

Narbonin, a novel 2S protein from *Vicia narbonensis* L. seeds: cDNA, gene structure and developmentally regulated formation

Van Hai Nong^{1,2}, Bernhard Schlesier^{1,*}, Ronald Bassünier^{1,3}, Alexander Repik¹,
Christian Horstmann¹ and Klaus Müntz¹

¹Institut für Pflanzengenetik und Kulturpflanzenforschung, D-06466 Gatersleben, Germany (*author for correspondence); ²Institute of Biotechnology, National Center for Natural Science and Technology, Hanoi, Vietnam; ³Present address: Dept. of Agronomy, Purdue University, West Lafayette, IN 47906, USA

Received 14 October 1994; accepted in revised form 1 February 1995

Key words: biosynthesis, gene structure, narbonin, recombinant protein, 2S globulin, seed storage protein, *Vicia narbonensis* L.

Abstract

cDNA and genomic clones encoding narbonin, a 2S globulin from the seed of narbon bean (*Vicia narbonensis* L.), were obtained using the polymerase chain reaction (PCR) and sequenced. The full-length cDNA as well as genomic clones contain a single open reading frame (ORF) of 873 bp that encodes a protein with 291 amino acids comprising the mature narbonin polypeptide (M_r ca. 33 100) and an initiation methionine. The deduced amino acid sequence lacks a transient N-terminal signal peptide. The genomic clones do not contain any intron. No homology was found to nucleic acid and protein sequences so far registered in sequence data libraries. The biosynthesis of narbonin during embryogenesis is developmentally-regulated and its pattern of synthesis closely resembles that of typical seed storage globulins. However, during seed germination narbonin was degraded very slowly, indicating that it may have other function than storage protein. Southern analysis suggests the existence of a small narbonin gene family. Narbonin genes were also found in four different species of the genus *Vicia* as well as in other legumes such as *Canavalia ensiformis* and *Glycine max*. In *Escherichia coli* a recombinant narbonin was produced which yielded crystals like those prepared from narbonin purified from seeds.

Introduction

Narbonin, a 2S globulin which accounts for up to 2.6% of the total globulins in *Vicia narbonensis* L. seeds, was isolated, purified and crystallized for the first time about two decades ago [33, 34].

Recently, the three-dimensional (3D) structure of narbonin at 0.18 nm resolution was determined, from which the protein sequence could be predicted [16]. Narbonin structure closely resembles that of the enzyme triosephosphate isomerase (TIM), which is more commonly referred to as an

α/β -barrel or TIM-barrel structure [9, 37]. More than twenty proteins with this structural motif have been investigated and all were found to exhibit enzymatic activity. Narbonin is so far the only case of a protein with this structural type that lacks any known enzymatic activity. Since seed storage globulins typically lack enzyme activity, narbonin is thought to serve as a storage protein that is degraded during seed germination in order to contribute to the nitrogen nutrition of the growing seedling.

In contrast to the wealth of information about the 3D structure of the protein, the complete primary sequence of the protein has not been determined and very little is known about structure and expression of the narbonin gene. Here we report the isolation and sequencing of cDNA and genomic clones encoding narbonin, using PCR-based techniques. In addition, the developmentally regulated expression patterns of the gene during embryogenesis of *Vicia narbonensis* were analysed and the overexpression of the cDNA in *Escherichia coli* was achieved which should aid further structure-function investigations of this protein.

Materials and methods

Plant material

Vicia narbonensis L. plants were grown in an environmentally controlled growth chamber with a day/night regime of 16 h light at 23 °C and 8 h dark at 18 °C. At appropriate stages of seed development cotyledons were harvested and immediately used or frozen in liquid nitrogen and stored at - 80 °C.

Isolation of plant RNA, DNA and cDNA synthesis

Total RNAs were isolated from plant materials by the method of Chomczynski and Sacchi [3] using the 'Total RNA' isolation kit (Promega) or by the method described by de Vries *et al.* [6]. Poly(A)⁺ mRNAs were obtained using Poly-

ATtract kit (Promega). DNA was isolated from 3–5-day old seedlings using the salting-out method [27] adapted for plant materials (C. Becker, personal communication). First-strand cDNAs were synthesized by the First Strand Synthesis kit (Stratagene). Both strands of cDNAs were obtained by the Capture Clone cDNA synthesis kit (Promega). All experimental procedures using kits were performed according to the manufacturers' protocols.

Oligonucleotide synthesis

A degenerate primer P 72 (5'-GAR TAY ATI GGI GTI AAR CCI AA-3') was designed on the basis of the peptide region EYIGVKPN which forms a part of the N-terminal sequence of narbonin. A pair of adapter primers: AP 59/AP 60 (5'-GGC CCT GCA GGC CAC CGC GGT GAG CT-3') was synthesized as complementary strands. AP 60 was phosphorylated and then annealed to AP 59 resulting in a cDNA adaptor. Several narbonin sequence-specific primers for PCR: P 94 (5'-CAA CGT AGC AAA CGG CTC ATC-3'), NP 10 (5'-CCA TAT GCC TAA G/ACC TAT CTT TCG GG-3'), NP 11 (5'-CGG ATC CTC ATC G/A TT TGG CGA GGA G-3'), P 106 (5'-ATC TCC TCC TCC TCC TTT CAC ATT GTC AAA ATG CC-3') and P 107 (5'-GAA AAT AAT AAC AAT GGG TCT TAT TTA ATA TTT TTA AAG GC-3') as well as various fluorescent dye-labelled primers for DNA sequencing were used. The oligonucleotides were synthesized on an A394 DNA/RNA Synthesizer (Applied Biosystems) and OPC-purified according to the manufacturer's protocols.

