

Translational and structural analysis of the shortest legume *ENOD40* gene in *Lupinus luteus*

Jan Podkowinski¹✉, Agnieszka Zmienko¹, Blazena Florek¹, Pawel Wojciechowski²,
Agnieszka Rybarczyk¹, Jan Wrzesinski¹, Jerzy Ciesiolka¹, Jacek Blazewicz¹,
Adam Kondorosi³, Martin Crespi³ and Andrzej Legocki¹

¹Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland; ²Institute of Computing Science, Poznan University of Technology, Poznań, Poland; ³Institut des Sciences du Vegetal, Gif sur Yvette, Cedex, France

Received: 02 October, 2008; revised: 04 February, 2009; accepted: 06 March, 2009
available on-line: 14 March, 2009

Two early nodulin 40 (*enod40*) genes, *ENOD40-1*, the shortest legume *ENOD40* gene, and *ENOD40-2*, were isolated from *Lupinus luteus*, a legume with indeterminate nodules. Both genes were expressed at similar levels during symbiosis with nitrogen-fixing bacteria. *ENOD40* phylogeny clustered the *L. luteus* genes with legumes forming determinate nodules and revealed peptide similarities. The *ENOD40-1* small ORF A fused to a reporter gene was efficiently expressed in plant cells, indicating that the start codon is recognized for translation. The *ENOD40-1* RNA structure predicted based on Pb(II)-induced cleavage and modeling revealed four structurally conserved domains, an absence of domain 4 characteristic for legumes of indeterminate nodules, and interactions between the conserved region I and a region located upstream of domain 6. Domain 2 contains Mg(II) ion binding sites essential for organizing RNA secondary structure. The differences between *L. luteus* and *Glycine max* *ENOD40* RNA models suggest the possibility of a switch between two structural states of *ENOD40* transcript.

Keywords: *ENOD40*, nitrogen fixation, phylogeny, transcript structure

INTRODUCTION

ENOD40 genes exhibit several intriguing structural features, such as the absence of a long conserved ORF, the presence of two small conserved ORFs, and regions corresponding to conserved secondary structures of the transcript. Due to its high expression during symbiotic nitrogen fixation, this gene is intensively studied in legumes. *ENOD40* is also present in plants from the *Poaceae* and *Solanaceae* families (Kouchi *et al.*, 1999; Compaan *et al.*, 2003; Larsen, 2003; Vlegghels *et al.*, 2003). Moreover, a recent search for RNA domains of secondary

structures similar to *ENOD40* domains identified a number of *ENOD40*-like genes from diverse angiosperm families, including *Brassicaceae* (Gultyaev & Roussis, 2007).

The mechanism of *ENOD40* gene activity seems to be of a dual mode: one relying on the encoded short peptide(s) and the second depending on the transcript structure. A comparison of *Papilionoideae* *ENOD40* genes has revealed two conserved regions, I and II, located within the 5' end and the central part of the cDNA, respectively, and up to six domains of conserved RNA secondary structure (Gultyaev & Roussis, 2007). The legume *ENOD40*

✉Corresponding author: Jan Podkowinski, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Z. Noskowskiego 12/14, 61-704 Poznań, Poland; phone: (48) 61 852 8503 ext. 148; fax: (48) 61 852 0532, e-mail: jpodkow@ibch.poznan.pl

Accession numbers: *L. luteus* early nodulin (*ENOD40-1*) mRNA, AF352375; *L. luteus* early nodulin (*ENOD40-1*) gene, AF352372; *L. luteus* *ENOD40-2* mRNA, AF352374; *L. luteus* *ENOD40-2* gene, AF352373.

Abbreviations: CaMV, cauliflower mosaic virus; DPI, days post infection; *ENOD*, early nodulin; *GUS*, beta-glucuronidase reporter gene; NOS, nopaline synthase; ORF, open reading frame; PCR, polymerase chain reaction; *pfu*, plaque forming units; sORF, small open reading frames coding for peptides shorter than 100 amino acids; UTR, untranslated region.

genes do not contain conserved coding sequences except for two short ORFs, ORF A and ORF B, that encode small peptides. ORF A of 10–13 codons is located in the conserved region I and partially overlaps ORF B. The growing evidence of the role of plant short peptides supports the hypothesis that either the single *ENOD40* peptide A or both A and B peptides may have biological functions (Barciszewski & Legocki, 1997; Linsey, 2001; Wen *et al.*, 2004). Both *ENOD40* ORFs match the novel class of genetic elements—sORS (short open reading frames coding for peptides shorter than 100 amino acids). ORF A translation was demonstrated for *Medicago truncatula* (Sousa *et al.*, 2001) and *Glycine max ENOD40*, in which the synthesis of both ORF A and ORF B peptides was proven *in vitro* (Rohrig *et al.*, 2002). The *G. max* ORF A and ORF B peptides were found to interact with sucrose synthase (Rohrig *et al.*, 2002) and the ORF A peptide covalently bound to cystein 264 of sucrose synthase increased the enzyme cleavage activity (the enzyme could function in both sucrose synthesis and degradation) (Rohrig *et al.*, 2004). ORF A and ORF B peptides antagonize *Zea mays* sucrose synthase phosphorylation decreasing the efficiency of the enzyme's proteolysis (Hardin *et al.*, 2003). This competition between *ENOD40* peptides and plant-specific calcium-dependent kinase, an enzyme involved in regulation of diverse cellular processes (Klimecka & Muszynska, 2007), points to an *ENOD40* regulatory function. In addition to the above *in vitro* data, *ENOD40* peptide A activity was also confirmed in *Arabidopsis thaliana*, where it inhibited expansion of protoplasts similarly to overexpression of the full-length *ENOD40* gene (Guzzo *et al.*, 2005).

