Published in Plant Science 117, issues 1-2, 83-93, 1996
which should be used for any reference to this work

The *tuf* gene family of soybean: structure and differential transcription[1]

Fabienne Maurer[a], Maximilien Murone[b], Erhard Stutz*[c]

[a]*Monash University, Department of Medicine, Box Hill Hospital, Box Hill 3128, Victoria, Australia*
[b]*ISREC, chemin des Boveresses 155, CH-1066 Epalinges, Switzerland*
[c]*Laboratoire de Biochimie végétale, Université de Neuchâtel, Rue Emile-Argand 11, CH-2007 Neuchâtel, Switzerland*

Abstract

The nuclear genome of soybean (*Glycine max* L.) contains a family of four *tuf* genes coding for the chloroplast translation elongation factor EF-Tu. Genomic and cDNA sequencing reveal that the four *tuf* genes belong to two subfamilies (type A and type B). *Tuf* genes A1/A2 are highly expressed in green organs (leaves and stems) while *tuf* genes B1/B2 are poorly expressed in leaves and stems. Both types of genes are about equally expressed in roots albeit at a low level. The 5′ flanking regions of *tuf*A1 and B1 are tested for promoter activities (GUS) in transformed green tobacco. The *tuf*A1 promoter is much stronger than the B1 promoter. Promoter deletion studies with *tuf*A1-derived constructs allowed identification of a DNA segment essential for gene transcription.

Keywords: Soybean; *Glycine max* L.; *Tuf* gene family; Chloroplast EF-Tu; Differential transcription

1. Introduction

We have recently identified and sequenced a soybean nuclear *tuf* gene (*tuf*A) coding for the chloroplast translation elongation factor EF-Tu [1]. Southern hybridization data suggested that the nuclear genome contains a 4-member *tuf* gene family and preliminary Northern experiments showed that *tuf*A mRNA concentration strongly increases upon illumination.

* Corresponding author, Tel.: +41 38 232205; fax: +41 38 232201; e-mail: Erhard.Stutz@Bota.unine.ch

[1] This paper is dedicated to the memory of Professor Hans Kössel who untimely passed away on December 24, 1995.

N. sylvestris has two *tuf* genes [2] per haploid genome and it was proposed [3] that these two genes were organ specifically expressed. One of the genes accounts for 75% of the *tuf* mRNAs in leaves while the other one seems to be preferentially expressed in roots. It was argued that differences in promoter structure may account for this finding. Two *tuf* genes were also found in *N. tabacum* [4]. A single *tuf* gene was found in *A. thaliana* [5] but several *tuf* genes seem to exist in other *Brassica* species [6].

In view of these results it was of interest to study the soybean *tuf* gene family in more detail. A second *tuf* gene (*tuf*B1) was completely sequenced and transcripts of all four types of *tuf* genes were

characterized by partial cDNA sequencing. Transcription of the four genes was studied under light and dark conditions and in different plant organs. To identify putative regulatory elements within the tuf genes a series of deletion mutants were prepared, containing progressively shorter sequences of the *tuf* promoters fused to the GUS reporter gene. GUS activities in transgenic tobacco were compared and results indicated that the *tuf*A1 promoter is much stronger than the *tuf*B1 counterpart confirming the in situ transcription studies.

2. Materials and methods

2.1. Plant material

Soybean (*Glycine max* L., cv. Ceresia [7] seedlings were grown in a phytotron at 25°C, with a 16 h light/8 h dark photoperiod and 70% relative humidity. Leaves were harvested during the light period after about 2 weeks of growth, immediately frozen in liquid nitrogen and stored at −80°C until use.

2.2. Isolation of DNA and RNA

Total DNA from soybean leaves (cv. Ceresia) was isolated following a procedure of Shure [8]. For RNA extractions, all solutions except those containing amines (Tris) were treated with diethyl pyrocarbonate (DEPC) according to Sambrook et al. [9]. Total soybean RNA was prepared according to Dean et al. [10] and as reported [1]. All RNA samples used in RT-PCR experiments were treated with 0.2 U/μg of RNase free DNase I (Boehringer Mannheim) for 1 h at 37°C in 50 mM Tris–HCl pH 7.4, 10 mM $MgCl_2$, 100 mM DTT, extracted twice with phenol/chloroform (1:1), precipitated with ethanol and resuspended in the appropriate buffer.

*2.3. Cloning and sequencing of genomic tuf*B1

From preliminary Southern hybridization experiments it was known that a B type *tuf* gene was located on an *Eco*RI fragment of about 7 kb length. We constructed a partial genomic *Eco*RI library, i.e. 50 μg of total DNA were digested with *Eco*RI (Boehringer), separated on a 0.6% agarose gel and fragments between 6 and 8 kb were electroeluted. A sample of 600 ng of the purified fragments were ligated to 1 μg of *Eco*RI λgt 10 arms for 20 h at 14°C in a reaction containing 50 mM Tris–HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 1 mM ATP and 1 U of T4-DNA ligase (Gibco-BRL). In vitro packaging reaction was done using the Promega Packagene Lambda A Packaging System according to the manufacturer's instructions. Library screening at high stringency was carried out following standard protocols [9] using *E. coli* C600 hfl as host strain [11]. As probe, we used a 379 bp *Kpn*I-*Xba*I *tuf* cDNA (B2) fragment [1]. A positive clone was isolated and purified. The 7 kb *tuf* insert was then subcloned into the phagemid pBluescriptII KS^+ (Stratagene). Sequencing followed standard protocols.

2.4. cDNA cloning and sequencing

A soybean (*Glycine max.* cv. Williams) seedling λgt 11 cDNA library purchased from Clontech was screened at high stringency using standard protocols [9]. The same cDNAB2 fragment as above was used as the probe for genomic cloning. Positive clones were purified and the corresponding inserts were subcloned into the phagemid pBluescript II SK^+ (Stratagene) and sequenced.