PCR cloning and sequencing

PCR assays were performed on a thermal cycler Autogene II (Grant) using a GenAmp kit (Perkin-Elmer Cetus). The assay conditions were: 96 °C for 0.8 min; 56 °C for 0.8 min; 72 °C for 1.4 min; 94.5 °C for 0.7 min; 32 cycles followed by 8 min incubation at 72 °C. PCR products were directly

cloned into the pCRII vector of the TA-Cloning kit (Invitrogen). Plasmid DNAs were purified by the Qiagen plasmid purification procedure (Diagen) and were sequenced on both DNA strands using an ALF DNA sequencer (Pharmacia) and fluorescent dye-labelled primers. The sequence data were analysed using PC/GENE software (IntelliGenetics).

Rapid amplification of cDNA ends (5' RACE)

Using the 5' RACE system (Gibco BRL) the procedure was performed basically according to the manufacturer's protocol, except that instead of a specific primer (GSP1), oligo-dT was used for the first-strand cDNA synthesis by SuperScript reverse transcriptase. After dC tailing of the cDNA, it was amplified by PCR as described above using the 'anchor primer' of the kit, and the narbonin-specific antisense primer NP 11.

DNA and RNA blotting techniques

Samples of 10 µg plant DNA were digested with appropriate restriction enzymes, electrophoresed on a 0.8% agarose gel, and blotted onto a Nytrans-Plus filter (Schleicher & Schuell) by capillary transfer. The procedures were carried out according to the laboratory manual of Sambrook *et al.* [30]. For the RNA analysis samples of about 1 µg total RNAs per slot were blotted onto a nitrocellulose membrane under vacuum using a Hybri-Slot Manifold system (Gibco BRL). The blotted membranes were prehybridized at 65 °C for 2 h, in 5 × Denhardt's solution, 5 × SSC, 0.5% Blocking reagent (Boehringer), 0.5% SDS and then hybridized at 65 °C, for 12–14 h, in the same solution containing the *Hind* III fragment of pNaF6 as probe, labelled with [α^{32} P]-CTP using the Megaprime Labeling kit (Amersham). The membranes were washed twice with 2 × SSC, 0.1% SDS, at 65 °C for 1 h, followed by several washings with 0.5 × SSC, 0.1% SDS under the same conditions. Membranes were exposed to Hyperfilm-MP X-ray films with

an intensifying screen (Amersham) at –80 °C for 1–5 days.

Expression in E. coli

An *Nde* I/*Bam* HI fragment containing the whole narbonin-coding sequence of the cDNA clone pNaF6 was generated by means of PCR using the primer pair NP10/NP11 and inserted into corresponding sites of the pET12a expression vector (Novagen). A single colony of *E. coli*, strain HMS 174 (DE3), harbouring the expression plasmid was grown in 3 ml LB medium containing 50 µg/ml ampicillin at 37 °C for 3–5 h to reach an OD_{600 nm} of ca. 0.3–0.6 units. Then IPTG was added to a final concentration of 0.5 mM and the expression was achieved under the control of the T7 promoter for 3–5 h. The bacterial cells were harvested, resuspended in 50 mM Tris, 10 mM EDTA pH 8.0, and disrupted by several cycles of sonication, each 20–30 s with an interval on ice, using a Vibra-Cell sonicator (BioBlock). The total lysate was assayed or separated into soluble and 'inclusion bodies' fractions by centrifugation for 30 min at 4 °C and at 13 000 rpm.

Protein analysis, purification and crystallization

Total proteins of different stages of the developing and germinating cotyledons were extracted with 0.1 M Tris-HCl pH 9.5, 0.27 M NaCl, 0.1% SDS. Protein samples were prepared in SDS gel loading buffer containing 0.2% 2-mercaptoethanol and denatured by heating for 5 min at 100 °C. Equal volumes of extracts were loaded on SDS gels according to Laemmli [24]. For western analysis proteins were blotted onto Immobilon-P membrane (Serva) or nitrocellulose membrane BA 85 (Schleicher & Schuell) using a Mini Trans-blot Cell (BioRad) and detected by antibodies raised against the native narbonin. Recombinant narbonin was purified from the soluble fraction of the cell lysate of *E. coli* expressing narbonin cDNA using 300–500 ml bacterial culture as starting material. The same purification

procedure was applied as for the native narbonin [33]. Protein crystallization was performed as described previously [17].

Results

PCR cloning of narbonin cDNAs

Using the degenerate primer P 72 in combination with oligo-dT, a nearly full-length cDNA fragment of about 1.1 kb was amplified. One of the clones sequenced was shown to be narbonin-specific because fragments of its derived protein sequence corresponded to known peptide regions of the protein (Fig. 1, pNaC18).

For the amplification of the 5' ends, the synthetic adaptor A 59/60 was ligated to the ends of the double strand cDNA, then the adaptor primer AP 60 and the internal narbonin-specific antisense primer P 94 derived from the sequence of the clone pNaC18 were used. A single PCR product of about 0.46 kb was obtained, cloned and sequenced. The sequences obtained from five clones cover the complete N-terminal peptide and contain a corresponding additional region, the shortest of 16 bp (pNaN14) and the longest of 33 bp (Fig. 1, pNaN21) upstream from a codon for proline, which is known to be the N-terminal amino acid of the mature narbonin polypeptide. In addition, cDNA fragments containing the whole coding sequence and the complete 5' end were amplified using the 5' RACE technique. The longest 5' end of the clones obtained consists of 44 bp upstream from the proline codon (Fig. 1, clone pNaRN6). Thus, using both the primer/adaptor strategy and the 5' RACE technique we have obtained similar 5' ends of narbonin transcripts. On the other hand, primer extension experiments did not give any longer extension product (data not shown). Hence, we conclude that the methionine codon found directly upstream from the N-terminal proline is the initiation codon for the gene.

A full-length cDNA clone pNaF6 was isolated using the pair of narbonin-specific primers P106/P107 designed after the two cDNA ends were known.

