

Isolation, characterization and expression of a gene coding for a 2S albumin from *Bertholletia excelsa* (Brazil nut)

E. S. Gander, K.-O. Holmstroem, G. R. De Paiva, L. A. B. De Castro, M. Carneiro and M.-F. Grossi de Sá

Embrapa/Cenargen, S.A.I.N., Parque Rural, 70.770 Brasilia-DF, Brazil

Received 20 August 1990; accepted in revised form 7 November 1990

Key words: *Bertholletia excelsa*, 2S albumin genes, gene regulation, methionine-rich storage protein, promoter

Abstract

Two genes, BE2S1 and BE2S2, coding for methionine-rich albumins of Brazil nut (*Bertholletia excelsa* H.B.K.) have been cloned and their sequence determined. The genes are members of a multigene family and one of them, i.e. BE2S1, codes for one of the dominant 2S isoforms. Its expression is highly regulated during seed development and with respect to tissue specificity. Sequence analysis has shown that the genes contain one intron and that the promoter of BE2S1 shows a canonical TATA motif. The transcription initiation site is located 26 nucleotides downstream from the TATA box. Sequence comparison of the promoter regions of 2S genes from *Brassica napus*, *Arabidopsis thaliana* and *B. excelsa* revealed the presence of TGCA palindromic sequences that appear to be arranged in a 2S-specific manner.

Introduction

The biosynthesis of plant storage proteins offers an interesting model system for studies concerning the regulation of gene expression during development and with respect to organ specificity [see reviews in 10, 12].

There are multiple classes of seed storage proteins; the 2S albumins are prominent among these in dicotyledons. In several species the genes coding for 2S albumins have been characterized; in the case of *B. napus* and *A. thaliana* they are encoded by multigene families [13, 14] and contain two subunits joined by disulphide

bridges. We are interested in a family of 2S albumins which are found abundantly in seeds of *B. excelsa* (Brazil nut). The amino acid composition [26], the sequences of various isoforms of Brazil nut 2S albumins and the nucleotide sequences of cDNAs coding for two isoforms have been described [1, 2, 6]. On the basis of these results it was concluded that the primary transcripts of these genes, like those of other 2S albumins, code for prepropolypeptides and that complex events at the translational and/or posttranslational level are required for the formation, transport and storage of the mature protein. The amino acid sequence of 2S albumin precursors from

various species have been compared [14]. The one from *B. excelsa* diverges substantially from other 2S species: the 2S albumins from *B. napus* and *A. thaliana* display homologies of 86 and 66% in their processed and mature region while a comparison of the corresponding regions of *B. excelsa* albumins with those of *A. thaliana* reveals homologies of only 22 and 24% respectively. However, the position of the cysteine residues in all 2S albumins described are conserved, suggesting that these residues are responsible for a similar tertiary structure.

Unlike the 2S albumins found in *Ricinus* [23], *Brassica* and *Arabidopsis*, the amino acid analysis of Brazil nut 2S albumins revealed the presence of 18% methionine residues. Amino acid and nucleotide sequence analysis subsequently showed that the methionine residues occur in two clusters in a domain of the large subunit between the 4th and 5th cysteine residue, the region least conserved in 2S sequences from different species [6, 14]. This is the only intercysteine stretch showing significant length variation between different species, supporting modelling studies suggesting that this region is located outside the domains critical for the stability of the secondary and tertiary structure (G. Neshich, pers. commun.).

The fact that the Brazil nut 2S albumins have a high methionine content make them interesting candidates for correction of methionine-deficient crop plants via recombinant DNA technology. Studies involving the introduction and expression of *B. excelsa* 2S structural sequences in host plants have demonstrated that the expression of Brazil nut cDNA using the phaseolin promoter in tobacco can increase the methionine concentration in tobacco seeds [1]. Similar results have been obtained using a 2S promoter from *A. thaliana* (De Clerq, pers. commun.) and Guerche *et al.* [9] have expressed a Brazil nut albumin in *Brassica* using the soybean lectin promoter.

However, all these studies involve exclusively the Brazil nut coding regions and thus far no promoter or entire 2S gene from *B. excelsa* has been analysed.

In this work the sequence of two Brazil nut 2S

genomic clones including their promoter regions are presented. In addition, along with the determination of the transcription initiation site, the developmental profile and the tissue-specific expression of one of these genes is demonstrated.

Materials and methods

Plant material

Brazil nut seeds and leaves were obtained from the 'Centro de Pesquisa Agropecuária do Tropicó Humido (CPATU/EMBRAPA)' in Belém, Pará, Brazil during the first quarter of the year. During this period embryos covering the entire developmental spectrum can be obtained. The material was stored at -80°C or used immediately.

Genomic library and DNA sequencing

A genomic library was established in the cloning vector Charon 35 using *Sau*3A partially digested Brazil nut seed DNA (13–15 kb fraction). The calf intestine phosphatase-treated fragments were ligated into *Bam*HI-created Charon 35 arms according to standard procedures [15]. The phages were plated out and transferred to nitrocellulose filters; recombinant phages were detected using radiolabelled nearly full-length cDNA [6]. DNA sequencing was performed using the chemical degradation procedure [16].