2.5. Gene expression

To determine relative transcript concentrations of the four *tuf* genes in different tissues and under light and dark conditions, we used an RT-PCR approach [12]. The reaction was carried out on total soybean RNA using a common RT-Tuf (5′ GGAATGAGGTTTGCTATTAG 3′) labeled primer and two unlabeled RT-Tuf A (5′ AAAGCAAAGTACAGAAGCTG 3′) and RT-Tuf B (5′ AAAACTCAATTGCACAAGCC 3′) primers designed to produce amplified fragments with small size differences. These primers have the same size and GC content. One μg of DNase I-treated total RNA was mixed with 100 pmol of RT-Tuf A and RT-Tuf B primers in a volume of 11.5 μl, heated to 95°C for 3 min and cooled on ice.

First strand synthesis was carried out in a volume of 25 μl containing the RNA/primer mix, 1 × RT-buffer (50 mM Tris–HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$), 1 mM dNTPs, 10 mM DTT, and 200 U of reverse transcriptase (SuperScript™ II RNase H⁻, Gibco-BRL) for 90 min at 45°C. The reaction was made up to 100 μl containing 1 × PCR-buffer (10 mM Tris–HCl pH 9, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Triton-X 100, 0.01% gelatin), an additional 0.2 mM dNTPs, 1 μM RT-Tuf A, 1 μM RT-Tuf B, 1 μM of labeled RT-Tuf primer and 0.5 U of Super Taq DNA polymerase (P.H. Stehelin & Cie AG, Basel). The PCR conditions were 4 min of denaturation at 94°C, followed by 30 cycles of 1 min denaturation at 94°C, 1 min hybridization at 55°C, and 1 min elongation at 72°C. In order to determine the relative amount of each transcript [13], 1 μl of RT-PCR products was analyzed on a sequencing gel after different PCR cycles alongside a DNA sequence ladder and autoradiographed. The relative intensity of the amplification signals was estimated on a bioimaging system (Bio Image® System, Millipore).

2.6. Antisense RNA probes and ribonuclease protection analysis

Radioactive antisense RNA probes were synthesized using 500 ng of proteinase K-treated linear DNA templates (Fig. 1), 1 × transcription buffer (40 mM Tris–HCl pH 8, 50 mM NaCl, 8 mM $MgCl_2$, 2 mM spermidine), 0.6 mM rNTP(-rATP), 15 mM DTT, 10 U of RNase inhibitor (RNasin, Promega), 50 μCi of [α-^{32}P] rATP 3000 Ci/mmol (Amersham) and 5 U of T7 RNA polymerase (Stratagene) during 40 min at 37°C and in a volume of 25 μl. Samples were treated with 40 U RNase free DNase I to eliminate DNA template, extracted twice with phenol/chloroform (1:1), precipitated with ethanol and resuspended in 20 μl of water. Fifty μg of total soybean RNA diluted in 30 μl of hybridization buffer (40 mM Pipes pH 6.4, 400 mM NaCl and 1 mM EDTA) were mixed with 1 μl of antisense probe, covered with mineral oil, denatured for 10 min at 85°C and hybridized overnight at 65°C. The mix was cooled to room temperature, and 300 μl of digestion buffer (10 mM Tris–HCl pH 7.4, 300 mM NaCl, 5 mM EDTA)

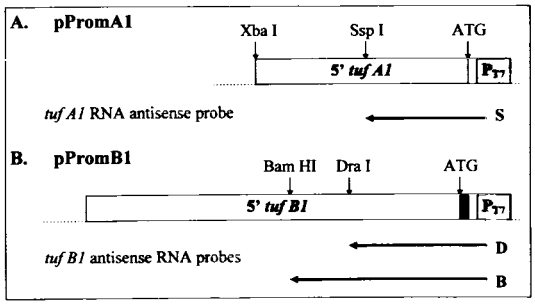

Fig. 1. Templates for synthesis of antisense RNA probes used for transcription start site analysis. (A) pPromA1 is a pBluescript II SK⁺ based construct containing 592 bp of the *tuf*A1 5′ region (positions −521 to +71). It was linearized with *Ssp*I and used as template. (B) pPromB1 contains 1230 bp of *tuf*B1 (positions −1156 to +74). It was linearized with *Dra*I or with *Bam*HI; both fragments were used as templates. P_{T7} = T7 promoter. Each RNA probe contains parts of the vector (55 nucleotides, white boxes) and *tuf*B1 RNA contains some coding sequences (11 nucleotides, black).

containing 40 μg/ml of RNase A (Sigma) and 60 U/ml of RNase T1 (Gibco-BRL) were added. The reaction was incubated for 1 h at 30°C, treated with 200 μg of proteinase K and 0.6% SDS during 30 min at 37°C, extracted with phenol/chloroform (1:1) and precipitated by adding 20 μg of yeast tRNA and ethanol. Protected hybrids were resuspended in 80% formamide, 10 mM EDTA and analyzed on a sequencing gel alongside a DNA sequence ladder.

2.7. Transgenic tobacco

Vectors containing promoter deletions (Fig. 2) were conjugated to *A. tumefaciens* LBA-4404 [14] by triparental mating with the helper plasmid pRK-2013 [15]. Introduction of the plasmids was confirmed by Southern blot hybridization (data not shown). We transformed tobacco (*N. tabacum*) by the leaf disk method [16]. Leaf disks of 5–7 mm diameter from sterile leaves of tobacco were cocultivated with the transformed *A. tumefaciens* LBA-4404 for 2 days on MS-medium [17] containing 1 μg/ml BAP (benzyl-amino-purine) and 0.1 μg/ml NAA (α-naphthaleneacetic acid). Leaf disks were transferred to the same medium as above,