Characterization of narbonin cDNA clones

The full-length cDNA clone pNaF6 is 1158 bp long and contains an ORF encoding 291 amino acids (an initiation methionine codon plus the complete narbonin mature polypeptide chain, M_r ca. 33 200). Accordingly, the mature polypeptide has M_r ca. 33 100, which agrees with the results obtained by SDS-PAGE. Overall, the sequence displays 70% identity with the amino acid sequence predicted from the X-ray derived 3D structure at 0.18 nm resolution [16]. The 30% mismatch was caused by difficulties in distinguishing between Asn, Asp and Leu as well as between Gln and Glu and between Thr and Val at this resolution [14]. The cDNA-derived amino acid sequence contains the three peptides analysed earlier, except for four glutamine residues that were determined as glutamic acids from the peptide sequence data (Fig. 1). These glutamine residues probably do exist in the primary translation products and were changed to glutamic acids either post-translationally or during protein purification as the result of protein deamidation [26]. In addition, the calculated pI of 5.26 of the deduced polypeptide from pNaF6 is in good agreement with the value of 5.17 ± 0.10 estimated for the native narbonin [33].

A total of nine mismatches in the nucleotide sequences of the cDNA clones analysed was registered. Seven of them are found at the third nucleotide position of the respective codons and so do not alter the encoded amino acids. The other two mismatches lead to conservative amino acid substitutions (Val-79→Ile and Ser→194-Gly). Notably, two nearly full length clones pNaC18 and pNaRN6 independently obtained from different cDNA preparations and PCR amplifications match perfectly, except for a single mismatch in the nucleotide at position 126. This is located at the third position of a codon and does not cause a change in the amino acid. Since the same nucleotide changes were obtained with different clones, of which both DNA strands had been sequenced, we believe that they reflect very closely related genes or allelic variants of the gene rather than PCR artefacts.

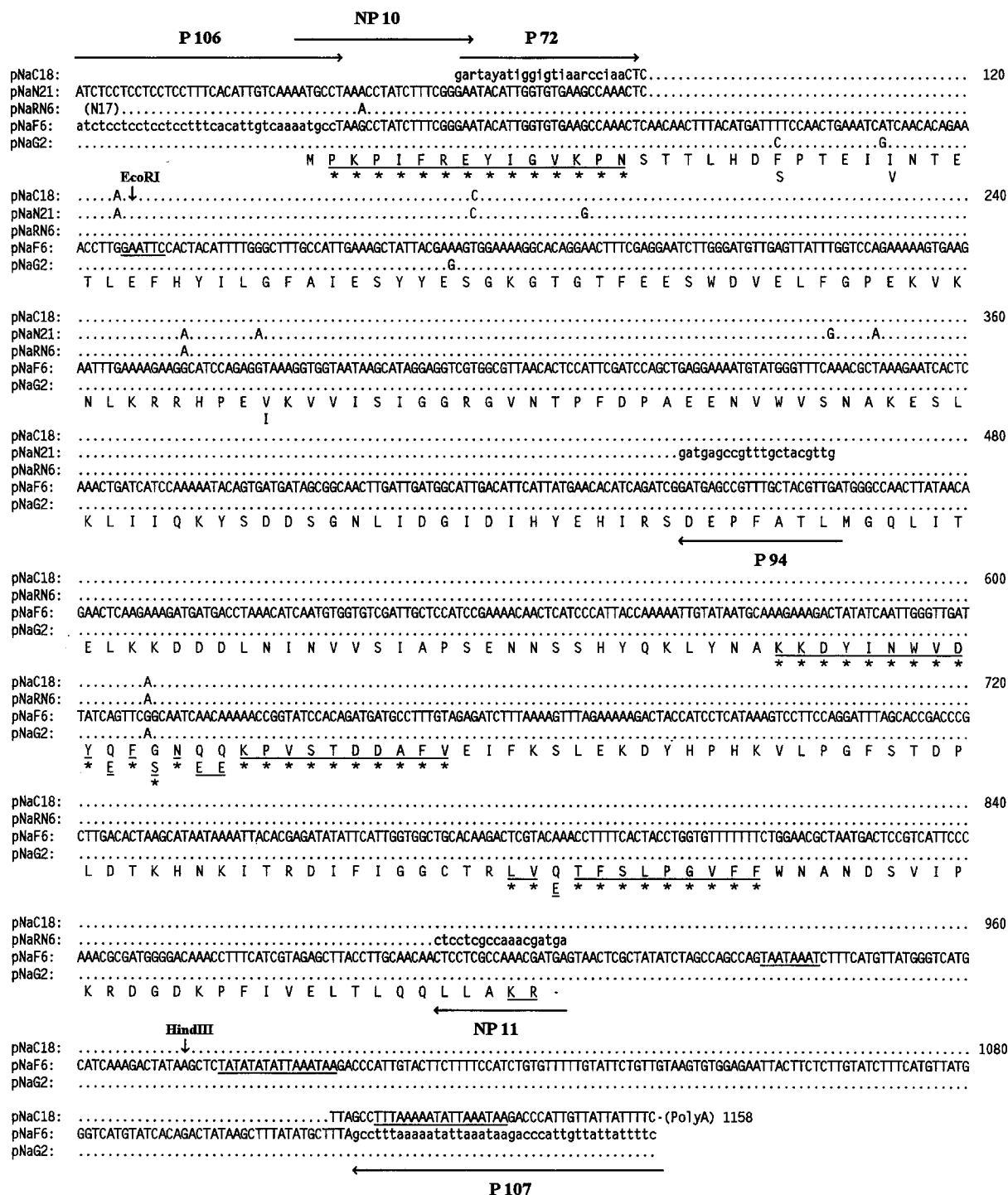


Fig. 1. Nucleotide sequences of the narbonin cDNA and genomic clones and the deduced amino acid sequences. Only the full-length cDNA clone pNaF6, together with the ends of other clones and the mismatches, are shown. Arrows indicate the primer positions. The nucleotides given in lower case are derived from the primers. The peptide fragments known from amino acid sequencing are underlined. Matched nucleotides and amino acids are represented by dots and stars respectively. The sequence of the region of 17 additional nucleotides, N(17), at the 5' end of clone pNaRN6 obtained by 5' RACE technique is: ACC AGA ACC AAT TTA TTA TCT CC. Three alternative potential polyadenylated signals are also underlined.