The interactions between sucrose synthase and *ENOD40* peptides A and B suggest a role for these peptides in photosynthate accumulation in sink regions and are congruent with their expression pattern. The activity of the *ENOD40* genes is associated with new organ formation with high expression observed during the development of the nodule, an organ specific for symbiotic nitrogen fixation (Asad *et al.*, 1994; Crespi *et al.*, 1994; Minami *et al.*, 1996; Fang & Hirsch, 1998; Wan *et al.*, 2007; Kumagai *et al.*, 2006; Murakami *et al.*, 2006). The gene is expressed also in mature nodules, meristems and cells of high mitotic or dedifferentiation activity (Kouchi & Hata, 1993; Yang *et al.*, 1993; Matvienko *et al.*, 1994; Vijn *et al.*, 1995; Papadopoulou *et al.*, 1996). Transgenic plants with silenced or overexpressed *ENOD40* gene provide evidence on a direct relationship between the gene expression and nodule formation efficiency (Charon *et al.*, 1997; 1999; Wan *et al.*, 2007).

ENOD40 expression not related to symbiotic interactions is associated with new organ initiation and developmental processes, such as formation of

lateral roots and shoots (Papadopoulou *et al.*, 1996; Varkonyi-Gasic & White, 2002). In the non-legume plant *Oryza sativa*, *ENOD40* expression is observed in young, lateral vascular bundles prior to the onset of leaf expansion (Kouchi *et al.*, 1999). Recently, an *ENOD40* function in cell elongation and cross-talk with ethylene signaling was demonstrated in a *Nicotiana tabacum* cell suspension (Ruttink *et al.*, 2006), *A. thaliana* (Guzzo *et al.*, 2005) and rice (Dey *et al.*, 2004).

Apart from the biological activity of the sORFs, the *ENOD40* gene has been proposed to be active at the RNA level based on the GC content and free energy of the transcript folding similar to those of noncoding RNAs (Crespi *et al.*, 1994; Sousa *et al.*, 2001). The finding that the putative RNA structural elements of *ENOD40*-like genes are more conserved than the peptide sequence suggests that the general function(s) of the genes might be determined by the RNA (Gulyaev & Roussis, 2007). This hypothesis is supported by *ENOD40* mRNA interactions with the nuclear RNA-binding protein MtrBPP1 (Campalans *et al.*, 2004). The interactions are independent of ORF A translation and result in the relocalization of the *ENOD40* RNA–MtrBPP1 complex into the cytoplasm.

To elucidate the molecular mechanism of *ENOD40* activity, the *G. max ENOD40-1* transcript secondary structure was studied (Girard *et al.*, 2003). A phylogenetic analysis and modeling of RNA folding identified six domains, named domains 1–6, of conserved secondary structure which was confirmed experimentally.

The function of the domains is not elucidated but numerous data point to their importance. The presence of domains 1, 2 and 3 in *ENOD40* genes from *Poaceae* and *Solanaceae* families suggests that their function is not limited to symbiotic nitrogen fixation (Gulyaev & Roussis, 2007). The conserved domain 1 is located in a region of biological activity in *M. truncatula ENOD40* mRNA (Sousa *et al.*, 2001). Domain 4, specific for legume plants with indeterminate nodules, is proposed to be involved in the determination of nodule type (Girard *et al.*, 2003).

In this study, we analyzed *Lupinus luteus L. ENOD40* genes, including one coding for the shortest legume *ENOD40* transcript. *L. luteus* is not only phylogenetically distant from the intensively studied legume plants such as *Medicago truncatula*, *Lotus japonicus* and *Glycine max*, but also exhibits peculiarities in symbiotic nodule development. The *Lupinus* nodule meristem is active along nodule ontogeny (persistent), a trait of indeterminate nodulation, but nodules originate from outer cortex cells, which normally form determinate nodules (characterized by non-persistent meristems). Furthermore, bacteria enter nodule primordia without infection threads and meristematic cells are infected with bacteria (Lo-

tocka *et al.*, 2000; González-Sama *et al.*, 2004; Strozycski *et al.*, 2007). A comparison of *L. luteus* ENOD40 expression patterns, genes and transcript structures with those of model legume plants, including *M. truncatula* of indeterminate nodule type and plants of determinate nodules, such as *L. japonicus* and *G. max*, may shed light on ENOD40 transcript structure-function relationships in nodulation.

MATERIALS AND METHODS

Plant growing and harvesting. *Lupinus luteus* L. var. Ventus was grown in 2 l pots with perlite under 16 h day at 23°C. Seed sterilization, seeding and broth composition were according to Strozycski *et al.* (2003). Plants grown under symbiotic conditions were infected with 10 ml of 7-day old *Bradyrhizobium* sp. (*Lupinus*) WM9 culture (YMB broth; Vincent, 1970) at OD₅₀₀ = 1.287 (1.2×10⁹ cfu/ml) days after seeding (4-leaf plants), and nitrate was excluded from the broth. The root sector 2 cm below the hypocotyl and 2 cm above the root tip was collected for early stages of symbiosis. The plant material was frozen in liquid nitrogen within 2 min after harvesting.

dsDNA probe labeling. Double stranded cDNA (100 ng) was ³²P-labeled with the HexaLabel kit (Fermentas) according to the manufacturer's protocol with the incubation at 37°C extended to 30 min. The final product was purified with the QIAquick PCR purification kit (QIAGEN).

Hybridization. Hybond N⁺ membrane (Amersham) was used for library screening and Southern and Northern hybridizations. The nucleic acids transfer followed by fixation at 80°C for 2 h was according to the producer's protocol. The hybridization in solution composed of 5×SSC, 5×Denhardt, 0.5% SDS was as suggested in Hybond N⁺ protocol. The results were read with a Typhoon 8600 phosphorimager and analyzed with ImageQuant software.

Genomic library screening. The *L. luteus* var. Ventus genomic library (EMBL3 vector, average insert length 15 kb) was screened with a fragment of *M. truncatula* ENOD40 cDNA as a probe at 55°C according to the protocol by Stratagene. A total of 4×10⁵ plaque forming units was screened on plates seeded at 260 pfu/cm².

cDNA library screening and cDNA clone isolation. The *L. luteus* var. Ventus cDNA library (UniZAP XR vector, Stratagene; average insert length 1.8 kb) prepared from roots during symbiosis with *Bradyrhizobium* sp. at 7, 14, and 21 days post infection (DPI) was screened with ENOD40-1 as a probe at 65°C according to the protocol by Stratagene. A total of 4×10⁵ pfu was screened on plates seeded at 260 pfu/cm².