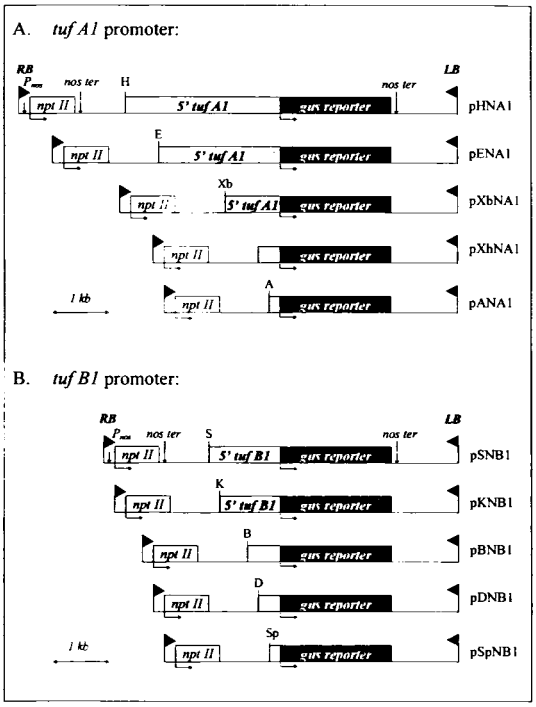

Fig. 2. Constructs for in vivo promoter activity studies in the binary vector pLP-100 (pROK-1 derived; Jefferson et al. [19]) carrying the *npt II* kanamycin resistance gene driven by nopaline synthase promoter (P_{nos}) and terminator (nos_{ter}). (A) Five progressive deletions of *tuf*A1 5′ flanking regions were fused to *E. coli* β-glucuronidase (GUS) coding sequence as follows: positions −2800 (approx.) to +71 (pHNA1), positions −2200 (approx.) to +71 (pENA1), positions −521 to +71 (pXbNA1), positions −226 to +71 (pXhNA1, *Exo*III/*Xho*I-generated), positions −80 to +71 (pANA1). (B) *tuf*B1 constructs: positions −1156 to +63 (pSNB1), positions −968 to +63 (pKNB1), positions −483 to +63 (pBNB1), positions −327 to +63 (pDNB1), positions −67 to +63 (pSpNB1). The constructs were introduced in pLP-100. Only the T-DNA regions are presented here. We used the following restriction sites: A, *Alu*I; B, *Bam*HI; D, *Dra*I; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; S, *Sac*I; Sp, *Spe*I; Xb, *Xba*I.

supplemented with 100 μg/ml cefotaxim and 100 μg/ml kanamycin. As soon as regenerated plantlets appeared, they were placed on a rooting medium (MS supplemented with 100 μg/ml cefotaxim and 50 μg/ml kanamycin). Plants were finally transferred to soil. T-DNA integrity, copy number and organization was determined by Southern blotting (data not shown).

2.8. Assays of GUS activity

Each independent *tuf*A1-derived transformant was selfed to obtain F1 seeds. Sterile seeds were grown up for 3 weeks on MS-medium containing 100 μg/ml kanamycin. Seedlings were frozen in liquid nitrogen and stored at −80°C. GUS activities were measured in whole extracts of 100 F1 plantlets following the procedure of Jefferson [18,19]. The fluorogenic reaction was performed in extraction buffer containing 1 mM MUG (4-methylumbelliferyl glucuronide) and 20 μl of extracts in a final volume of 2 ml. Protein concentrations were determined by the dye-binding procedure of Bradford [20] with a kit supplied by BioRad Laboratories.

3. Results

3.1. Evidence for two tuf gene subfamilies

Previous Southern hybridization data using two different *tuf* cDNA probes indicated that most likely two types of *tuf* genes exist on the soybean nuclear genome. We arbitrarily called the first sequenced gene *tuf*A1 (accession number X66062; [1]). A second *tuf* gene (*tuf*B1) was retrieved from a cloned 7.2 kb genomic *Eco*RI fragment from which 3079 bp were sequenced containing the entire coding part and adjacent regions (accession number X89058). To characterize additional genes we screened a soybean seedling Lambda gt11 cDNA library using cDNA B2 fragment [1] as probe. Positive clones were partially sequenced and four distinct *tuf* related transcripts were identified. Sequence alignment revealed that these clones could be subdivided into two closely related subfamilies confirming our previous results [1]. In Fig. 3 we aligned the four partial gene sequences covering about 52% of the C-terminal coding part and the entire 3′ terminal non-translated region. Two of the four cDNA sequences are identical with either the genomic *tuf*A1 or *tuf*B1 sequence while the other two belong to either the A or B subfamily. Sequence identities are 92–94% between members of a subfamily but only about 78% between the subfamilies. However, the regions between stop codon and poly(A) start