The C-terminal region of the cDNA-derived polypeptide sequence is two amino acid residues longer than that predicted on the basis of the X-ray diffraction analysis. This part of molecule was hidden in the 3D structure and its precise determination was rather difficult and presumably incorrect. C-terminal protein sequencing has confirmed that the C-terminal amino acid sequence is Lys-Arg-COOH.

Cloning and characterization of a narbonin gene

Pairs of narbonin-specific oligonucleotides P106/P107 or NP10/NP11 designed on the basis of the

narbonin cDNA clones were used to amplify genomic sequences from the total plant genomic DNA by PCR. Products having the same size as those obtained with a cDNA library were amplified. Correspondingly, no intron was found in the sequenced genomic clone pNaG2 (Fig. 1). Several mismatches which do not result in amino acid changes exist at the same positions as those found in the cDNA clones (see above). Two additional changes induce amino acid substitutions, one of which is non-conservative (Phe-22→Ser) while the other is conservative (Ile-27→Val). To verify the sequence two additional genomic clones (pNaG3 and pNaG9) from the same PCR experiment were sequenced. The sequences of all

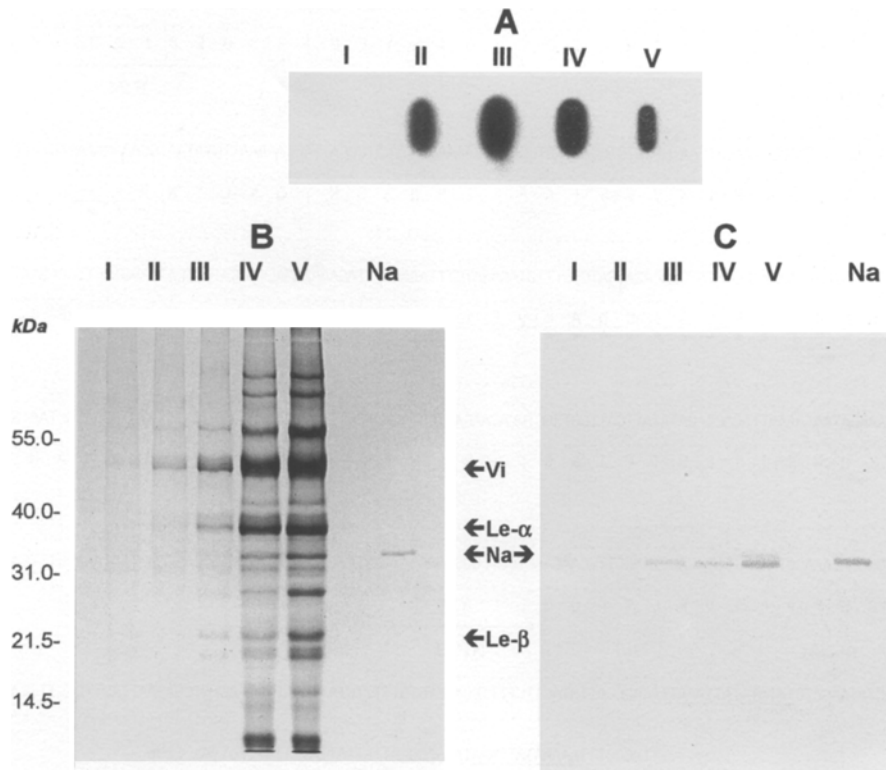


Fig. 2. Developmental pattern of narbonin gene expression: occurrence of the narbonin-specific mRNAs (A) and the accumulation of narbonin in the developing *Vicia narbonensis* seeds (B/C). Following developmental stages: I, 1–4 mm (18 DAF); II, 4–6 mm (21–24 DAF); III, 6–8 mm (25–30 DAF); IV, 8–11 mm (31–36 DAF); V, 6–8 mm (42–48 DAF) were analysed. Samples of about 1 µg total RNAs from different stages were transferred by vacuum blotting onto a nitrocellulose filter. The slot-blot filter was hybridized with the [^{32} P]-labelled *Hind* III fragment of pNaF6 as probe. SDS-PAGE was performed under reducing conditions on 1 mm thick 11.25% gel slabs with 0.5 µg of narbonin as the standard. The proteins were either stained with Coomassie dye (B) or transferred onto nitrocellulose membrane and detected by polyclonal antibodies raised against native narbonin (C). Na, V, Le-α and Le-β denote narbonin, vicilin, legumin-α and legumin-β, respectively.

three genomic clones were identical. Thus, the sequences of the cDNA and the genomic clones we obtained do not completely match each other indicating that they may be derived from different genes or alleles. Cloning of the promoter part of the gene in order to characterize in more detail the regulation of the narbonin gene expression is currently under work.

Expression of narbonin gene in developing seeds

Narbonin mRNA and protein were isolated at five different stages during seed development. Narbonin-specific transcripts first appeared at 21–24 days after flowering (DAF) and reached their maximum 25–30 DAF. Thereafter, the level of narbonin mRNAs decreased but they were still clearly detectable in the mature seeds (Fig. 2A). Correspondingly, narbonin protein was at first found in seeds at 21–24 DAF. It continuously accumulated with a pattern similar to that of vicilin and legumin, the major seed storage globulins (Fig. 2B/C).

Degradation of narbonin in germinating seeds

The mobilization of globulins during the germination of *Vicia narbonensis* seeds was followed for a period of 24 days after imbibition (DAI). By 24 DAI the plantlets have developed three to four leaves and their total dry weights are equal to or larger than that of the original seeds. SDS-PAGE and western blots show that degradation patterns of legumin and vicilin during the early and mid stages of storage protein mobilization are in general similar to that of storage globulins from field beans [25]. In contrast to the time course of legumin and vicilin, the level of narbonin decreased very slowly and the protein is still easily detected in cotyledons 24 DAI (Fig. 3). This result indicates that narbonin is not a true seed storage protein.