Northern analysis of total seed and leaf RNA

Total RNA from seeds at different developmental stages and from leaves was prepared using the CTAB method [24]. After the analysis of the samples on 1.5% agarose-glyoxal gels [17] and transfer to nitrocellulose filters, the blots were probed with 5×10^6 CPM of radiolabelled 2S cDNA.

S1 nuclease mapping

RNA (50 μg /reaction) from seeds and leaves was hybridized to labelled probes (8×10^4 CPM/

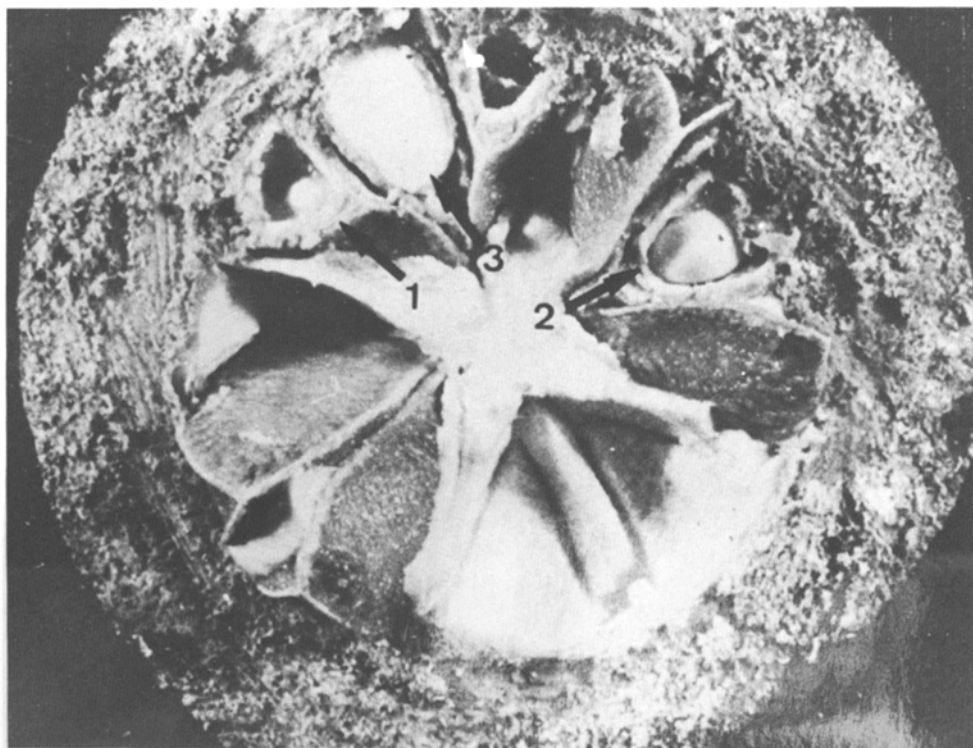
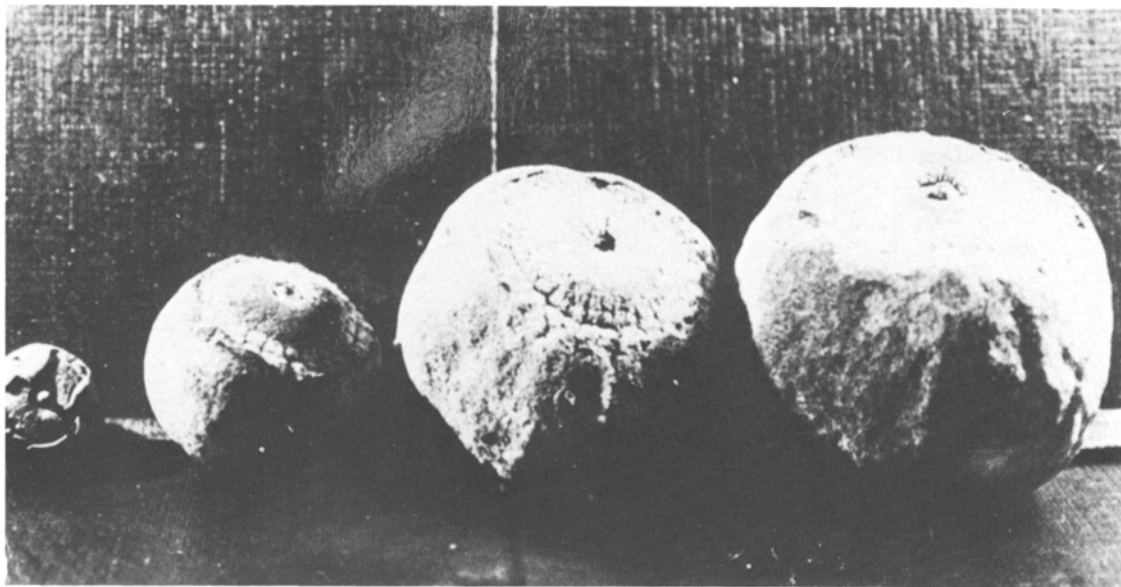