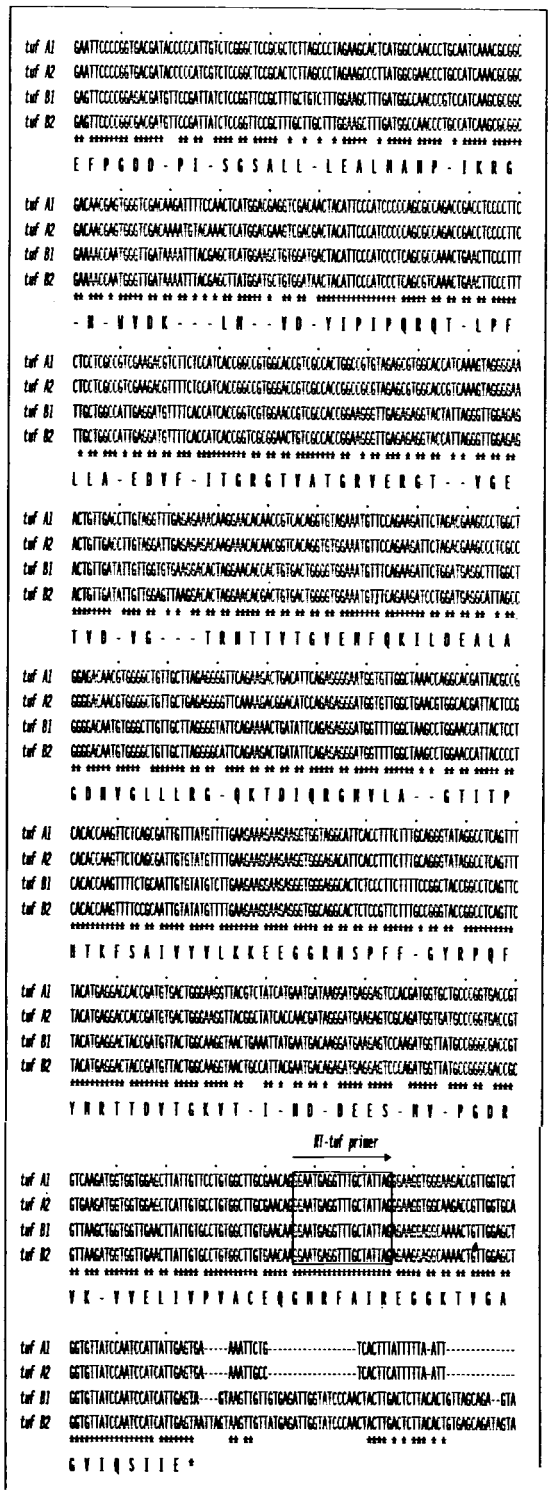

Fig. 3. Comparison of four partial soybean *tuf* gene sequences. Asterisks mark nucleotide sequence identity. The first decoded amino acid corresponds to position 161 of the mature EF-Tu. Conserved amino acids are given (one letter code). Stop codons are in bold letters. Positions (boxes) and polarity (arrows) of the primers used in the RT-PCR experiments are given. Poly(A) sites are shown and putative polyadenylation motifs are underlined.

show a very pronounced sequence divergence between the subfamilies. Multiple poly(A) start sites were identified and putative polyadenylation signals are marked.

Amino acids which are conserved in all four *tuf* genes are given in the bottom line. Amino acid sequence identity is about 92% between EF-Tu (A1) and EF-Tu (B1).

3.2. *Expression of tuf*A and *tuf*B genes

The *tuf*A and *tuf*B genes have distinct 3′ terminal sequences allowing monitoring of the relative steady state concentration of the *tuf* mRNAs under strictly comparable conditions [12]. The analytical protocol allowed us to discriminate between the two *tuf*B genes (amplified B1 and B2 fragments are 173 and 179 bp, respectively, see Fig. 3) but not between the two *tuf*A genes (amplified fragments A1 and A2 are both 139 bp). In Fig. 4 the results from dark and light grown

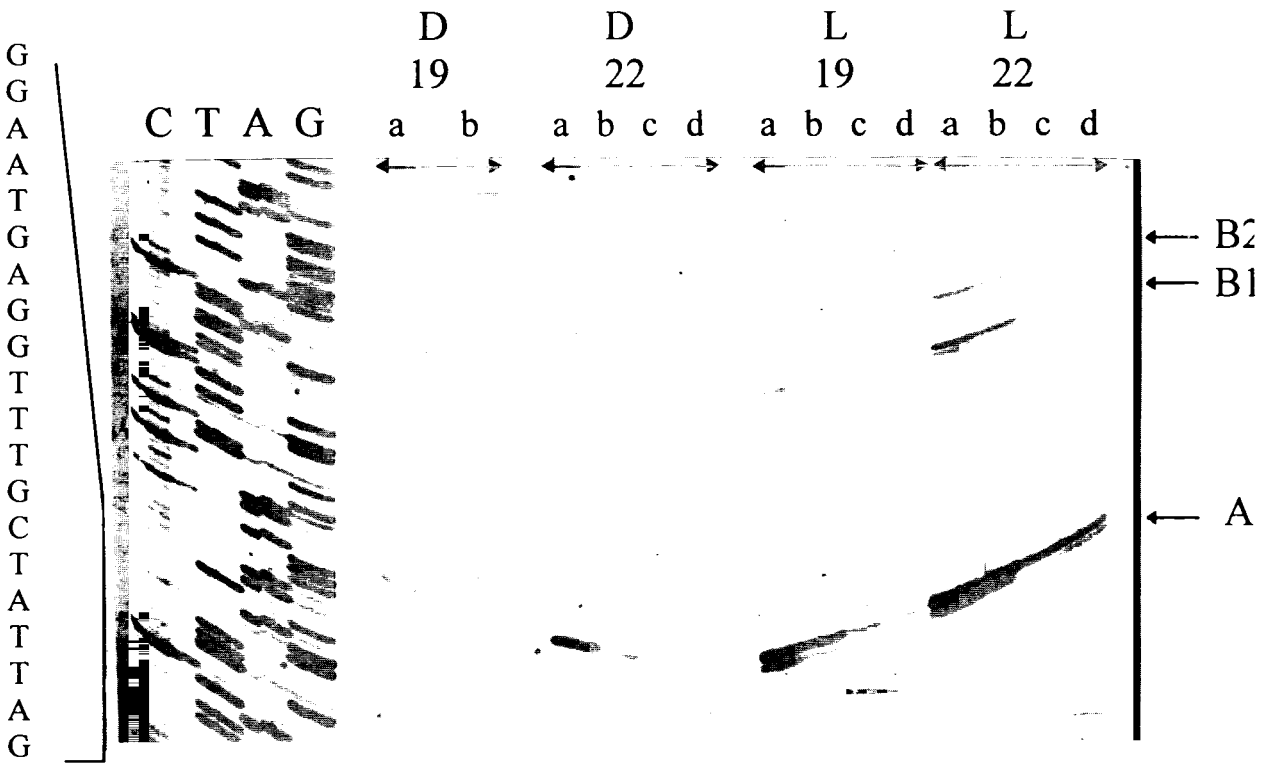

Fig. 4. *Tuf* gene transcription in dark and light grown whole soybean seedlings. Relative concentrations of RT-PCR products were analyzed after 19 or 22 PCR cycles, starting RNA concentrations were (in ng): lanes (a) 200, (b) 100, (c) 40, (d) 20. Arrows mark transcripts: (A) is *tuf*A1/A2 and (B1) (B2) are *tuf*B1, B2. Sequence ladder corresponds to the sequence of the *tuf*A1 gene using RT-Tuf A primer. Negative controls without reverse transcription prior to PCR amplification do not produce any detectable signals (data not shown).

seedlings are given. *Tuf*A and B genes are weakly expressed in the dark but after illumination *tuf*A mRNA concentration sharply increases. These results confirm previous Northern experiments which suggested that light leads to a higher steady state concentration of *tuf*A mRNA [1]. We did not test whether this increase was due to a higher transcription rate or higher transcript stability.