Southern hybridization. Five microgram of genomic DNA digested with EcoRV and HindIII (Fermentas) was separated on a 0.7% agarose gel and transferred to Hybond N⁺. The hybridization was carried out at 68°C. The dsDNA probe was PCR-amplified with Taq DNA polymerase (Fermentas), pSK Bluescript targeted primers JP203 CGGCCAGTGAATTGTAATACG and JP204 ATGACCATGATTACGCCAAGC, and 1 ng of cDNA cloned in pSK Bluescript as a template. The PCR product was purified with the QIAquick PCR Purification Kit (QIAGEN).

Northern hybridization. RNA was isolated from frozen plant material according to Logemann *et al.* (1987). Formaldehyde/agarose gel electrophoresis (Sambrook & Russell, 2002) was used to analyze up to 5 µg of RNA per lane. The gel was either stained in 0.5 µg/ml ethidium bromide for RNA visualization or processed to transfer RNA on Hybond N⁺ which was then hybridized at 68°C using the same procedures as for Southern hybridization.

Semi-quantitative RT-PCR analysis of ENOD40 gene expression. First strand cDNA was synthesized with 1 µg of total RNA as a template, oligo (dT) primer and RevertAid M-MuLV Reverse Transcriptase (Fermentas) according to the producer's protocol. One microliter of the reverse transcription reaction was used as a template for PCR amplification with 1 U Taq DNA polymerase (Fermentas) in a final volume of 40 µl; the amplification reaction contained 1.6 mM MgCl₂, 0.2 mM dNTP, 0.4 µM forward primer, 0.4 µM reverse primer, and 1×Taq buffer (Fermentas). The primers for ENOD40-1 amplification were JP4-05 AGAGACTGTTTATAATATTA ACTG and JP4-06 GATATGCATACAAAAA-CATTC; the primers for ENOD40-2 were JP4-07 CT-CATAACTCATAAGGGCATAA and JP4-08 GGGA-TAGCATACATAGTATTAA. The amplification program was one cycle at 94°C for 1 min, three cycles of 94°C for 40 s, 53°C for 1 min, and 72°C for 1 min, followed by 21 cycles of 94°C for 40 s, 51°C for 1 min, and 72°C for 1 min, and a final extension at 72°C for 5 min. A 10 µl aliquot of the reaction was analyzed on a 0.7% agarose gel stained in 0.5 µg/ml ethidium bromide after electrophoresis. The gel was scanned with Typhoon 8600 and ImageQuant software was used for quantitative analysis.

Analysis of translational efficiency of selected ORFs by transient expression. Constructs for analysis of translation efficiency of ENOD40-1 ORFs were generated in the pSK Bluescript vector which contained the CaMV 35S promoter, GUS coding sequence (*uidA* gene) and NOS terminator from pBI121. Restriction enzyme digestion (*Pst*I, *Bst*XI, *Eco*RI, *Bam*HI, *Xba*I), dephosphorylation and ligation were performed according to Fermentas. *Pwo* DNA polymerase (Promega) was used for DNA amplification and blunt-ending.

The construct pTE20 comprised the CaMV promoter, GUS coding sequence without the starting codon, and NOS terminator. The remodeled spacer between the CaMV promoter and ATG-less GUS coding sequence containing a *Bam*HI and an *Xba*I site proximal and distal to the CaMV promoter, respectively, allowed the directional insertion of the *ENOD40-1* cDNA fragment.

The pTE30 construct contained the CaMV 35S promoter followed by the 5' UTR and a fusion of ORF A and the ATG-less GUS coding sequence.

The 1–86 bp fragment of *L. luteus ENOD40-1* cDNA was amplified with the forward primer te01 **ATGGATCCTGGAAATTCTCCAAAACCA** (*ENOD40-1* non-specific region in bold, *Bam*HI site underlined) and reverse primer te02 **ATTCTAGAACCATGGATGGATTTTTGC** (*ENOD40-1* non-specific region in bold, *Xba*I site underlined), and cloned into the *Bam*HI and *Xba*I sites of pTE20 to generate the pTE30 construct.

The pTE32 construct contained the CaMV 35S promoter followed by ORF A without the 5' UTR, and the ATG-less GUS.

The *ENOD40-1* ORF A 51–86 bp fragment was amplified with the forward primer te09 **ATGGATC-CATGGA**ACTCTCTTGGCAA (*ENOD40-1* non-specific region in bold, *Bam*HI site underlined) and the reverse primer te02, and cloned into the *Bam*HI and *Xba*I sites of pTE20, to generate the pTE32 construct.

The pTE40 construct contained the CaMV 35S promoter followed by ORF B with its complete upstream region and ATG-less GUS.

The 1–90 bp fragment of *ENOD40-1* was amplified with the forward primer te01 and reverse primer te03 **ATTCTAGATTCAAGA**ACCATG-GATGGA (*ENOD40-1* non-specific part in bold, *Xba*I site underlined), and cloned into *Bam*HI and *Xba*I sites of pTE20 to produce the pTE40 construct.

The pTE50 construct contained the CaMV 35S promoter followed by the *ENOD40-1* cDNA region 1–229 bp, ORF 330 (named according to the position of starting codon), and ATG-less GUS.

The 1–404 bp fragment of *ENOD40-1* cDNA, containing the ORF 330 and its upstream region, was amplified as above with the forward primer te01 and reverse primer te04 **ATTCTAGACATACAT-TCTACTGCAACA** (*ENOD40-1* non-specific region in bold, *Xba*I site underlined), and cloned into the *Bam*HI and *Xba*I sites of pTE20 to generate pTE50. The ORF 330–329 bp long upstream region included ORF A and ORF B.