Fig. 1. Fruits and seeds of *B. excelsa*. a (top). Fruits at various developmental stages. b (bottom). Cross-section through an apparently mature fruit.

reaction) in 30 μ l of a solution containing 50% formamide, 0.4 M NaCl, 0.04 M PIPES (pH 6.4) and 0.001 M EDTA [5] at 47 °C for 18 h. At this temperature RNA:DNA rather than DNA:DNA hybridization is favoured. The

DNA probe was prepared by end-labelling a DNA fragment from a Brazil nut gene with T4 polynucleotide kinase and ^{32} P-ATP. The samples were heated for 10 min at 85 °C in order to ensure denaturation prior to hybridization. After 18

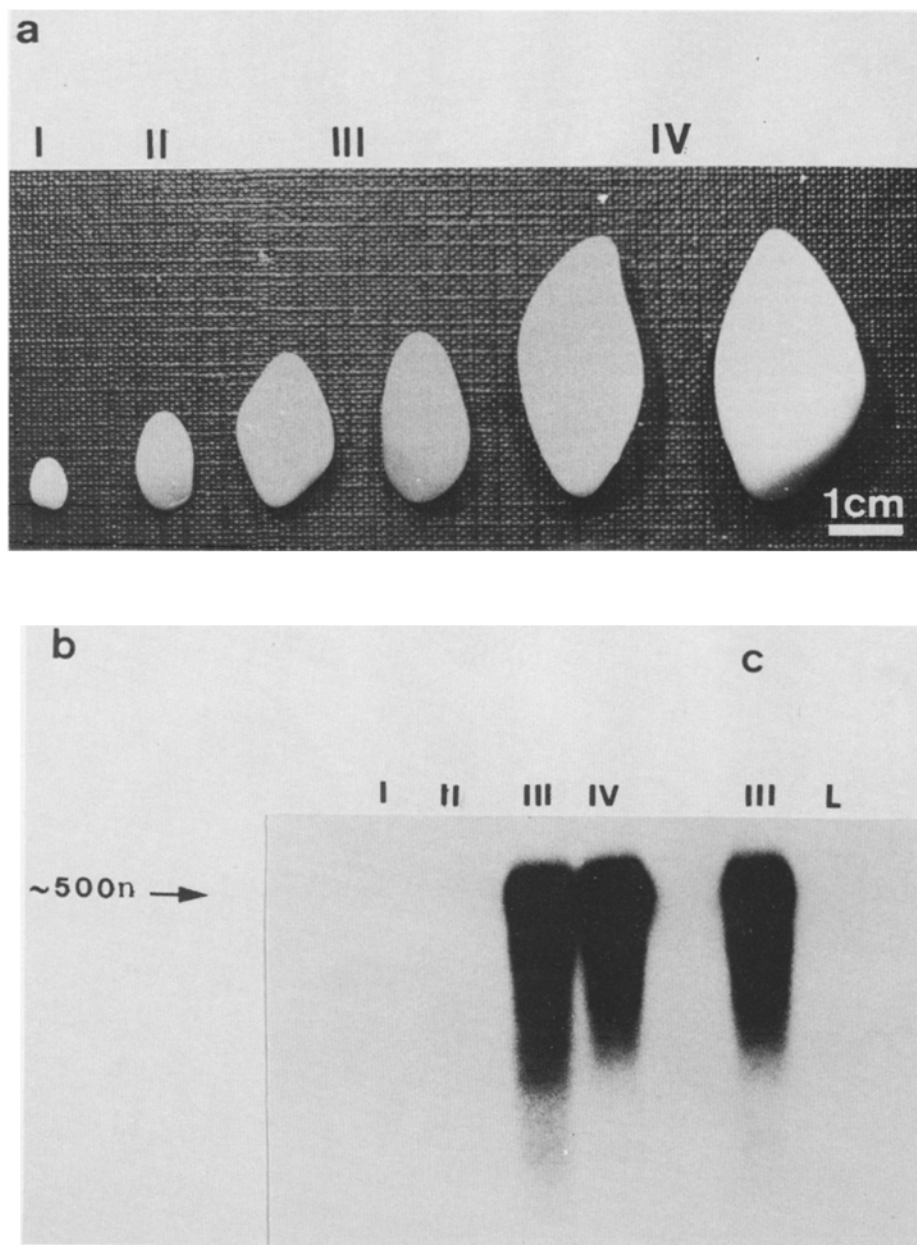


Fig. 2. Northern blot of *B. excelsa* seed and leaf RNA. Four seed developmental stages (a) and leaves were used for RNA extraction. After fractionation on glyoxal gels the RNA was transferred to Hybond N membranes (Amersham) and probed with labelled 2S cDNA (b and c). Columns I, II, III and IV represent RNA from stages I, II, III and IV, respectively and 'L' represents leaf RNA.

hours the samples were digested with 30 units of S1 nuclease [25] in 0.3 ml of S1 buffer [7] for 30 min at 37 °C. Protected DNA fragments were resolved on 8% urea-polyacrylamide gels [21].

