, agrees well with that determined by electrophoresis in denaturing polyacrylamide gels (12 kDa).

Fig. 5 shows a hydropathy plot [57] of the putative thioredoxin f precursor. This plot is consistent with biochemical findings [27] that thioredoxins are both hydrophilic and components of the stroma. The presumed transit sequence is dissimilar in structure to those of the luminal proteins [55, 56] but clearly resembles those of matrix components, e.g. [52, 53].

CONCLUSIONS

This is the first report of the complete primary structure of an f-type thioredoxin and of the isolation of a higher plant cDNA clone coding for thioredoxin. The protein, although quite different from other thioredoxins in its primary structure, must have maintained a very similar three-dimensional structure. Some surface residues should be responsible for the specificity of interaction with target enzymes. Thioredoxin f is synthesized in the cytosol as a precursor protein of larger size which is imported into the chloroplast stroma where the signal peptide is cut off and the N-terminal blocked.

This work was supported by the *Schweizerischer Nationalfonds* (grants 3.639-0.84 and 3.626-0.87), the Roche Research Foundation, by Grants in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan, by the *Deutsche Forschungsgemeinschaft* (grants He 693; SFB-184) and the *Fonds der Chemischen Industrie*.

REFERENCES

1. Holmgren, A. (1985) *Annu. Rev. Biochem. 54*, 237–271.
2. Buchanan, B. B., Wolosiuk, R. A. & Schürmann, P. (1979) *Trends Biochem. Sci. 4*, 93–96.
3. Buchanan, B. B. (1980) *Annu. Rev. Plant Physiol. 31*, 341–374.
4. Cséke, C. & Buchanan, B. B. (1986) *Biochim. Biophys. Acta 853*, 43–63.
5. Johnson, T. C., Cao, R. Q., Kung, J. E. & Buchanan, B. B. (1987) *Planta 171*, 321–331.
6. Florencio, F. J., Yee, B. C., Johnson, T. C. & Buchanan, B. B. (1988) *Arch. Biochem. Biophys., 266*, 496–507.
7. Holmgren, A. (1968) *Eur. J. Biochem. 6*, 475–484.
8. Meng, M. & Hogenkamp, H. P. C. (1981) *J. Biol. Chem. 256*, 9174–9182.
9. Mathews, W. R., Johnson, R. S., Cornwell, K. L., Johnson, T. C., Buchanan, B. B. & Biemann, K. (1987) *J. Biol. Chem. 262*, 7537–7545.
10. Clement-Metral, D., Holmgren, A., Cambillau, C., Jörnvall, H., Eklund, H., Thomas, D. & Lederer, F. (1988) *Eur. J. Biochem. 172*, 413–419.
11. Gleason, F. K., Whittacker, M. M., Holmgren, A. & Jörnvall, H. (1985) *J. Biol. Chem. 260*, 9567–9573.
12. Ip, S.-M., Rowell, P., Aitken, A. & Stewart, W. D. P. (1984) *Eur. J. Biochem. 141*, 497–504.
13. Maeda, K., Tsugita, A., Dalzoppo, D., Vilbois, F. & Schürmann, P. (1986) *Eur. J. Biochem. 154*, 197–203.
14. Höög, J.-O., Von Bahr-Lindström, H., Josephson, S., Wallace, B. J., Kushner, S. R., Jörnvall, H. & Holmgren, A. (1984) *Biosci. Rep. 4*, 917–923.
15. Wallace, B. J. & Kushner, S. R. (1984) *Gene 32*, 399–408.
16. Lim, C.-J., Geraghty, D. & Fuchs, J. A. (1985) *J. Bacteriol. 163*. 311–316.