Preparation of tungsten beads coated with DNA. Tungsten beads (15 mg) washed with 0.1 N nitric acid, deionized water and 96% ethanol were suspended in 1.0 ml of deionized water. A total of 150 µl of the suspension was combined with 30 µl of the construct (1 µg/µl DNA concentration), vortexed

and mixed. The DNA coated beads were combined with 150 µl of 2.5 M calcium chloride, vortexed, supplemented with 60 µl of 0.1 M spermidine and vortexed again for 5 min. The DNA coated tungsten beads were spun down, washed with 96% ethanol and resuspended in 180 µl of 96% ethanol.

Plant transformation by leaf bombardment.

Nicotiana tabacum leaves (grown *in vitro*) were bombarded with 10 µl of the DNA-coated tungsten beads using 0.5 MPa of helium at a 5 cm distance between the cartridge and the leaves, two shots per leaf. The leaves were grown for two more days and then stained in a solution of 1 mM X-GlcA (5-bromo-4-chloro-3-indolyl β-D-glucuronide cyclohexylammonium salt, Sigma), 0.1 M Na₂HPO₄, 0.1 M NaH₂PO₄, 0.01 M EDTA, 0.5 mM K₄Fe(CN)₆, 0.5 mM K₃Fe(CN)₆, and 0.1% Triton X-100 for 24 h at 37°C. Following the staining, the leaves were washed with 50% and 96% ethanol.

Probing RNA structure with Pb(II)-induced cleavage

T7 RNA polymerase template and RNA synthesis. The *ENOD40-1* cDNA was cloned in pSK Bluescript with the 5' end proximal to the T7 RNA polymerase promoter and the 3' end at the *Xba*I site to produce the pHG2 construct. The 8 bp spacer ATTA CGAG separated the T7 RNA polymerase promoter from the 5' end of the insert. Two micrograms of linearized pHG2 DNA (*Xba*I digestion followed by proteinase K treatment and QIAquick PCR purification) was used as a template for RNA synthesis with the T7 RNA polymerase (Fermentas). The reaction was composed of 0.3 U/µl of T7 RNA polymerase, 2 mM NTP, and 0.5–1.25 U/µl ribonuclease inhibitor (Fermentas), incubated for 3 h at 37°C, and stopped with EDTA pH 8.0, which was added to a final concentration of 20 mM. RNA was precipitated with ethanol and dissolved in water. The analyzed RNA lacked the poly A tail, as T7 RNA polymerase often produces populations of truncated molecules on a template with a homo-polymeric region at the end.

Pb(II)-induced cleavage and primer extension by reverse transcriptase. RNA concentration was evaluated measuring absorbance (with spectrophotometer or phosphorimager Typhoon); low RNA concentrations mentioned below were obtained as dilutions from samples of higher concentration. *ENOD40-1* RNA (0.8 µM) was supplemented with carrier tRNA to a total RNA concentration of 4 µM, incubated at 65°C for 3 min in the 2× reaction buffer, and cooled to 25°C. An equal volume of lead(II) acetate solution of two times higher concentration than the final concentration was added and the reaction was performed in the mixture composed of (final concentrations is given for all com-

ponents) 10 mM Tris/HCl, pH 7.2, 40 mM NaCl, 10 mM MgCl₂, 0.4 μM of *ENOD40-1* RNA, 1.6 μM of carrier tRNA and 0.0 mM, 0.5 mM, 1 mM or 2 mM Pb(II) ions. The Pb(II)-induced cleavage reactions were performed with increasing concentrations of Pb(II) ions to saturate all possible Pb(II) binding sites. The reaction mixture was incubated at 25°C for 10 min, stopped with EDTA added to a final concentration of 5 mM, precipitated with ethanol and dissolved in water. The primer extension reaction used to determine positions of Pb(II)-induced breaks was carried out 30 min at 42°C in a final volume of 10 μl in 50 mM Tris/HCl, pH 8.3, 75 mM KCl, 10 mM DTT, 3 mM MgCl₂, 1 mM dNTP, 50 U M-MLV (Promega), 1 μM ³²P 5' end labeled reverse transcription primer (see below), usually 2–5 × 10⁵ c.p.m. ³²P, and 1.0 μM of total cleaved RNA (total RNA is composed of *ENOD40* transcript and carrier tRNA). The reaction was stopped with an equal volume of 7 M urea, 20 mM EDTA. Sequencing reactions were carried out in parallel to facilitate interpretation of the primer extension results; in each of the four sequencing reactions, one of the dideoxynucleotides was added to a final concentration of 0.025 mM and the other components of the reactions were as above. The reverse transcription primers were as follows: RT-1 ATGCGTGCTTATTCAAGAAC, RT-2 TGGTGATTAGAGAAGCCAATA, RT-3 TGGAGTCAAGCCTTTTTGTG, RT-4 GTAACATCTCAAAGGAGTGC and RT-5 ATACAAAGAGACTGTTATAATAT-TAACTG. The last 24 nucleotides at the 3' end of the *ENOD40-1* transcript were not accessible for the analysis due to interaction with the reverse transcriptase primer. The cleavage sites were detected by separating the products of the reverse transcriptase primer extension reaction on a gel along with the products of dideoxy sequencing reactions. The electrophoresis was run on 12% polyacrylamide, 7 M urea, 1 × TBE.

RNA structure modeling. *ENOD40-1* RNA structure prediction was performed with the Zuker/Turner mfold program version 3.1 (Zucker, 2003; <http://www.bioinfo.rpi.edu/applications/mfold/>). The standard default parameters were used during the modeling, with the exception of forcing domain 1 (nt 109–137), domain 2 (nt 153–288), domain 3 (nt 310–327) and domain 6 (nt 440–497) to form the structures proposed by Girard *et al.* (2003). Models were validated with the following data: Pb(II)-induced cleavage of *LIENOD40-1* transcript (see above), *LIENOD40-1* regions corresponding to *G. max ENOD40-1* single stranded regions (26–31), (78–84), (332–337) and (355–336) (Girard *et al.*, 2003), and pairing between *ENOD40* conserved regions.