Southern analysis

DNA from Brazil nut seeds (10 µg) was restricted with *Hind*III and *Eco*RI and sized on 0.8% agarose gels. The DNA was transferred to nitrocellulose and the filters were prehybridized for 3 h at 60 °C in 5 × SSC, 5 × Denhardt's. Hybridization was done against 3 × 10⁶ CPM of nick-translated, ³²P-labelled 2S cDNA in 5 × SSC, 5 × Denhardt's, 10% dextran sulphate and 0.1 mg ml⁻¹ salmon sperm DNA at 60 °C overnight. The filters were washed in 5 ×, 2 ×, and 0.1 × SSC containing 0.2% SDS at 60 °C exposed to Kodak X-ray films.

Results

Developmental and organ-specific expression of 2S genes

B. excelsa belongs to the family Lecythidaceae which includes about 200 neotropical species. It grows in the Amazon basin and can reach a height of 50 to 60 m. Brazil nut fruits take from 9 to 12 months to ripen. In Fig. 1a fruits at various developmental stages are shown. The seeds develop in groups of 12 to 25 in the horizontal plain of the fruit as is illustrated in a cross-section of a 'mature' fruit (Fig. 1b). From this figure it is also evident that seed development is not synchronized; among mature seeds (arrow 3), individuals at various developmental stages are encountered (arrows 1 and 2). The embryos belong to the macropodial type, i.e. no cotyledons are differentiated and almost the entire seed mass consists of hypocotylar tissue. In contrast to many other plants where either the endosperm or the cotyledons serve as storage organs, in *B. excelsa* this function is taken over by the hypocotyl.

Most plant storage protein genes are strictly regulated in time and in space, but over a much shorter period of development than that available to *B. excelsa*. Therefore it was of interest to evaluate at what stage of seed development the Brazil nut 2S albumin genes are expressed. Total RNA from seeds at different developmental stages (Fig. 2a) and from leaves was subjected to northern blot analysis (Fig. 2b) using as a probe nearly full-length cDNA coding for one of the 2S albumin isoforms [6].

It is evident that the expression of 2S albumin genes is restricted to seeds that are close to maturity, since a signal is detected only in RNA from stages III and IV; Fig. 2c demonstrates that the genes are not expressed in leaves. The size of the transcript, about 500 nucleotides, corresponds

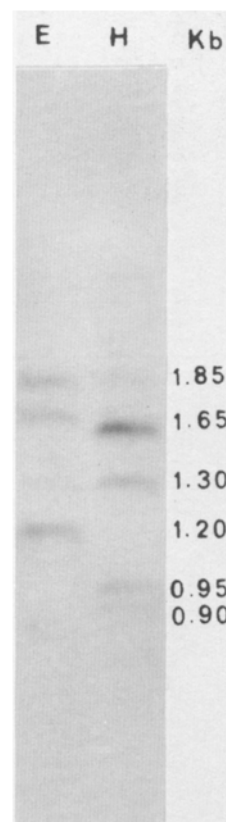


Fig. 3. Genomic Southern blot of *B. excelsa* DNA. *B. excelsa* DNA was digested with *Eco*RI (E) and *Hind*III (H) and sized on a 0.8% agarose gel together with digested lambda DNA as a size marker. Hybridization was against a labelled 2S cDNA.

Nucleotide sequence of gene BE2S1

Fig. 4. Nucleotide sequence of BE2S1 and BE2S2. Fig. 4a: BE2S1; Fig. 4b: BE2S2. The TATA box is underlined, transcription initiation site is marked with an asterisk; PRE: start of 'pre'-sequence, PRO: start of 'pro'-sequence; SS: small subunit, LS: large subunit; the sequence coding for IPF is boxed, methionine- and arginine-rich stretches are marked with dashed lines. A vertical