A similar RT-PCR experiment is shown in Fig. 5 using, respectively, total RNA extracts from leaves, stems and roots. (1) After 20 PCR cycles there is no signal in the root sample but bands are detected in the stem and leaf samples. (2) *Tuf*A mRNA predominates in leaf and stem tissues but *tuf*B1/B2 mRNAs slightly prevail in root tissue. (3) *Tuf* gene mRNA concentration is about 20 times lower in roots than in stems and leaves. The relative *tuf* mRNA concentrations as calculated from the scanned autoradiographs are summarized in Table 1.

3.3. Transcription start sites

Transcription start sites were determined by ribonuclease protection assays (Fig. 6). According to the length of the largest protected fragments (strongest bands, arrows) *tuf*A and B mRNAs start, respectively, 71 and 64 positions upstream of the start codon. The observed mapping patterns of *tuf*A and B are identical irrespective of growth conditions (T, 16) and probe length (*tuf*B). In line with the previous in situ transcription results, the

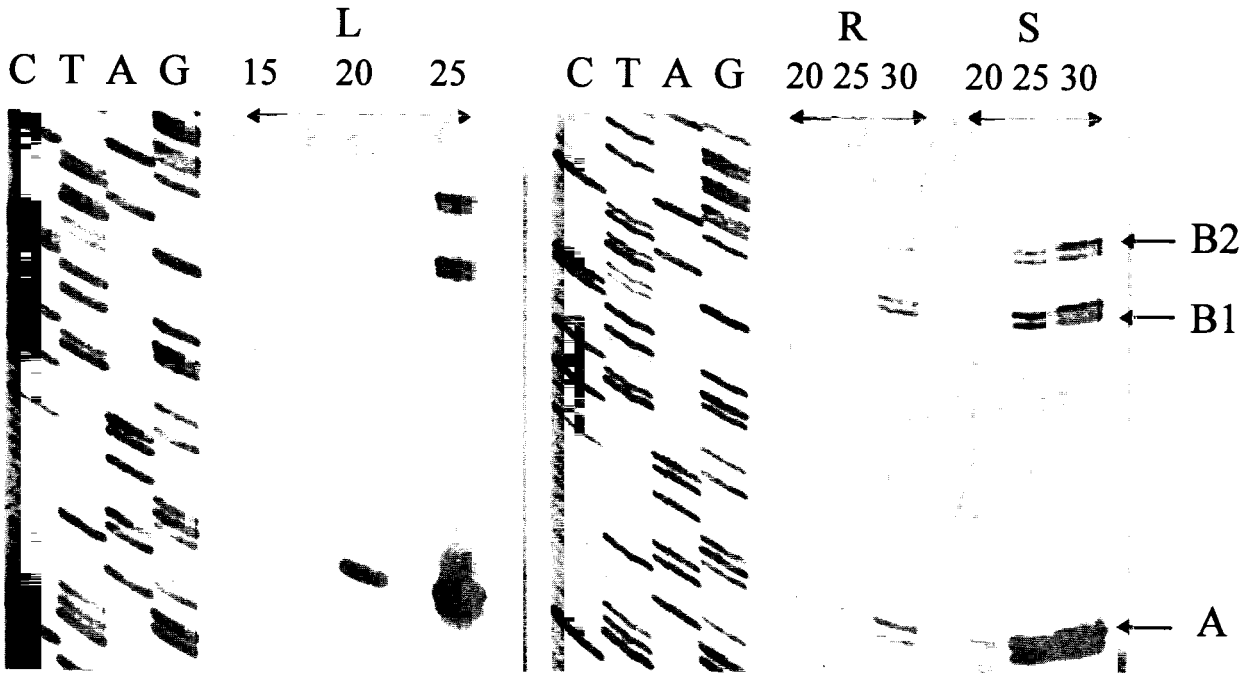

Fig. 5. *Tuf* gene transcription in leaves (L), roots (R), and stems (S). The amplified products were measured after 15, 20, 25 (leaves) or after 20, 25, 30 (stems, roots) PCR cycles. Symbols, sequence ladder and control are as in Fig. 4.

*tuf*A bands are much stronger than those of *tuf*B. The *tuf*A lanes show in addition to the top three strong bands (71, 69, 68 nucleotides) two strong bands in the range of 62, 60 nucleotides. Two relatively strong bands are seen in all four B lanes. Small variation in fragment length can be due to microheterogeneity of start sites and/or experimental artefacts. Both, the multiple longer (cluster of faint bands) and the more abundant shorter *tuf*A fragments could be due to sequence heterogeneity between members of the subfamilies leading to mismatched zones (note that the 5′ flanking regions of A2 and B2 were not sequenced). Corresponding shorter and longer relatively faint bands exist also in the B lanes.

3.4. Search for active promoter regions

The transcription studies reveal that the genes of the two *tuf* subfamilies are differentially expressed which can be due to differences in promoter structure. Indeed, the upstream regions of *tuf*A1 and *tuf*B1 (positions −1 to −500) have only 39% sequence similarity. We tested promoter strength in constructs with partially deleted promoter regions from the *tuf*A1 and B1 genes using transformed tobacco plants and GUS as a reporter gene.

The *tuf*A1 constructs were tested in transformed F1 tobacco plants and in transformed calli (ENA1) while the *tuf*B1 constructs were only tested in tobacco plants derived from transformed calli. We show in Fig. 7 the results obtained with 7–9 independent *tuf*A1 derived transformants. (1) Highest average GUS-activity is obtained with construct ENA1 with an insert of 2.2 kb; longer (HNA1) and shorter (XbNA1, XhNA1) promoter segments are less active in the average. (2) Activities vary considerably between individual transformants. (3) Activity is lost in all those transformants carrying the construct ANA1 which has an 80 bp insert. This suggests that the sequence from position −80 to −227 (Fig. 8) carries one