Phylogenetic analysis. Phylogenetic analyses were conducted in MEGA4 (Tamura *et al.*, 2007). The accession numbers of the 23 analyzed *ENOD40*

sequences, including *L. luteus ENOD40-1* and *ENOD40-2*, are given in Table 1. *G. max ENOD40-2* (X86442) fragment 1744–2960 bp and *L. japonicus ENOD40* (AF013594) fragment 792–1561 bp were taken for analysis, while the rest of the sequences were used without trimming. The *ENOD40* evolutionary history was inferred using the Neighbor-Joining method (Saitou & Nei, 1987), and the optimal tree was generated with a bootstrap test using 500 replicates (Felsenstein, 1981). The evolutionary distances were computed using the Maximum Composite Likelihood method. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 183 positions in the final dataset.

RESULTS

Two *ENOD40* genes are present in the *L. luteus* genome and contain a 10-nt promoter-proximal region conserved in *Papilionoideae* and *Poaceae*

We screened a *L. luteus* genomic library and isolated two genomic clones containing *ENOD40* genes and their promoter regions, which we named *ENOD40-1* (AF352372.1) and *ENOD40-2* (AF352373.1). Moreover, screening cDNA library resulted in six cDNA clones — four of them represented *ENOD40-1* (AF352375), the shortest legume *ENOD40* gene, and two others represented *ENOD40-2* (AF352374). The *L. luteus ENOD40* cDNAs are 85% identical to one another and possess two conserved regions, I and II. The conserved region I contains ORF A encoding a 12 amino-acid peptide (MELSWQKSIHGS in *ENOD40-1* and MKLFWQKSINGS in *ENOD40-2*) and the first three codons of the extremely short ORF B, which is composed of four (*ENOD40-1*) or five (*ENOD40-2*) codons.

Promoter analysis revealed a unique 90 bp region conserved in *L. luteus ENOD40-1* and *ENOD40-2* located directly upstream of the *ENOD40-1* putative transcription start site. The ACTA element, a 10-nucleotide motif proximal to the transcription start site with the consensus sequence of TTC(T/C)CCACTA, is also present in the *ENOD40* genes of *Papilionoideae* and *Poaceae*. Putative TATA boxes in the conserved region are 40 bp upstream of the ACTA element. Promoters of both genes have regions highly similar to distinct regions of *G. max ENOD40-2* promoter, located several hundred base pairs upstream of the ACTA element, and inverted (*ENOD40-1*) or direct (*ENOD40-2*) repeats situated around 1.5 kb above the ACTA element.

The *ENOD40-1* and *ENOD40-2* cDNA probes hybridized to two genomic DNA fragments: a long-

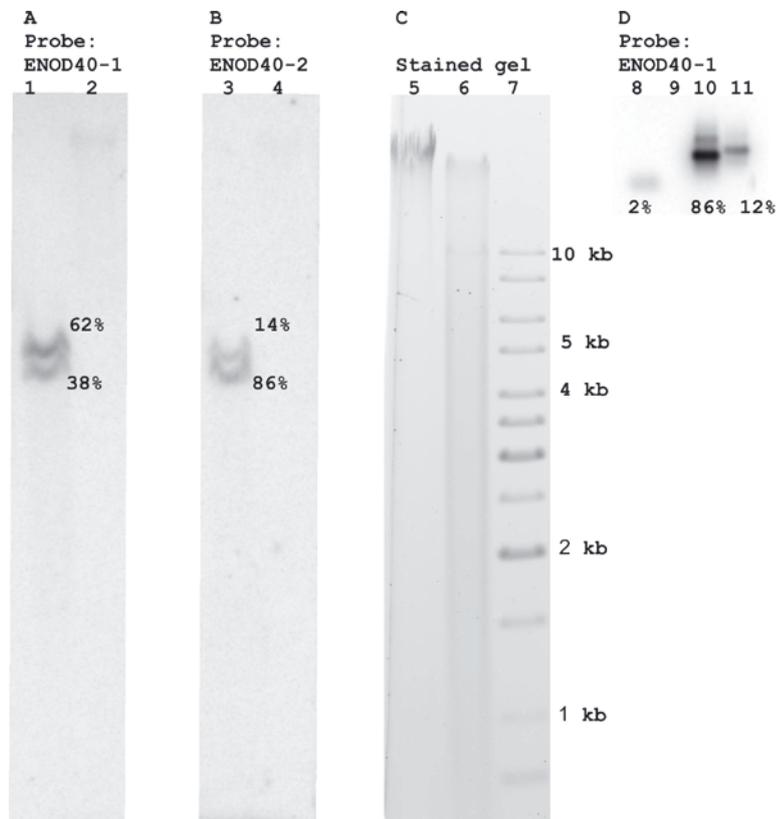


Figure 1. Southern analysis of *Lupinus luteus* ENOD40-1 and ENOD40-2 genes.

Five micrograms of genomic DNA was loaded per lane, transferred to a membrane, and the membrane was hybridized with the ENOD40 cDNA probe at 68°C (high stringency conditions). The distribution of radioactive probe between the bands on each blot is shown as percentage. **A.** Genomic DNA was digested with *EcoRV* and *HindIII* (lane 1) and the complete cDNA of ENOD40-1 was used as a probe. Undigested genomic DNA was run in lane 2. **B.** Genomic DNA was digested with *EcoRV* and *HindIII* (lane 3) and the complete cDNA of ENOD40-2 was used as a probe. Lane 4 contained undigested genomic DNA. **C.** The ethidium bromide stained gel of undigested genomic DNA (lane 5), genomic DNA digested with *EcoRV* and *HindIII* (lane 6), and 100 ng of the 1 kb Fermentas marker (lane 7). **D.** The complete cDNA of ENOD40-1 was used as a probe against the most specific region of ENOD40-1 (3' end fragment of cDNA; 50 ng, lane 8), the most specific region of ENOD40-2 (3' end fragment of cDNA; 50 ng, lane 9), the complete ENOD40-1 cDNA (50 ng, lane 10), and the complete ENOD40-2 cDNA (50 ng, lane 11).

er fragment (4.6 kb) that hybridized more efficiently with the ENOD40-1 probe, and a shorter one (4.0 kb) that hybridized more intensely with the ENOD40-2 probe (Fig. 1). Attempts of Southern analysis of genomic DNA with ENOD40-1 and ENOD40-2 specific fragments were unsuccessful, most likely due to the length of the fragments (49 bp and 121 bp), which did not support efficient hybridization. Hybridization with the short ENOD40-1 specific fragment produced a signal forty times less intense than the full-length ENOD40-1 cDNA (Fig. 1D). Hybridization of the same amount (50 ng) of both cDNAs with the whole cDNA of ENOD40-1 (Fig. 1D) and ENOD40-2 (not shown) proved their cross-hybridization and revealed differences between the efficiency of hybridization and cross-hybridization (Fig. 1D). The Southern hybridization results, together with control of the efficiency of cDNAs cross-hybridization, suggest that there are only two genes of high nucleotide similarity to ENOD40-1 and ENOD40-2 in the *L. luteus* genome.