Fig.4b

Nucleotide sequence of gene BE2S2

```

GCCACCTCGCCTGCGCCTGTCCTTACCGCCCTCCTXGTACTTTCTCGTGC
AAAGTTGCCACCTCCCCCTGCTGTCCTTACAGTCATCCXGTACTTTC
CCATGCACAGTTGCTAGCCTGCAAAACGCACCATAAATAGAGATTCTCTC
CACCATCCTCTTCGCCCCACCACGGCTGCAGATCATACACCCACACCCAAG
TACAGCTAGAATCACATCACC PRE ATG GCA AAG ATG TCA GTT GTG
                        M A K M S V V
GCA GCA GCC CTT CTC GCT TC CTT GTC CTC GGC CAA GCC
A A A L L A L L V L G Q A
ACT GCC PRO TTC AGG ACC ACT GTC TC ACG ACG TTG GAG GAG
T A F R T T V T T T L E E
GAG SS CAG GAG GAG AAC CCC CGC GGA CGG TCC GAG CAG CAG
E Q E E N P R G R S E Q Q
TGC CGC GAG CAG ATG GAG AGG CAG CAG CAG CTG AAC CAC
C R E Q M E R Q Q Q L N H
TGC CGG ATG TAC TTA AGA CAG CAG GTTAATTATTGTGCAAGA
C R M Y L R Q Q
AAGTATTGAACTATTTTATATTTCCGAATTTAATGACG...TTTTTCCGT

CAAACGAACCCGGGTGCAG SS ATG GAG GAG AGC CCG TAC CAG AAT
                        M E E S P Y Q N
CCC CGG CCC LS CTG AGA CGG GGG GAG GAG CCG CAC CTG GAC
P R P L L R R G E E P H L D
GAG TGC TGT GAG CAG CTG GAG AGG ATG GAC GAG ATG TGC
E C C E Q L E R M D E M C
AGA TGT GAA GGC TTA AGG ATG ATG TTG AGG AGG CAA CGG
R C E G L R M M L P R Q R
GAG GAG ATG GAG CTC CAG GGC GAG CAG ATG CAG AGG ATA
E E M E L Q G E Q M Q R I
ATG AGG AAG GCC GAG AAC CTC CTC TCC CGC TGC AAC CTC
M R K A E N L L S R C N L
AGC CCA CAG AGA TGC CCC ATG GGT GGC TAC ACG GCC TGG
S P Q R C P M G G Y T A W
CTC TAA TCAGCTAGCCCTGCGCATATGTTAATAAATCACAGACCAACTACC
L •
TGTTTTGAATGCAGAGGTTAGGGTGAGATTAATAATAGTAATAAAGCACATTA
TCGTGTGTTTTCTGTCCAAGCTCCCCCCCCCCCCACCCTCTCTCGTTTCC
ACCATCTTGATGCCTTCATGAAAAAAAAAATTCTTTAATTACATTATAATT
TACTAAATGGAGAAAAAGAAGACTACGAGAGAAGCAAAACAGAAAAACGA

```

arrow denotes the end of the mature large subunit; stop codons are marked with dots and polyadenylation signals are pointed out. Horizontal arrows denote the start of coding regions.

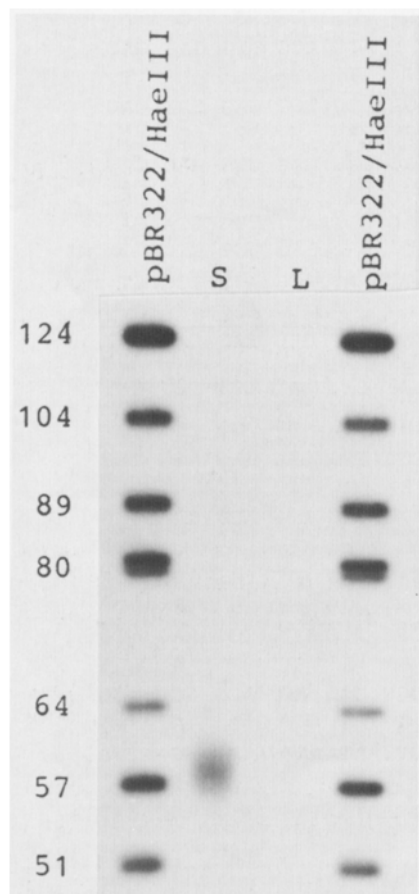
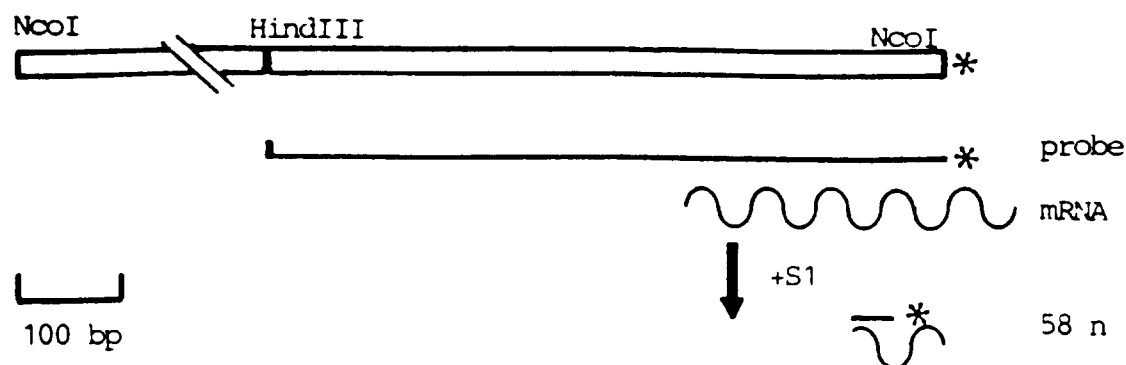
a.