Southern analysis

After digestion of genomic DNA with three different restriction enzymes and hybridization under stringent conditions, Southern analyses

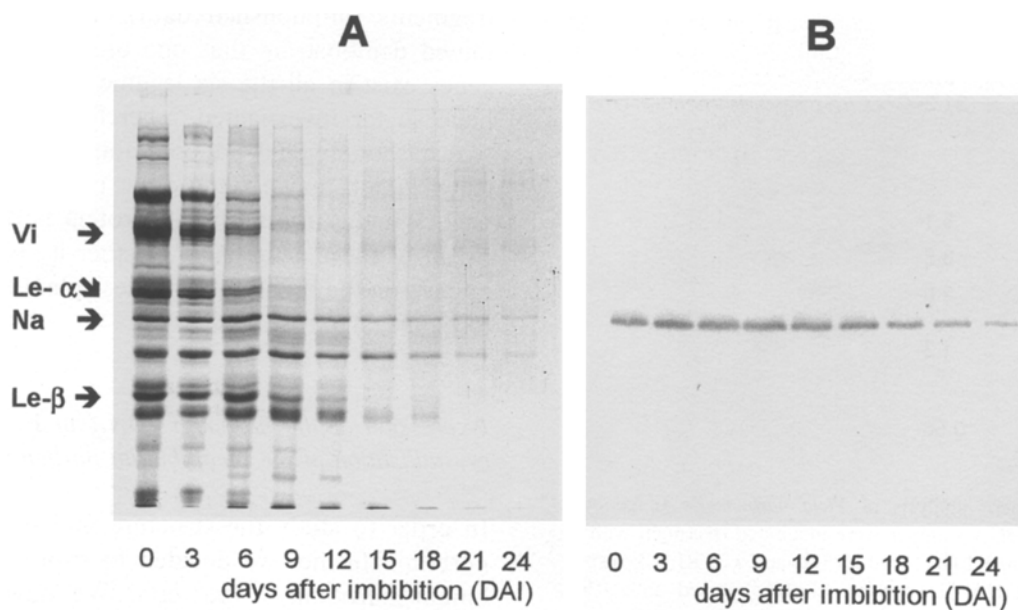


Fig. 3. Time course of narbonin degradation. Seeds were planted directly in wet soil. Cotyledons were harvested every 3 days over a 24 day period after imbibition. Total proteins were analysed by SDS-PAGE (A) and western blot (B). Methods and labels are the same as described in Fig. 2.

yielded two bands which strongly hybridized with the cDNA probe (Fig. 4). These are fragments of about 4.8 and 4.3 kb in *Bam* HI digests, 15.5 and 2.8 kb in *Hind* III digests, and 4.0 and 1.5 kb in *Eco* RI digests, respectively. In addition, several less prominent bands were observed, that may arise either from cross reaction with a related gene or represent unspecific hybridization. The largest number of weaker bands were found in the *Eco* RI digest, whereby two bands were relatively small. The latter are probably hybridization products of the small fragment of the probe generated by an internal *Eco* RI site at position 128. Assuming that genomic fragments are intron-free and knowing that the cDNA probe (0.97 kb *Hind* III fragment of pNaF6) does not contain any *Hind* III or *Bam* HI sites, the existence of a small narbonin gene family with at least two members can be implied.

Occurrence of narbonin gene(s) in other plant species

Besides *V. narbonensis*, significant amounts of narbonin protein were found in the seeds of *V. pannonica*. The protein could be crystallized. The

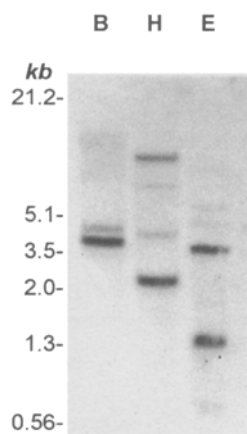


Fig. 4. Southern analysis of *Vicia narbonensis* L. genomic DNA. The DNA samples were incubated overnight with an excess of *Bam* HI (B), *Hind* III (H) and *Eco* RI (E), respectively. The DNA digests were electrophoresed on a 0.8% agarose gel, blotted onto a Nytrans-Plus filter. The nearly complete insert of the cDNA clone pNaF6 (0.97 kb *Hind* III fragment), labelled with [α^{32} P]-CTP was used as hybridization probe.

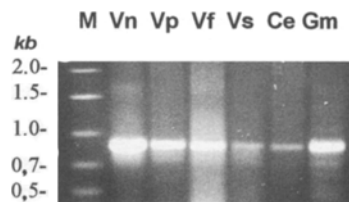


Fig. 5. PCR amplification of narbonin-specific sequences from genomic DNAs of various legume plants. Samples of about 0.1 μ g total DNA were used as PCR templates. After 32 cycles of amplification 5 μ l aliquots of PCR products were electrophoresed on a 1% agarose gel, stained with ethidium bromide and visualized on an UV transilluminator. M, DNA markers (BioRad); Vn, *Vicia narbonensis*; Vp, *Vicia pannonica*; Vf, *Vicia faba*; Vs, *Vicia sativa*; Ce, *Canavalia ensiformis*; Gm, *Glycine max*.

twelve N-terminal amino acid residues were identical to those of narbonin from *V. narbonensis* but there are differences in the primary structure as detected by peptide mapping [32].

Using the narbonin-specific primers NP10/ NP11 we have succeeded in amplifying narbonin genomic DNA fragments from four *Vicia* species as well as from jack bean (*Canavalia ensiformis*) and soybean (*Glycine max*) (Fig. 5). We have cloned and sequenced all of these genomic DNA fragments (unpublished data). The results obtained demonstrate that one or more narbonin genes exist in all the six legume plants that we have so far investigated. Therefore, we believe that narbonin genes are common or universal in legumes and may exist in other plant families as well. Whether the narbonin protein is produced by all of these genes and whether it specifically appears in the seeds of all of these plants remains to be investigated.