ENOD40 phylogeny clusters *L. luteus* genes with genes of plants with determinate nodules

We analyzed the phylogenetic relationships between *L. luteus* ENOD40-1 and ENOD40-2 cDNA sequences and 21 legume ENOD40 homologues (accession numbers in Table 1). The *L. luteus* genes ENOD40-1 and ENOD40-2 are paralogs clustered

within group I (plants forming determinate nodules), based on transcriptional unit phylogeny (Fig. 2). The position of the *L. luteus* genes within group I is not clear, but, importantly, the results indicate that the *L. luteus* genes are clustered with genes from plants with determinate nodules and are distinct from group II (genes from plants with indeterminate nodules), whereas *L. luteus* nodules are of persistent meristem. The specificity of *Lupinus* nodules, in which persistent meristem seems to be of different origin than that of plants from group II, such as *Medicago*, *Pisum* or *Trifolium*, supports the ENOD40 phylogeny (Golinowski *et al.*, 1987; Lotocka *et al.*, 2000; González-Sama *et al.*, 2004; Strozycki *et al.*, 2007). In addition, ENOD40 of *Sesbania rostrata*, a plant showing intermediate type of nodules, is clustered within group I (Ndoye *et al.*, 1994; Goormachtig *et al.*, 1997).

The ORF A and ORF B peptides from the ENOD40 genes reveal similarities in amino-acid sequence and length within the clusters obtained from the ENOD40 transcriptional unit phylogeny (Table 1). Also, the higher diversity of group I compared to group II is demonstrated by peptide A and B analysis (Table 1) as well as transcriptional unit phylogeny (Fig. 2). The distant position of *T. repens* ENOD40-3 and *M. truncatula* ENOD40-2 is well established by both peptide comparison and gene phylogeny. The nucleotide diversity of the two genes opens the possibility that similar ENOD40 genes may be also present in other legume plants.

Table 1. Papilionoideae ENOD40 peptides A and B.

ENOD40 peptides encoded by two conserved ORFs A and B are clustered according to phylogeny of transcriptional units (Fig. 2) and reveal similarities within the clusters. *L. luteus* peptides are clustered with peptides from *Papilionoideae* of determinate nodules (group I), although the lupin nodule is of the indeterminate type. *T. repens* ENOD40-3 and *M. truncatula* ENOD40-2 are distinct from groups I and II. Note that conservation of the first three amino acids of ORF B is a result of the two ORFs overlapping.

Group I, plants with determinate nodules		
Accession number, organism, clone	ORF A peptide of 12 amino acids	ORF B peptide of 4, 5, 8, 10, 11, 24, 30 amino acids
X69154, <i>G. max</i> , ENOD40-1	MELCWLTTIHGS	MVLEEAWRERGVREGGAHSSSHSLT
X86442, <i>G. max</i> , ENOD40-2	MELCWLTTIHGS	MVLEEAWRERGVREGGAHSSSHSLT
D13504, <i>G. max</i> , GmN36b	MELCWLTTIHGS	MVLEEAWRERGVREGGAHSSSHSLT
D13503, <i>G. max</i> , GmN36a	MELCWQTSIHGS	MVLEEAWR
X86441, <i>Ph. vulgaris</i> , ENOD40	MKFCWQASIHGS	MVLKKGHEKGV
AF061818, <i>V. radiata</i> , ENOD40	No seq.	No seq.
Y12714, <i>S. rostrata</i> , ENOD40	MKLCWQKSIHGS	MVLKETWEEV
AF013594, <i>L. japonicus</i> , ENOD40	MRFCWQKSIHGS	MALEEAWRGMCEVRVVFHPSHTILCILQFV
AJ271788, <i>L. japonicus</i> , ENOD40-2	MRFCWQKSIHGS	MALEEAWRGMCEVRVVFHPSHTLLCILQFV
AJ271787, <i>L. japonicus</i> , ENOD40-1	MKLCWQISIHGS	MVLK
AF352375, <i>L. luteus</i> , ENOD40-1	MELSWQKSIHGS	MVLE
AF352374, <i>L. luteus</i> , ENOD40-2	MKLEWQKSINGS	MVLDS
Group II, plants with indeterminate nodules		
Accession number, organism, clone	ORF A peptide of 13 amino acids	ORF B peptide of 14, 15 amino acids
AJ000268, <i>T. repens</i> , ENOD40	MKLLCWQKSIHGS	MVLKTNMERSVRGLY
AF426839, <i>T. repens</i> , ENOD40-2	MKLLCWQKSIHGS	MVLKTNMERSVRGLY
AF426838, <i>T. repens</i> , ENOD40-1	MKLLCWQKSIHGS	MVLKTNMERSVRGLY
X83683, <i>V. sativa</i> , ENOD40	MKLLCWQKSIHGS	MVLKIKHGEKCERVN
X81064, <i>P. sativum</i> , ENOD40	MKFLCWQKSIHGS	MVLKNKYGVK CERAN
L32806, <i>M. sativa</i> , ENOD40	MKLLCWQKSIHGS	MVLKTNMERSVRGY
X80263, <i>M. sativa</i> , ENOD40	MKLLCWQKSIHGS	MVLKTNMERSVRGY
X80264, <i>M. truncatula</i> , ENOD40	MKLLCWQKSIHGS	MVLKTNMERSVRGY
X80262, <i>M. truncatula</i> , ENOD40	MKLLCWQKSIHGS	MVLKTNMERSVRGY
Distant ENOD40 genes		
Accession number, organism, clone	ORF A peptide of 11, 12 amino acids	ORF B peptide of 8, 28 amino acids
AF426840, <i>T. repens</i> , ENOD40-3	MDLCWQKSIHGS	MVLKGRMEKLCERVLMTLHTLHLQQFA
<i>M. truncatula</i> ENOD40-2 (Wan <i>et al.</i> , 2007)	MNLCWQKSIYD	