17. Tabor, S., Huber, H. E. & Richardson, C. C. (1986) in *Thioredoxin and glutaredoxin systems: structure and function* (Holmgren, A., Brändén, C.-I., Jörnvall, H. & Sjöberg, B.-M., eds) pp. 285–300, Raven Press, New York.
18. Lim, C.-J., Fuchs, J. A., McFarlan, S. C. & Hogenkamp, H. P. C. (1987) *J. Biol. Chem. 262*, 12114–12119.
19. Lim, C.-J., Gleason, F. K. & Fuchs, J. A. (1986) *J. Bacteriol. 168*. 1258–1264.
20. Alam, J. & Curtis, S. E. (1986) *VIIth Int. Congress on Photosynthesis*, abstr. 308–231, Brown University (USA).
21. Muller, E. G. D. & Buchanan, B. B. (1989) *J. Biol. Chem. 264*. 4008–4014.
22. Jones, S. W. & Luk, K.-C. (1988) *J. Biol. Chem. 263*, 9607–9611.
23. Wollman, E. E., D'Auriol, L., Rimsky, L., Shaw, A., Jacquot, J.-P., Wingfield, P., Graber, P., Dessarps, F., Robin, P., Galibert, F., Bertoglio, J. & Fradelizi, D. (1988) *J. Biol. Chem. 263*. 15506–15512.
24. Tsugita, A., Maeda, K. & Schürmann, P. (1983) *Biochem. Biophys. Res. Commun. 115*, 1–7.
25. Bartlett, S. G., Grossmann, A. R. & Chua, N.-H. (1982) in *Methods in chloroplast molecular biology* (Edelman, M., Hallick, R. B., Chua, N.-H., eds) pp. 1081–1091, Elsevier, Amsterdam, New York, Oxford.
26. Westhoff, P., Jansson, Ch., Klein-Hitpass, L., Berzborn, R., Larsson, Ch. & Bartlett, S. G. (1985) *Plant Mol. Biol. 4*, 137–146.
27. Schürmann, P., Maeda, K. & Tsugita, A. (1981) *Eur. J. Biochem. 116*, 37–45.
28. Crestfield, A. M., Moore, S. & Stein, W. H. (1963) *J. Biol. Chem. 238*, 622–627.
29. Habeeb, A. F. S. A., Cassidy, H. G. & Singer, S. J. (1958) *Biochim. Biophys. Acta 29*, 587–593.
30. Hewick, R. M., Hunkapiller, M. W., Hood, L. E. & Dryer, W. J. (1981) *J. Biol. Chem. 256*, 7990–7997.
31. Kamp, M. & Tsugita, A. (1987) *J. Biochem. (Tokyo) 102*, 243–246.
32. Tittgen, J., Hermans, J., Steppuhn, J., Jansen, T., Jansson, C., Andersson, B., Nechushtai, R., Nelson, N. & Herrmann, R. G. (1986) *Mol. Gen. Genet. 204*, 258–265.
33. Lautner, A., Klein, R., Ljungberg, U., Reiländer, H., Bartling, D., Andersson, B., Reinke, H., Beyreuther, K. & Herrmann, R. G. (1988) *J. Biol. Chem. 263*, 10077–10081.
34. Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) *Proc. Natl Acad. Sci. USA 74*, 5463–5467.
35. Whittacker, M. M. & Gleason, F. K. (1984) *J. Biol. Chem. 259*. 14088–14093.
36. Langlotz, P., Wagner, W. & Follmann, H. (1985) *Z. Naturforsch. 41c*, 275–283.
37. Soulié, J.-M., Buc, J., Meunier, J.-C., Pradel, J. & Ricard, J. (1981) *Eur. J. Biochem. 119*, 497–502.
38. Buc, J., Rivière, M., Gontero, B., Sauve, P., Meunier, J.-C. & Ricard, J. (1984) *Eur. J. Biochem. 154*, 485–488.
39. Schürmann, P., Tsugita, A., Maeda, K., Dalzoppo, D. & Vilbois, F. (1986) in *Thioredoxin and glutaredoxin systems: structure and function* (Holmgren, A., Brändén, C.-I., Jörnvall, H. & Sjöberg, B.-M., eds) pp. 31–38, Raven Press, New York.
40. Stryer, L., Holmgren, A. & Reichard, P. (1967) *Biochemistry 6*. 1016–1020.
41. Clement-Metral, J. D., Höög, J.-O. & Holmgren, A. (1986) *Eur. J. Biochem. 161*, 119–126.