Expression of the narbonin cDNA in *E. coli* and crystallization of the recombinant narbonin

In order to study the structure and function of narbonin further we decided to express recombinant narbonin in bacteria. We have so far achieved expression levels of more than 50% of total bacterial cell proteins (Fig. 6). The narbonin produced in bacteria is present in both the soluble

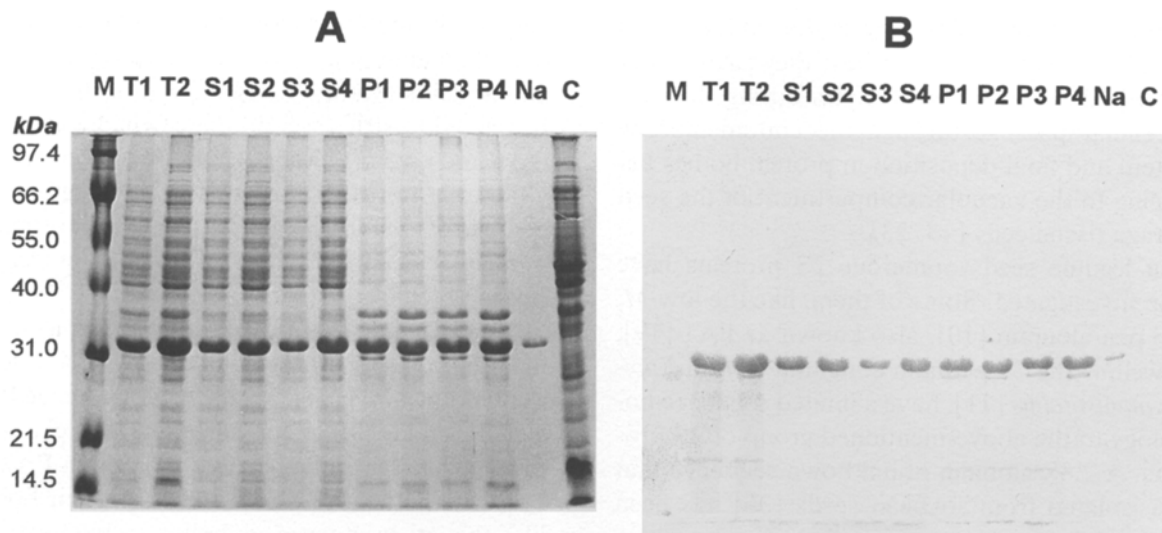


Fig. 6. Expression of the narbonin cDNA in *E. coli*. SDS-PAGE Coomassie-stained protein patterns (A) and the immunoblot of the same gel using antibodies raised against native narbonin (B). Shown are: M, molecular weight markers (Serva); T1-T2, total cell lysates; S1-S4, supernatants; P1-P4, pellets after centrifugation at 13 000 rpm for 30 min at 4 °C (where the numbers 1 to 4 are the numbers of the corresponding clones analysed); Na, native narbonin; C, total cell lysate of the control *E. coli*, strain HMS 174 (DE3), without narbonin expression plasmid.

and the 'inclusion bodies' fractions in approximately equal amounts. Assuming that narbonin would more likely be correctly folded in the soluble fraction, we purified this fraction and have used it to raise antibodies. Under the same conditions as used with native narbonin, the purified recombinant protein readily formed crystals (Fig. 7) with the same form [15].

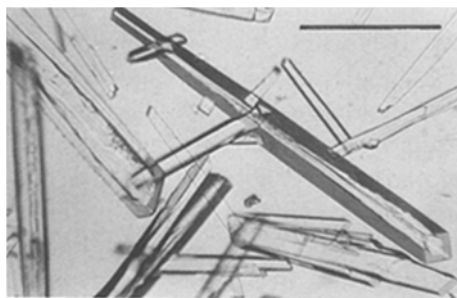


Fig. 7. Crystalline preparation of the recombinant narbonin. Crystals were grown by pH shift during dialysis as described previously [32]. The bar in the upper right corner indicates 0.5 mm length.

Discussion

Due to their quantitative predominance, the 11S and the 7S globulins (also termed legumins and vicilins, respectively) found in legume seeds and other angiosperms have been extensively investigated (for review see [2]). Much less is known about the minor components of seed proteins, often classified as the 2S protein fraction. Proteins with sedimentation coefficients of approximately 2S are not only found in the globulin but also in the albumin fraction according to Osborne [29]. In addition to similarities in their solubility properties, upon which the Osborne classification scheme is based, many 2S proteins exhibit sequence similarities and are structurally related to the cereal prolamins [31]. This group of 2S proteins includes biologically active proteins, like trypsin/ α -amylase inhibitors [12] or 'allergenic' proteins [20], as well as storage proteins, such as the napin of rapeseed [8], the 2S albumin of sunflower [21] or the sulfur-rich 2S albumin from Brazil nut [1]. They are synthesized as precursors comprised of an N-terminal transient signal pep-

tide which directs them into the endoplasmic reticulum, and, in most cases, they are further processed by limited proteolysis during their intracellular passage through the endomembrane system and final deposition in protein bodies belonging to the vacuolar compartment of the seed storage tissue cells [13, 23].

In legume seeds numerous 2S proteins have been investigated. Some of them, like the low- M_r '2S' pea albumin [10], also known as PA1 [19], as well as the 2S globulin conglutin δ from *Lupinus angustifolius* [11], have a limited sequence homology to the above-mentioned group of 2S proteins. A 2.8S globulin of unknown sequence that was isolated from soybean seeds [36] has been reported to have some trypsin/ α -chymotrypsin inhibitor activities [22]. The major pea albumins PMA-L/S, also named as PSA [5] or PA2 [18], clearly belong to another type of 2S proteins, which appear to be located in the cytoplasm. The absence of an N-terminal signal peptide in the cDNA-derived amino acid sequence of the proteins seems to support this compartmentation [18]. Preliminary functional studies indicate that the pea albumins are much more slowly degraded and therefore persist much longer than the major storage globulins, legumin and vicilin, in the cotyledons of germinating seeds. This argues against their function as storage proteins.