Analysis of the 23 legume ENOD40 sequences for conserved regions identified eight motifs (position in bp according to *L. luteus* ENOD40-1): motif 51–91 corresponds to conserved region I, motif 150–157 at the basis of domain 2, motif 172–181 within domain 2, motifs 282–312 and 330–335 within conserved region II, and motifs 424–433, 447–464 and 480–493 within domain 6. Each of the motifs has more than 60% conserved positions (occupied by the same nucleotide) when analysed within group I or group II. The most conserved motifs are the short motif 150–157, of eight out of nine positions conserved in genes from both groups I and II, and motif 330–335 of all six residues conserved in all legume ENOD40 genes except for those from *S. rostrata*. A function may be assigned to some of the motifs based on the transcript structure: domain 2 depends on interactions of motifs 150–157 and 172–181 with the conserved region II, similarly, motif 447–464 interacts with motif 480–493 to form domain 6. The role of the other motifs is not yet clear, but they

may be useful for the transcript secondary structure validation.

ENOD40-1 is only expressed in symbiotic root nodules whereas ENOD40-2 is also expressed in non-symbiotic organs

Northern hybridization analyses showed that *L. luteus* ENOD40 genes were expressed at early stages of symbiosis starting from the third day post-infection (3 DPI) with *Bradyrhizobium* sp. until 14 DPI, the oldest tested stage of symbiosis. The highest level of ENOD40 transcripts was found in the mature nodules at 14 DPI (Fig. 3A–C). A decrease of ENOD40 gene expression was observed between the third and eighth day after infection. As the Northern technique could not distinguish the two *L. luteus* ENOD40 genes due to cross-hybridization of the probes (see Fig. 1D), the expression pattern of ENOD40 genes was confirmed by reverse transcription followed by PCR amplification, in which

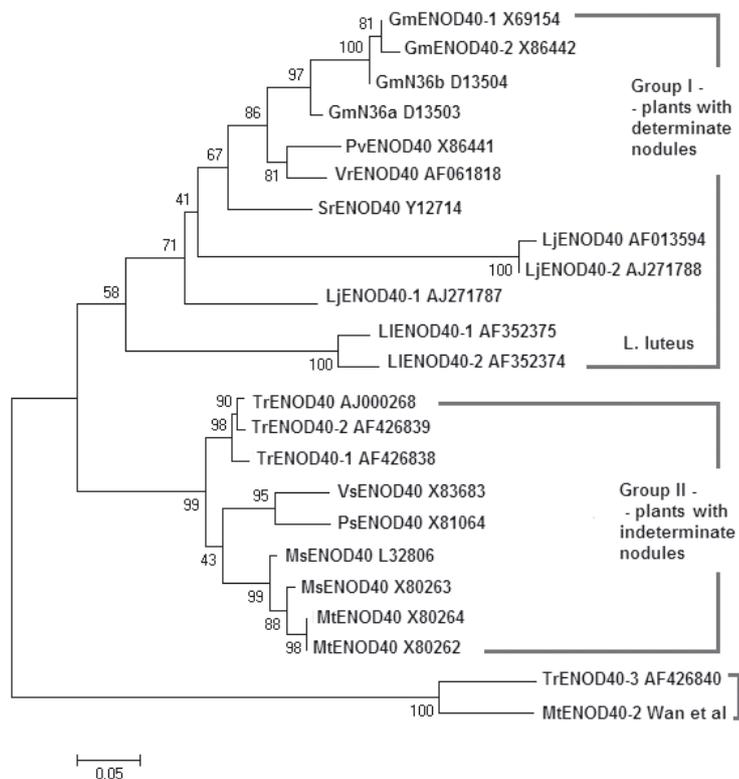


Figure 2. Evolutionary relationships of *ENOD40* genes.

The evolutionary history of *ENOD40* genes was inferred using the Neighbor-Joining method. The optimal tree with a branch length of 1.92971421 is shown. The percentages of replicate trees in which the associated taxa cluster together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale with branch length in units of the number of base substitution per site. *L. luteus* *ENOD40-1* and *ENOD40-2* are clustered within group I, plants of determinate nodules, similar to *S. rostrata* *ENOD40*. The genes are labeled with accession numbers and abbreviations: Mt, *M. truncatula*; Ms, *M. sativa*; Tr, *T. repens*; Vs, *V. sativa*; Ps, *P. sativum*; Lj, *L. japonicus*; Ll, *L. luteus*; Sr, *S. rostrata*; Gm, *G. max*; Pv, *P. vulgaris*; Vr, *V. radiata*; the list of the full names of the genes with species and accession numbers is given in Table 1.

gene-specific primers allowed the specific analysis of *ENOD40-1* and *ENOD40-2* (Fig. 3D–F). The results showed that both genes exhibited the same expression pattern during development of the symbiotic organ. The higher sensitivity of RT-PCR allowed the detection of very low levels of *ENOD40-2* transcripts in uninfected roots at 14 DPI and in root tips (Fig. 3E). Notably, *ENOD40-1* transcripts were not detected in uninfected roots or in shoot tips, thus expression in non-symbiotic organs differentiated the *L. luteus* *ENOD40* genes.