Fig. 5. Nuclease S1 mapping of the transcription initiation site. The *Nco*I–*Nco*I fragment containing the entire available 2S promoter region of BE2S1 was end-labelled. The 1.6 kb *Hind*III–*Nco*I fragment was annealed, after denaturation, to 30 ng of either total seed or total leaf RNA. S1 nuclease-resistant DNA fragments were sized on a 8% polyacrylamide-urea gel using end-labelled pBR322/*Hae*III DNA fragments as size markers (b). Column S: 58 nucleotide fragment protected by total seed RNA. Column L: hybridization with total leaf RNA (= negative control).

well with mRNA coding for a prepolypeptide of 146 amino acid residues.

B. excelsa 2S proteins are encoded by a multigene family

In order to establish whether the 2S albumins of Brazil nuts are encoded by a multigene family,

Table 1. Comparison of 2S promotor regions with promoters from constitutively and light-regulated genes. The sequences were aligned according to their respective TATA boxes. Dots represent nucleotides; thus dotted stretches represent real distances. Unbroken lines indicate relative distances. Arabic numerals indicate the number of nucleotides; Roman numerals denote the domains where TGCA palindromes are accumulated. Sources: *AT2S1*: Krebbers *et al.*, 1988; *napA*: Josefsson *et al.*, 1987; *gNa*: Scofield and Crouch, 1987; *BE2S1* and *BE2S2*: this work; *EF-1αA1*: Axelos *et al.*, 1989; *EF-1αA4*: B. Lescure, pers. commun.; *rbcS-3A* and *rbcS-3C*: Fluhr *et al.*, 1986; G-box *rbcS* consensus: Giuliano *et al.*, 1988.

5'	IV	III	II	I	3'
<i>AT2S1</i> :.....GTTTTTG	282—TGCA...TGCA	94—TGCATGtA...ACGT	TGCA	16—TATAAAT	
<i>napA</i> :.....ACGTTTTTTTTT	75—ACGT.....TGCA	91—TGCATGCA...ACGT	TGCA	16—TATAAAT	
<i>gNa</i> :.....	TGCA	67—TGCATGCA...ACGT	TGCA	16—TATAAAT	
<i>BE2S1</i> :.....ACGTTTTTG...ACGT.TGCA	ACGT	48—ACGT...ACGTGCA		17—TAAATAT	
<i>BE2S2</i> :.....	TGCA—38	AxGT.....TGCA	TGCA—10—TAAATAG		
<i>EF-1αA1</i> :.....TGCA	283—			TATAAT	
<i>EF-1αA4</i> :.....TGCA				TATATT	
<i>rbcS-3C</i> :.....TGCA	77—ACGT.....	TGCA		TAATAT	
G-box <i>rbcS</i> :	TCTTACGTGGCAPPyPy				

Southern blots with *Hind* III and *Eco* RI restricted genomic DNA from *B. excelsa* were performed (Fig. 3). Although it is unknown if these enzymes cut within other genes (the genes described below have neither *Eco* RI nor *Hind* III sites), it is reasonable to conclude from the hybridization pattern that at least 5 genes coding for 2S albumins are present in the genome of *B. excelsa*.

Cloning and sequence analysis of genes coding for B. excelsa 2S albumins

A genomic library constructed in the vector Charon 35 was screened using the nearly complete cDNA described by De Castro *et al.* [6]. Of 20 initially positive clones, two remained after high-stringency washing. These were subcloned into pUC18 and sequenced.

One of these, BE2S1 (Fig. 4a), is 100% homologous to an abundant cDNA [9] while the other, BE2S2 (Fig. 4b), shows a sequence homology of 67%. As expected, both genes encode a prepolypeptide consisting of a small (SS) and a large (LS) subunit joined together by a peptide of five (BE2S1) and seven (BE2S2) amino acids that are post- or co-translationally processed (= internal processed fragment). In addition the previously observed amino and carboxyl terminal processed fragments are present [14]. In both genes the coding region for the small subunit is interrupted by an intron of 89 and 87 nucleotides respectively. The introns conform to the GT-AG rule and in both a branch site (TTTAAT) starting 47 nucleotides downstream from the left junction is present. Within the sequence coding for the large subunit of BE2S1 two regions are remarkable (underlined): they consist of 7 and 6 triplets respectively coding for either methionine or arginine. In gene BE2S2 these regions, though identifiable, are not particularly enriched in methionine and in fact, while in BE2S1 the methionine concentration reaches 18%, in BE2S2 only 10% of the amino acids are methionines.