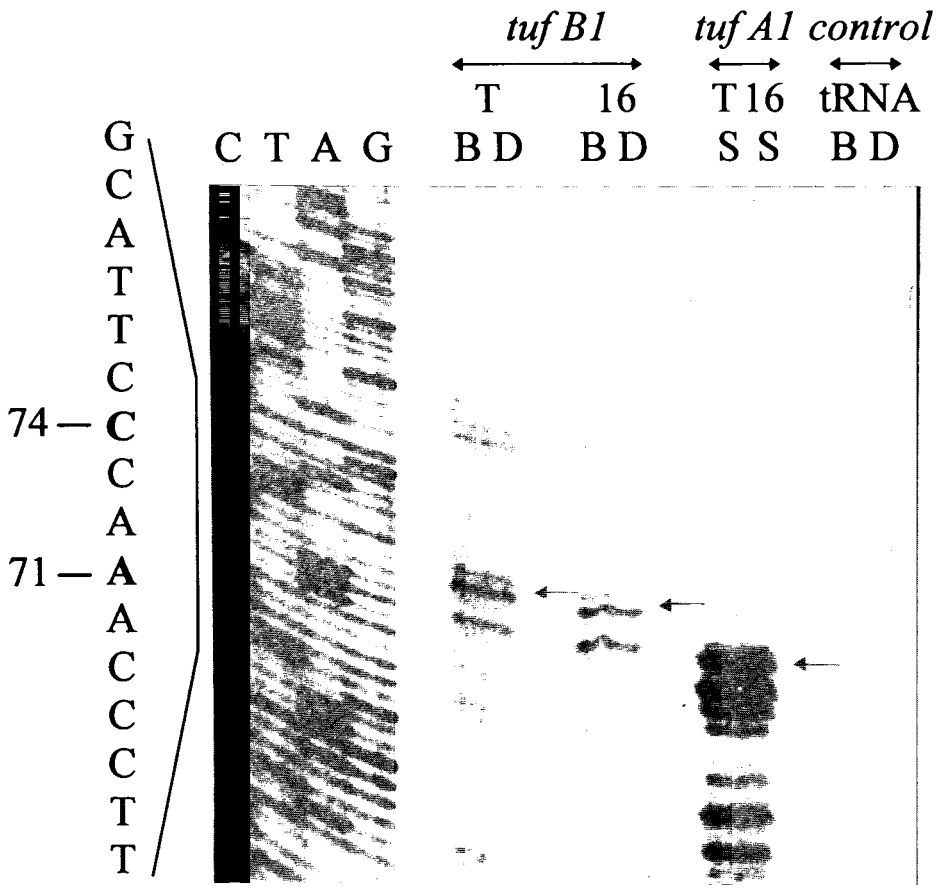

Fig. 6. Transcription start sites were determined by ribonuclease protection analysis using a single *tuf*A (S) and two different *tuf*B (B or D) ^{32}P-labeled antisense RNA probes (see Fig. 1). We tested, respectively, total RNA from light grown (T) or dark grown seedlings exposed to light for 16 h [16]. The size marker is given by the corresponding sequence of *tuf*B1 gene. The longest protected fragments are indicated by arrows. Length of fragments was determined from the sequence ladder and transcription start sites with respect to the start codon for A1 and B1 were calculated knowing the 3′ end of the antisense RNA probe. Yeast tRNA was used in the negative controls.

(several) *cis*-element(s) essential for promoter activity.

We tested this and some other upstream fragments for factor binding capacity (gel shift experiments) using bean nuclear extracts (Ph.D. thesis of F.M, 1995)[1]. The −80 to −227 region definitely interacted with nuclear proteins (data not shown), however, cross-competition experiments using sub-fragments of this region did not yield results allowing to closer delimit factor binding domains.

[1]F. Maurer, Structure et Fonction d'uine Famille de gènes nucleáirs codant pour le facteur d'elongation EF-Tu chloroplastique chez le soja (glycine max.). Thèse de doctorat, Université de Neuchâtel, Faculté des Sciences (1995).

The promoter activity tests with *tuf*B1 constructs (transformed calli) were less clear (not shown) and can be summarized as follows: (1) plants transformed with the constructs SNB1 (1.1 kb insert) or KNB1 (0.95 kb insert) showed definite, but low GUS-activity. However, GUS-

Table 1
Relative steady state concentrations (%) of *tuf* mRNA

	*tuf*A	*tuf*B1	*tuf*B2
Leaves	80	12	8
Stems	75	15	10
Roots	45	35	20

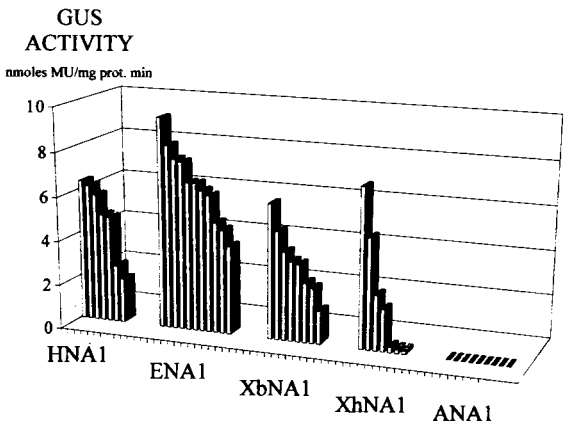

Fig. 7. GUS-activities (nmol MU/mg protein/min) in extracts of transgenic tobacco carrying *tuf*A1 derived promoters. Hundred light grown seedlings (F1) of each independent transformant were pooled and assayed.

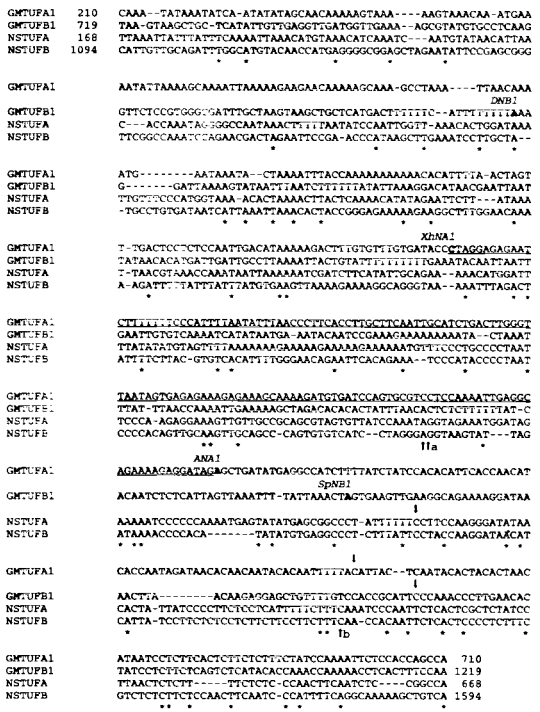

Fig. 8. Sequence alignment (CUSTALV) of 500 positions upstream of the start codon of the soybean (*G. max*) and tobacco (*N. sylvestris*) *tuf* genes. Top and bottom numbers correspond to sequences deposited in the database under accession numbers: X66062 (GMTUFA1); X89058 (GMTUFB1); D11469 (NSTUFA); D11470 (NSTUFB). Fragments fused to GUS reporter gene (Fig. 2) are indicated with the first base in bold. Transcription start sites are marked by vertical arrows; two start sites (a,b) are shown for NSTUFB [3]. The XhNA1-ANA1 fragment is underlined. The first position of the start codon is in bold; asterisks mark common sequences.