Narbonin, which was isolated from the mature seeds of the narbon bean (*Vicia narbonensis* L.) is an additional representative of the heterogeneous group of legume 2S proteins. According to its solubility it was classified as a globulin [33, 34]. Here we report the characterization of a number of cDNA and genomic clones encoding narbonin, on the cDNA-derived amino acid sequence of this polypeptide, and on the formation of narbonin and its mRNA in the cotyledons of the developing narbon beans. We failed to find any DNA or protein sequence homologous to narbonin in either the EMBL Data Library (release 39) or the SwissProt Database (release 29) of June 1994. Since narbonin was shown to have a TIM-barrel structure [16] an alignment of the narbonin polypeptide sequence with those other proteins exhibiting the TIM motif was performed. Only

very weak similarity was found, mainly in charges of certain amino acids (data not shown). The few amino acids that are conserved in narbonin may be related to structural domains which determine the same type of folding in these apparently evolutionarily unrelated proteins [9]. Therefore, we conclude from our sequence data that narbonin is a member of a new family of proteins present in legume seeds.

The sequence data also show that narbonin, unlike the major storage globulins, such as legumin and vicilin, lacks a consensus sequence for a signal peptide. Whether narbonin is a cytosolic protein, similar to the pea major albumin PA2 [5, 18], or whether it is localized in protein bodies like the other storage globulins, remains to be elucidated. Some plant proteins, such as β -amylase [28] or lipoxygenase [35], which also lack signal peptides but are located nevertheless in vacuoles, have been reported. A direct path to the lysosomal or vacuolar compartment, in which so-called peptide recognition proteins [7] are thought to be involved, has been proposed for the transport of such proteins [4].

Although narbonin was shown to be formed and accumulated in the developing seeds in a similar stage-dependent pattern as that exhibited by the major storage globulins in the germinating seeds, it breaks down only very slowly. Therefore, its function as nitrogen reserve for germinating seedlings is highly questionable.

Southern blot analysis has shown that a small narbonin gene family exists. Furthermore, narbonin gene fragments were found in six different legume species indicating that similar genes may be distributed within the Leguminosae plant family. Searching for other narbonin-like genes and their expression products in different plants may help to clarify the structural origin as well as function of narbonin.

Narbonin is known to be easily crystallized [33, 34] and was the first 2S globulin whose crystal structure had been determined at 0.18 nm resolution. The 3D structure model of narbonin [16] was obtained before the primary structure of narbonin protein or gene was completely known. Our sequence data provide additional informa-

tion for the refinement of the 3D structure model of narbonin. In this respect, comparison of structural data obtained from the crystalline preparations of the native as well as of the 'recombinant' narbonins should be of interest. This work as well as investigations of the intracellular transport and localization of narbonin are now in progress.

Acknowledgements

We thank Monika Wiesner and Annegret Wolf for excellent technical assistance. We are grateful to Susanne König for the DNA sequencing, Dr Renate Manteuffel for the preparation of the antiserum against native narbonin, Dr Regine Kraft (Max-Delbrück-Centrum für Molekulare Medizin, Berlin-Buch) for the C-terminal protein sequencing and Rose-Marie Gillandt for the oligonucleotide synthesis. We wish also to thank Dr David Waddell for the critical reading of the manuscript. This work was supported, in part, by grants Ba 1215/1-1 and SFB 363/A-5 from the Deutsche Forschungsgemeinschaft (DFG) as well as by grant 233A0911 from the government of Sachsen-Anhalt, Germany.