Determination of the site of transcriptional initiation

In order to determine the transcriptional initiation site, nuclease S1 mapping was performed using as probe an end-labelled coding strand of 1636 nucleotides spanning the 5' upstream region starting at the initiation codon of the 2S albumin open reading frame (Fig. 5). Seed and leaf total RNA were hybridized to the probe under conditions favouring RNA:DNA over DNA:DNA hybridization and subsequently treated with S1 nuclease. The resulting protected fragment, not detected when leaf RNA was used, was 58 bp long, demonstrating that transcription starts 55 nucleotides upstream of the initiation codon (Fig. 4a, asterisk). In gene BE2S1 a canonical TATA box is located 25 bp upstream from this position (Fig. 4a, underlined); in gene BE2S2 this sequence is slightly modified, i.e. from TAAATAT to TAAATAG. Further particularities of the promoter regions as well as a comparison with the promoters of 2S genes from other species will be dealt with in the discussion and in Table 1.

Discussion

In this paper we present data on the isolation and characterization of two genes, BE2S1 and BE2S2, from *B. excelsa* that code for 2S albumins. They belong to a multigene family responsible for the various 2S isoforms encountered in *B. excelsa* seeds. One of them, BE2S1, codes for an isoform with a methionine content of 18%. This gene shows 100% homology with a cDNA prepared from an abundant mRNA species [9] suggesting that it is actively transcribed. Southern blots, however, indicate that at least four other genes are present. One of them, BE2S2, has also been isolated and characterized. It codes for a putative 2S albumin with a methionine content of only 10% and its sequence homology with the cDNA is 67%. No stop codons interrupt its coding regions indicating that BE2S2 could be transcribed and translated. However, so far no isoform corresponding to its sequence has been described [2].

The degree of homology in the coding region of various 2S genes has been discussed elsewhere [14]; it has also been pointed out that 2S genes from *B. napus* and *A. thaliana* contain no introns [13, 14, 22]. In contrast to these genes, both *B. excelsa* 2S genes described here have an intervening sequence located in the coding region of the small subunit.

Northern analysis (Fig. 2) clearly shows that 2S mRNA accumulates in substantial amounts only in mature seeds. Not even traces of these molecules can be demonstrated in preparations of total leaf RNA. This is an indication that the expression of Brazil nut 2S genes, like that of 2S genes from other species, is highly regulated with respect to organ specificity and during seed development. Such specificity of expression is presumably due to sequences in promoter and/or enhancer regions involved with positive or negative control of transcription through interaction with specific DNA-binding proteins. One such sequence, the G-box of a gene coding for ribulose-1,5-biphosphate carboxylase/oxygenase, centres around the palindrome TGCA/ACGT [11]. Such palindromes are also present in the promoter region of *S. napus* and *A. thaliana* 2S genes [13, 14] as well as in all other 2S genes examined (Table 1). Upon alignment with the TATA box, four TGCA domains stand out (Table 1); they are located between 10 to 16 (domain I), 25 to 28 (domain II) and 38 to 40 (domain III) nucleotides upstream from the TATA box. In all genes a stretch of at least 38 to at most 94 nucleotides separates domain III from the next region, domain IV, where TGCA motifs again occur with high frequency. The first three domains are located at spatially the same or nearly the same position in all 2S genes described. Interestingly, gene BE2S1 domain I is absent. Finally, at variable distances in the various promoter regions a pyrimidine stretch is close to (in AT2S1) or associated with a TGCA motif.

For comparison, we have also analysed the promoter regions of two constitutively expressed genes coding for protein elongation factors of *A. thaliana* (EF1 α -A1 and EF1 α -A4) and two light-regulated genes (*rbcS-3A* and *rbcS-3C*). In

all these promoters the situation is drastically different: the TGCA palindromes occur with no greater frequency than could be expected statistically ($\text{Pr}_{\text{TGCA}} = 0.003$) (Table 1).

In contrast to that finding in constitutively and light-regulated promoters, a preliminary survey of promoter regions of other storage protein genes, i.e. patatins [19, 20] and G1-glycin and β -conglycinin 4 (R. B. Goldberg, pers. commun.) reveals the presence of TGCA sequences at elevated frequencies. However, no spacing pattern comparable to that found in the 2S promoters could be discerned.

It appears that the massed occurrence of TGCA palindromes in the promoter regions of storage proteins is a common feature and the tropical spacing pattern described here an intrinsic characteristic of genes coding for 2S albumins.

As mentioned in the introduction, the mature 2S proteins from the phylogenetically closely related *Arabidopsis* and *Brassica* (order: Capparales) display a homology of about 70%. On the other hand, the evolutionarily distant 2S proteins from *B. excelsa* (order: Lecythidales) and *R. communis* (order: Euphorbiales) are only 24 and 36% homologous with the proteins of the other group [20]. We believe therefore that the close to 100% homology with respect to the TGCA spacing in all 2S promoters cannot be explained with a mere phylogenetic relationship but rather reflects the functional homology of these sequences.

Ongoing functional tests in our lab are aimed to provide evidence as to the biological significance if any of these TGCA palindromes.

Acknowledgements

We thank Drs. E. Krebbers and C. R. Clements for critical reading of the manuscript and J. Seurinck for assistance in sequencing. K.-O. Holmstroem was supported by a grant from the Brazilian National Research Council (CNPq) and from the 'Sverige Jordbruks Forskningsråd' (SJFR). The project was financed in part by a World Bank loan to the Brazilian government.