activities of these constructs were much lower than those of control samples containing the ENA1 construct (transformed calli). (2) Plants containing the constructs BNB1, DNB1 and SpNB1 did not yield a clear result, i.e. GUS activity was in any case very low prohibiting an evaluation comparable to the one given in Fig. 7 for the *tuf*A1 gene. The conclusion is that the *tuf*B1 upstream region (up to 1.1 kb) displays much weaker promoter activity in transformed tobacco than the corresponding *tuf*A1 segment. This result correlates positively with the in situ transcription data.

4. Discussion

4.1. Tuf gene structure

We have shown by Southern experiments [1] using two different soybean cDNA probes, coding both for the C-terminal part of EF-Tu (about 500 bp), that the nuclear genome contains four *tuf* genes. The cDNA sequencing studies of this report allowed grouping of the four genes into two subfamilies A1, A2 and B1, B2. Only two of the four genes were isolated and sequenced: the *tuf*A1 gene was isolated from a 6 kb *Hin*dIII fragment [1] and *tuf*B1 from a 7.2 kb *Eco*RI fragment. The two fragments did not overlap (restriction site analysis) and none of these fragments contained additional *tuf* gene sequences meaning that the *tuf* genes are at least 6–7 kb apart. Soybean is an amphidiploid plant ($2n = 40$) originating from two progenitors each contributing an equal number of chromosomes [21]. It is very likely that the two progenitors already had two *tuf* genes (A and B type) as seen for example in *N. sylvestris* [3]. Duplication of each type of *tuf* gene may have occurred after fusion of the progenitors.

The export of *tuf* genes to and successful integration into the nuclear genome occurred relatively late, most likely about 400–450 million years

ago within the charophycean lineage before the emergence of land plants [5]. The soybean EF-Tus have still a high amino acid sequence identity with algal [22,23] chloroplast encoded EF-Tus (70–80%), a very high sequence identity with higher plant EF-Tus (range of 90%) but only 52% with the functionally analogous cytosolic soybean EF-1α [24].

The *tuf* cDNA analysis allowed the identification of several poly(A) start sites. Poly(A) start sites vary from gene to gene and multiple start sites were found in a single gene (e.g. *tuf*B2, Fig. 3). Flexibility of plant polyadenylation is well-documented, (e.g. [25,26]) but the mechanisms of cleavage-polyadenylation site recognition and possible impact on mRNA stability are not understood in detail. Particularly, plants lack conserved *cis* elements near the cleavage site and near-upstream elements are poorly defined [27,28]. We underlined in Fig. 3 putative polyadenylation signals near the sequenced poly(A) tracts.

4.2. Transcription of tuf genes

Sugita et al. [3] reported that the *tuf*B mRNAs of *N. sylvestris* have three distinct transcription start sites which are located, respectively, 219, 169 and 72 bases upstream of the start codon while a single start site was found in case of *tuf*A (position −118). In soybean the untranslated leaders, as calculated from the longest protected bands, are about 71 (A1) and 64 (B1) nucleotides long. The cluster of major fragments varying in length from 71 down to 60 nucleotides (*tuf*A) can be the result of minor start site differences, however, we can not exclude that the smaller and larger protected fragments are due to hybrid mismatches, i.e. the 5′ flanking regions of *tuf*A1 and A2 may have slightly diverged.

The soybean *tuf*A genes correspond to the tobacco (*N. sylvestris*) *tuf*B gene with respect to gene activity (high activity in leaves) while the subfamily B rather behaves like the tobacco A gene (the gene nomenclature is not yet cleared). To trace possible phylogenetic relationships between functionally similar soybean and *N. sylvestris tuf* gene 5′ flanking regions we aligned and compared (Fig. 8) the sequences (position −1 to −500 upstream of the start codon) from soybean and tobacco using the program ALIGN (CUSTALV). Neither the alignment nor an analysis by the neighbour-joining method [29] did yield a statistically significant grouping. As a corollary of this study we found that the transcribed 5′ flanking region of the *N. tabacum tuf* gene [4] matches perfectly that of the *tuf*A but not of the B gene of *N. sylvestris*.

Our data as well as those from *N. sylvestris* show that the two types of *tuf* genes yield the same gene product (EF-Tu, 92–94% sequence identity, all functional domains intact) but the genes respond differently to light and are differentially expressed in leaves, stems and roots. EF-Tu is an essential factor for chloroplast protein synthesis and one would expect an increase in the steady state concentration of *tuf* mRNAs in green tissues. The in situ experiments show that the sharp increase is mainly due to the *tuf*A mRNAs (80% of *tuf* mRNA). The question arises whether the *tuf*A promoter contains light responsive and/or leaf specific *cis* elements not seen in the *tuf*B 5′ flanking region. We searched for light responsive elements [30] upstream of the transcription start sites of both the soybean *tuf*A1 and the functionally analogous *N. sylvestris tuf*B gene, but we could not detect any significant common domains indicative for light responsive elements.

According to the GUS-tests with transformed tobacco, deletion of the DNA segment XhNA1-ANA1 results in complete loss of promoter activity, suggesting that this segment contains essential target sequences for transcription factors, while the sequences further upstream enhance the average transcription activity. We searched for an XhNa1-ANA1-like segment within the 5′ flanking regions of the other three *tuf* genes listed in Fig. 8, however without success.