References

1. Altenbach SB, Pearson KW, Sun SSM: Nucleotide sequences of the Brazil nut methionine-rich 2S albumin gene family. *Plant Physiol* 98: 1520-1522 (1992).
2. Casey K, Domoney C, Ellis N: Legume storage proteins and their genes. *Oxford Surv Plant Mol Cell Biol* 3: 1-95 (1986).
3. Chomczynski P, Sacchi N: Single-step method of RNA isolation by acid guanidinium-thiocyanate phenol-chloroform extraction. *Anal Biochem* 102: 156-159 (1987).
4. Chrispeels MJ, Staehelin LA: Budding, fission, transport, targeting, fusion-frontiers in secretion research. *Plant Cell* 4: 1008-1016 (1992).
5. Croy RRD, Hoque MS, Gatehouse JA, Boulter D: The major albumin proteins from pea (*Pisum sativum* L.). Purification and some properties. *Biochem J* 218: 795-803 (1984).
6. De Vries S, Hoge H, Bisseling T: Isolation of total and polysomal RNA from plant tissue. In: Gelvin SB, Schilperoort RA (Eds) *Plant Molecular Biology Manual*, Chapter B/6. Kluwer Academic Publishers, Dordrecht (1991).
7. Dice JF: Peptide sequences that target cytosolic proteins for lysosomal proteolysis. *Trends Biochem Sci* 15: 305-309 (1990).
8. Ericson ML, Rødin J, Lenman M, Glimelius K, Josefsson L-G, Rask L: Structure of the rapeseed 1.7 S storage protein, napin, and its precursor. *J Biol Chem* 261: 14576-14581 (1986).
9. Farber GK: An α/β -barrel full of evolutionary trouble. *Curr Opin Struct Biol* 3: 409-412 (1993).
10. Gatehouse JA, Gilroy J, Hoque MS, Croy RRD: Purification, properties and amino acid sequence of a low- M_r abundant seed protein from pea (*Pisum sativum* L.). *Biochem J* 225: 239-247 (1985).
11. Gayler KR, Kolivas S, Macfarlane AJ, Lilley GG, Baldi M, Blagrove RJ, Johnson ED: Biosynthesis, cDNA and amino acid sequences of a precursor of conglutin δ , a sulphur-rich protein from *Lupinus angustifolius*. *Plant Mol Biol* 15: 879-893 (1990).
12. Halford NG, Morris NA, Urwin P, Williamson MS, Kasarda DD, Lew EJL, Kreis M, Shewry PR: Molecular cloning of the barley seed protein cMd: a variant member of the alpha-amylase/trypsin inhibitor family of cereals. *Biochim Biophys Acta* 950: 435-440 (1988).
13. Hara-Nishimura I, Takeuchi Y, Inoue K, Nishimura M: Vesicle transport and processing of the precursor to 2S albumin in pumpkin. *Plant J* 4: 793-800 (1993).
14. Hennig M, Pfeffer-Hennig S, Dauter Z, Wilson KS, Schlesier B, Nong VH: Crystal structure of narbonin at 1.8 Å resolution. *Acta Cryst D* 51: 177-189 (1995).
15. Hennig M, Schlesier B: Narbonin, a 2S seed globulin from *Vicia narbonensis* L. Biochemical characterization and X-ray structure. In: Schwenke KD, Mothes R (Eds) *Food Proteins: Structure and Functionality*, pp. 109-117. VCH Verlagsgesellschaft, Weinheim, FRG, (1993).
16. Hennig M, Schlesier B, Dauter Z, Pfeffer S, Betzel C, Höhne W-E, Wilson KS: A TIM barrel protein without enzymatic activity? Crystall-structure of narbonin at 1.8 Å resolution. *FEBS Lett* 306: 80-84 (1992).
17. Hennig M, Schlesier B, Pfeffer S, Höhne W-E: Narbonin, a 2S globulin from *Vicia narbonensis* L. Crystallization and preliminary crystallographic data. *J Mol Biol* 215: 339-340 (1990).
18. Higgins TJV, Beach LR, Spencer D, Chandler PM, Randall PJ, Blagrove RJ, Kortt AA, Guthrie RE: cDNA and protein sequence of a major pea seed albumin (PA2: $M_r \sim 26,000$). *Plant Mol Biol* 8: 37-45 (1987).
19. Higgins TJV, Chandler PM, Randall PJ, Spencer D, Beach LR, Blagrove RJ, Kortt AA, Inglis AS: Gene structure, protein structure, and regulation of the synthesis of a sulfur-rich protein in pea seeds. *J Biol Chem* 261: 11124-11130 (1986).
20. Izumi H, Adachi T, Fujii N, Matsuda T, Nakamura R, Tanaka K, Urisu A, Kurosawa Y: Nucleotide sequence of a cDNA clone encoding a major allergenic protein in rice seeds: homology of the deduced amino acid sequence

- with members of alpha-amylase/trypsin inhibitor family. FEBS Lett 302: 213–216 (1992).
21. Kortt AA, Calldwell JB, Lilley GG, Higgins TJV: Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (*Helianthus annuus* L.). Eur J Biochem 195: 329–334 (1991).
 22. Koshiyama I, Kikuchi M, Fukushima D: 2S globulins of soybean seeds. 2. Physicochemical and biological properties of protease inhibitors in 2S globulins. J Agric Food Chem 29: 340–343 (1981).
 23. Krebbers E, Herdies L, De Clercq A, Seurinck J, Leemans J, Van Damme J, Segura M, Gheysen G, Van Montagu M, Vandekerckhove J: Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family. Plant Physiol 87: 859–866 (1988).
 24. Laemmli UK: Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. Nature 227: 680–685 (1970).
 25. Lichtenfeld C, Manteuffel R, Müntz K, Neumann D, Scholz G, Weber E: Protein degradation and proteolytic activities in germinating field beans (*Vicia faba* L. var. *minor*). Biochem Physiol Pflanzen 174: 255–274 (1979).
 26. Liu DT: Deamidation: a source of microheterogeneity in farmaceutical proteins. Trends Biotechnol 10: 364–369 (1992).
 27. Miller SH, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. Nucl Acids Res 16: 1215–1215 (1988).
 28. Monroe JD, Salminen MD, Preiss J: Nucleotide sequence of a cDNA clone encoding β -amylase from *Arabidopsis thaliana*. Plant Physiol 97: 1599–1601 (1991).
 29. Osborne TB: The Vegetable Proteins, 2nd ed., Longmans, London (1924).
 30. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
 31. Shewry PR, Tatham AS: The prolamin storage proteins of cereal seeds: structure and evolution. Biochem J 267: 1–12 (1990).
 32. Schlesier B, Hennig M, Kraft R, Horstmann C: Biochemical characterization and preliminary X-ray data of narbonin from *Vicia pannonica*. In: Schwenke KD, Mothes R (Eds) Food Proteins: Structure and Functionality, pp. 118–120. VCH Verlagsgesellschaft, Weinheim, FRG (1993).
 33. Schlesier B, Manteuffel R, Rudolph A, Behlke J: Studies on seed globulins from legumes. VII. Narbonin, a 2S globulin from *Vicia narbonensis* L. Biochem Physiol Pflanzen 173: 420–428 (1978).
 34. Schlesier B, Scholz G: Studies on seed globulins from legumes. II. A crystalline protein from the globulin fraction of *Vicia narbonensis* L. Biochem Physiol Pflanzen 166: 367–369 (1974).
 35. Tranbarger TJ, Franceschi VR, Hilderbrand DF, Grimes HD: The soybean 94-kilodalton vegetative storage protein is a lipxygenase that is localized in paraveinal mesophyll cell vacuoles. Plant Cell 3: 973–987 (1991).
 36. Vaintraub IA, Shutov AD: Isolation and characterization of 2.8 S protein from soya seeds. Biokhimiya 34: 984–992 (1969).
 37. Wierenga RK, Noble MEM, Davenport RC: Comparison of the refined crystal structures of liganded and unliganded chicken, yeast and trypanosomal triosephosphate isomerase. J Mol Biol 224: 1115–1126 (1992).