References

- Altenbach SB, Pearson KW, Sun SSM: Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. *Plant Mol Biol* 8: 239–250 (1987).
- Ampe C, Van Damme J, De Castro LAB, Sampaio MJ, Van Montagu M, Vanderkerckhove J: The amino acid sequence of the 2S sulphur-rich proteins from seeds of Brazil nut (*Bertholletia excelsa* H.B.K.). *Eur J Biochem* 159: 597–604 (1986).
- Axelos M, Bardet C, Liboz T, La Van Thai A, Curie C, Lescure B: The gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1 α : molecular cloning, characterization and expression. *Mol Gen Genet* (1989).
- Carneiro M: Control of gene expression: transcription and mRNA stability. Ph.D. thesis, Lausanne University, Switzerland (1985).
- Casey J, Davidson N: Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nucl Acids Res* 4: 1539–1552 (1977).
- De Castro LAB, Lacava Z, Aramayo RA, Sampaio MJ, Gander ES: Evidence for a precursor molecule of Brazil nut 2S seed proteins from biosynthesis and cDNA analysis. *Mol Gen Genet* 206: 338–343 (1987).
- Favarolo J, Freisman R, Kamen R: Transcription maps of polyoma virus-specific RNA: Analysis by two-dimensional nuclease S1 mapping. *Meth Enzymol* 65: 718–749 (1980).
- Fluhr R, Moses P, Morelli G, Coruzzi G, Chua N-H: Expression dynamics of the pea *rbcs* multigene family and organ distribution of the transcripts. *EMBO J* 5: 2063–2071 (1986).
- Guerche P, De Almeida RP, Schwarzstein MA, Gander ES, Krebbers E, Pelletier G: Expression of 2S albumine from *Bertholletia excelsa* in *Brassica*. *Mol Gen Genet* 221: 306–314 (1990).
- Goldberg RB, Barker SJ, Perez-Grau L: Regulation of gene expression during plant embryogenesis. *Cell* 56: 149–160 (1989).
- Giuliano G, Pichersky E, Malik VS, Timko MP, Scolnik PA, Cashmore AR: An evolutionary conserved protein binding sequence upstream of a plant light-regulated gene. *Proc Natl Acad Sci USA* 85: 7089–7093 (1988).
- Higgins TJV: Synthesis and regulation of major proteins in seeds. *Ann Rev Plant Physiol* 35: 191–221 (1984).
- Josefsson L-G, Lennon N, Ericson ML, Rask L: Structure of a gene encoding the 1.7 S storage protein, napin, from *Brassica napus*. *J Biol Chem* 262: 12196–12201 (1987).
- Krebbers E, Herdies L, De Clercq A, Seurinck J, Leemans J, Van Damme J, Segura M, Gheysen G, Van Montagu M, Vanderkerckhove J: Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family. *Plant Physiol* 87: 859–866 (1988).
- Loenen WAM, Blattner FR: Lambda Charon vectors (Ch32, 33, 34 and 35) adapted for DNA cloning in recombination-deficient hosts. *Gene* 26: 171–179 (1983).
- Maxam AM, Gilbert W: A new method for sequencing DNA. *Proc Natl Acad Sci USA* 74: 560–564 (1977).
- McMasters GK, Carmichael GG: Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc Natl Acad Sci USA* 74: 4835–4838 (1977).
- Okamuro JK, Jofuku KD, Goldberg RB: Soybean seed lectin and flanking nonseed protein gene are developmentally regulated in transformed tobacco plants. *Proc Natl Acad Sci USA* 83: 8240–8244 (1986).
- Pikaard CS, Mignery GA, Ma DP, Park WD: Sequence of two apparent pseudogenes of the major potato tuber protein, patatin. *Nucl Acids Res* 14: 5564–5566 (1986).
- Rosahl S, Schmidt R, Schell J, Willmitzer L: Isolation and characterization of a gene from *Solanum tuberosum* encoding patatin, the major storage protein of potato tubers. *Mol Gen Genet* 203: 230–236 (1986).
- Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463–5467 (1977).
- Scofield SR, Crouch ML: Nucleotide sequence of the napin storage protein family from *Brassica napus*. *J Biol Chem* 262: 12202–12208 (1987).
- Sharief F, Li SS-L: Amino acid sequence of small and large subunits of seed storage protein from *Ricinus communis*. *J Biol Chem* 257: 14753–14759 (1982).
- Taylor B, Powell A: Isolation of plant DNA and RNA. *Focus* 4: 4–6 (1982).
- Vogt WM: Purification and further properties of single-strand-specific nuclease from *Aspergillus oryzae*. *Eur J Biochem* 33: 192–200 (1973).
- Youle R, Huang AHC: Occurrence of low molecular weight and high cysteine containing albumin storage protein in oilseeds of diverse species. *Am J Bot* 68: 44–48 (1981).