Acknowledgements

We thank Drs R. Medcalf, Melbourne and J.-M. Neuhaus, Neuchâtel, for critically reading the manuscript and for helpful suggestions. We thank Ch. Bonny for his help to initiate this study. This work is part of the doctoral thesis of Fabienne Maurer and received support from the Fonds

national suisse de la recherche scientifique (to E.S.).

References

[1] C. Bonny and E. Stutz, Soybean (*Glycine max* L.) nuclear DNA contains four *tuf*A genes coding for the chloroplast-specific translation elongation factor EF-Tu. Chimia, 47 (1993) 247–249.

[2] Y. Murayama, T. Matsubayashi, M. Sugita and M. Sugiura, Purification of chloroplast elongation factor Tu and cDNA analysis in tobacco: the existence of two chloroplast elongation factor Tu species. Plant Mol. Biol., 22 (1993) 767–774.

[3] M. Sugita, Y. Murayama and M. Sugiura, Structure and differential expression of two distinct genes encoding chloroplast elongation factor Tu in tobacco. Curr. Genet., 25 (1994) 164–168.

[4] V.M. Ursin, C.K. Becker and C.K. Shewmaker, Cloning and nucleotide sequence of a tobacco chloroplast translational elongation factor, EF-Tu. Plant Physiol., 101 (1993) 333–334.

[5] S. Baldauf and J. Palmer, Evolutionary transfer of the chloroplast *tuf*A gene to the nucleus. Nature, 344 (1990) 262–265.

[6] S.L. Baldauf, J.R. Manhart and J.D. Palmer, Different fates of the chloroplast *tuf*A gene following its transfer to the nucleus in green algae. Proc. Natl. Acad. Sci. USA, 87 (1990) 5317–5321.

[7] A. Schori, A. Fossati, A. Soldati and P. Stamp, Cold tolerance in soybean (*Glycine max* L. Merr.) in relation to flowering habit, pod set and compensation for lost reproductive organs. Eur. J. Agron., 2 (1993) 173–178.

[8] N. Fedoroff, S. Wessler and M. Shure, Isolation of the transposable maize controlling elements Ac and Ds. Cell, 35 (1983) 235–242.

[9] J. Sambrook, E. Fritsch and T. Maniatis, Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, New York, 2nd edn., 1989.

[10] C. Dean, P. Van den Elzen, S. Tamaki, P. Dunsmuir and J. Bedbrook, Differential expression of the eight genes of the petunia ribulose bisphosphate carboxylase small subunit multi-gene family. EMBO J., 4 (1985) 3055–3061.

[11] T.V. Huynh, R.A. Young and R.W. Davis, Constructing and screening of cDNA libraries in lambda gt10 and lambda gt11, in: Glover D.M. (Ed.), DNA Cloning, a Practical Approach, IRL Press Limited, Oxford, 1987, pp. 49–78.

[12] C. Simpson, T. Sawbridge, G. Jenkins and J. Brown, Expression analysis of multigene families by RT-PCR. Nucl. Acids Res., 20 (1992) 5861–5862.

[13] T.E. Golde, S. Estus, M. Usiak, L.H. Younkin and S.G. Younkin, Expression of beta amyloid protein precursor mRNAs: recognition of a novel alternatively spliced form and quantification in Alzheimer's disease using PCR. Neuron, 4 (1990) 253–267.

[14] A. Hoekema, P.R. Hirsch, P.J.J. Hooykaas and R.A. Schilperoort, A binary plant vector strategy based on separation of vir- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. Nature, 303 (1983) 179–180.

[15] D.H. Figurski and D.R. Helinski, Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. Proc. Natl. Acad. Sci. USA, 76 (1979) 1648–1652.

[16] R.B. Horsch, J.E. Fry, N.L. Hoffmann, D. Eicholtz, S.G. Rogers and R.T. Fraley, A simple and general method for transferring genes into plants. Science, 227 (1985) 1229–1231.

[17] T. Murashige and F. Skoog, A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant, 15 (1962) 473–497.

[18] R. Jefferson, Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol. Biol. Rep., 5 (1987) 387–405.

[19] R. Jefferson, T. Kavanagh and M. Bevan, GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J., 6 (1987) 3901–3907.

[20] M. Bradford, A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of the protein-dye binding. Anal. Biochem., 72 (1976) 248.

[21] J.A. Lackey, Chromosome numbers in the Phaseoleae (Fabaceae:Faboideae) and their relation to taxonomy. Am. J. Bot., 67 (1980) 595–602.

[22] P.E. Montandon and E. Stutz, Nucleotide sequence of a *Euglena gracilis* chloroplast genome region coding for the elongation factor Tu; evidence for a split mRNA. Nucl. Acids Res., 13 (1983) 5877–5892.

[23] J.C. Watson and S.J. Surzycki, Extensive sequence homology in the DNA coding for elongation factor Tu from *Escherichia coli* and the *Chlamydomonas Reinhardtii* chloroplast. Proc. Natl. Acad. Sci. USA, 79 (1982) 2264–2267.

[24] F. Aguilar, P.E. Montandon and E. Stutz, Two genes encoding the soybean translation elongation factor eEF-1α are transcribed in seedlings leaves. Plant Mol. Biol., 17 (1991) 351–360.

[25] C. Dean, S. Tamaki, P. Dunsmuir, M. Favreau, C. Katayama, H. Dooner and J. Bedbrook, mRNA transcripts of several plant genes are polydenylated at multiple sites in vivo. Nucl. Acids Res., 14 (1986) 2229–2241.

[26] J.F. Manen and P. Simon, A possible explanation for the multiple polyadenylation sites in transcripts coding for a winged-bean leghemoglobin. Planta, 191 (1993) 289–292.

[27] E. Wahle and W. Keller, The biochemistry of 3′-end cleavage and polyadenylation of messenger RNA precursors. Annu. Rev. Biochem., 61 (1992) 419–440.

[28] A.G. Hunt, Messenger RNA 3′ end formation in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol., 45 (1994) 47–60.

[29] N. Saitou and M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol., 4 (1987) 406–425.

[30] R. Fluhr, C. Kuhlemeier, F. Nagy and N. Chua, Organ-specific and light-induced expression of plant genes. Science, 232 (1986) 1106–1111.