42. Gleason, F. K. & Holmgren, A. (1981) *J. Biol. Chem. 256*, 8306–8309.
43. Holmgren, A. (1981) *Biochemistry 20*, 3204–3207.
44. Holmgren, A. (1972) *J. Biol. Chem. 247*, 1992–1998.
45. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251–276.
46. Eklund, H., Cambillau, C., Sjöberg, B.-M., Holmgren, A., Jörnvall, H., Höög, J.-O. & Brānden, C.-I. (1984) *EMBO J. 3*, 1443–1449.
47. Huber, H. E., Russel, M., Model, P. & Richardson, C. C. (1986) *J. Biol. Chem. 261*, 15006–15012.
48. Lim, C.-J., Gleason, F. K., Jacobson, B. A. & Fuchs, J. A. (1988) *Biochemistry 27*, 1401–1408.
49. Pla, A. & Lopez-Gorge, J. (1981) *Biochim. Biophys. Acta 636*, 113–118.
50. Soulié, J.-M., Buc, J., Rivière, M. & Ricard, J. (1985) *Eur. J. Biochem. 152*, 565–568.
51. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular cloning: a laboratoty manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
52. Von Heijne, G., Steppuhn, J. & Herrmann, R. G. (1989) *Eur. J. Biochem. 180*, 535–545.
53. Jansen, T., Reiländer, H., Steppuhn, J. & Herrmann, R. G. (1988) *Curr. Genet. 13*, 517–522.
54. Von Heijne, G. (1986) *J. Mol. Biol. 189*, 239–242.
55. Tyagi, A., Hermans, J., Steppuhn, J., Jansson, C., Vater. F. & Herrmann, R. G. (1987) *Mol. Gen. Genet. 207*, 288–293.
56. Jansen, T., Rother, C., Steppuhn, J., Reinke, H., Beyreuther, K., Jansson, C., Andersson, B. & Herrmann, R. G. (1987) *FEBS Lett. 216*, 234–240.
57. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.
58. Westhoff, P., Nelson, N., Bünemann, H. & Herrmann, R. G. (1981) *Curr. Genet. 4*, 109–120.
59. Habeeb, A. F. S. A. (1972) *Methods Enzymol. 25*, 457–464.

Bibliography

Original Contibutions

1. Kamo,M.,Tsugita,A.,Wiessner,C., Wedel,N., Bartling,D.,Herrmann,R.Aguilar,F.,Gardet-Salvi,L. and Schürmann,P. Primary structure of spinach chloroplast thioredoxin f. **Eur. J. Biochem.** 182:315-322 (1989)
2. Aguilar,F., P-E.Montandon and E. Stutz. Two genes encoding the soybean elongation factor eEF-1α are transcribed in seedling leaves. **Plant Molecular Biology** 17:351-360 (1991)
3. Aguilar,F.,B.Brunner,L.Gardet-Salvi,E.Stutz and P.Schürmann. Biosynthesis of active spinach-chloroplast thioredoxin f in transformed *E.coli*. **(in preparation)**

Abstracts

A-1 Aguilar.F.,Stutz,E. and schürmann,P. The primary structure of spinach chloroplast thioredoxin f. **USGEB**, 1987.

A-2 Aguilar,F.,Brunner,B.,Stutz,E. and Schürmann,P. Synthesis of active spinach thioredoxin f in transformed E. coli. **2nd International Congress of Plant Molecular Biology**, 1988.

A-3 Aguilar,F.,Montandon,P-E. and Stutz,E. Structure and organizytion of EF-1α in soybean. **USGEB**, 1990.

A-4 Job-Cei,C.,Aguilar,F. and Schürmann,P. Isolation and characterization of cDNA clones coding for soybean thioredoxins. **USGEB**, 1990.

Le texte complet de la thèse est déposé au Laboratoire de Biochimie Végétale. Ch. de Chantemerle 18, 2000 neuchâtel.