Reporter gene fused to *ENOD40-1* ORF A and B, but not to the longest ORF, is translated

The translation of *ENOD40-1* ORF A (start codon at position 51) that encodes the 12 amino-acid peptide A, ORF B (start codon at position 79) that encodes the 4 amino-acid peptide B, and ORF 330, the longest ORF of 25 codons (start codon at position 330) was tested in *N. tabacum* leaves using a transient expression assay. ORF A and ORF B are the two first ORFs of *L. luteus* *ENOD40-1* with start codons in different reading frames; ORF 330 is not only the longest ORF of *ENOD40-1*, but its start codon is located in motif 330–335, which is conserved in all legume *ENOD40* genes apart from *S. rostrata*. The experimental constructs contained the specific *ENOD40-1* fragments fused to an ATG-less GUS coding sequence under the control of the CaMV 35S promoter. Removing the start codon from the

reporter gene coding sequence eliminated potential false positive results, as the observed GUS activity was proof of translation starting from the ORF, and not GUS, start codon.

The results demonstrated high expression of ORF A, with efficiency depending on the 5' UTR (Fig. 4A, B). The construct including the short ORF B showed much lower expression (Fig. 4C). No expression of ORF 330 was observed under these experimental conditions (Fig. 4D). Of the three analyzed ORFs, ORF A starting codon is in a context (length of the upstream region, sequence composition in the vicinity of the codon, structural elements, etc.) promoting efficient initiation of translation. Also, the relationship between the 5' UTR and translation efficiency suggests that ORF A serves as a template for peptide synthesis in the native *ENOD40-1* mRNA molecule.

The secondary structure of the *ENOD40-1* transcript

The secondary structure of the *ENOD40-1* RNA was analyzed using Pb(II)-induced cleavage. This method is well suited for probing RNA structure, as Pb(II) ions preferentially induce RNA chain breakages in single-stranded regions of high flexibility (Ciesiolka *et al.*, 1998). The location of all Pb(II)-induced cleavage sites within the *ENOD40-1* transcript is displayed in Fig. 5, and the radiogram of region 165–290 (domain 2) is shown in Fig. 6. Apart from many observed weak cleavages, three strongly

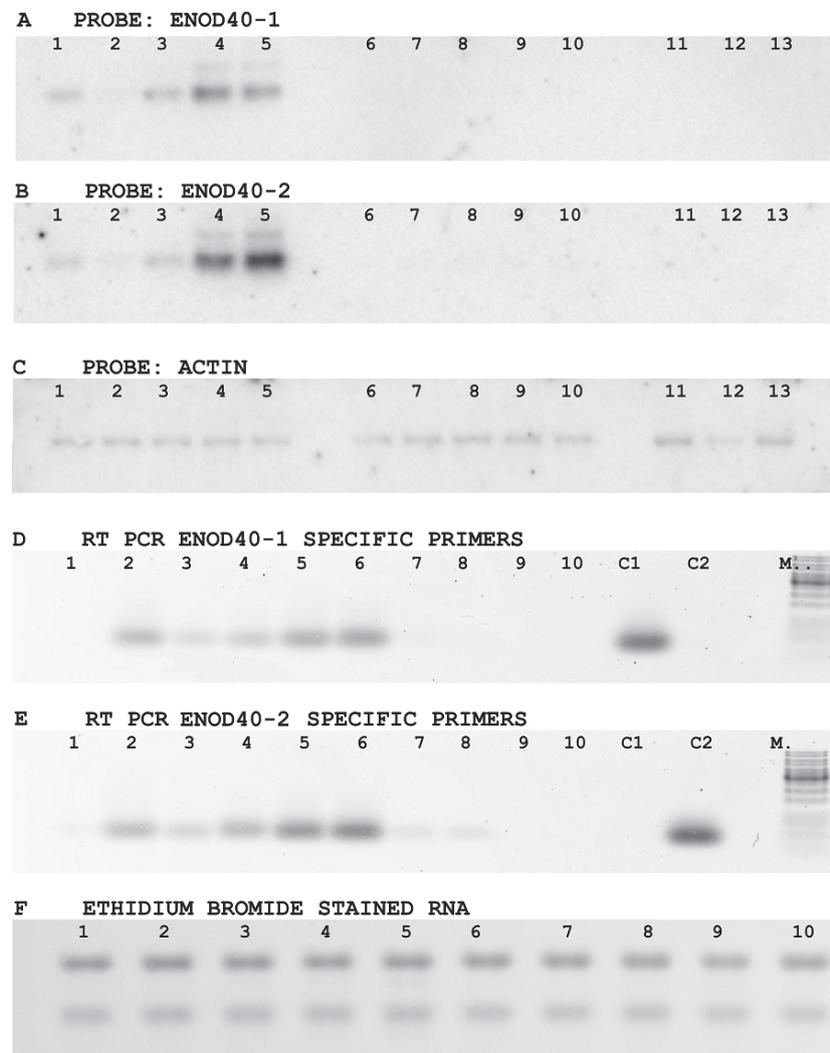


Figure 3. Expression of *L. luteus* ENOD40-1 and ENOD40-2.

Panels A, B and C show Northern hybridization with 5 μ g of total RNA per lane: lane 1, infected plant 3 DPI; lane 2, infected plant 5 DPI; lane 3, infected plant 8 DPI; lane 4, infected plant 11 DPI; lane 5, infected plant 14 DPI; lane 6, uninfected plant 3 DPI; lane 7, uninfected plant 5 DPI; lane 8, uninfected plant 8 DPI; lane 9, uninfected plant 11 DPI; lane 10, uninfected plant 14 DPI; lane 11, root tip (0.5 cm); lane 12, shoot tip (0.5 cm); and lane 13, inflorescence. The complete cDNA of ENOD40-1 and ENOD40-2 was used as a probe in A and B, respectively; hybridization with actin (C) confirmed equal loading of RNA samples. (D–F) RT-PCR analysis of ENOD40-1 and ENOD40-2 expression with gene-specific primers for ENOD40-1 (D) and ENOD40-2 (E); lane 1, uninfected plant 3 DPI; lane 2, infected plant 3 DPI; lane 3, infected plant 5 DPI; lane 4, infected plant 8 DPI; lane 5, infected plant 11 DPI; lane 6, infected plant 14 DPI; lane 7, uninfected plant 14 DPI; lane 8, root tip (0.5 cm); lane 9, shoot tip (0.5 cm); lane 10, inflorescence; lane C1, 0.05 ng of ENOD40-1 cDNA as PCR template; lane C2, 0.05 ng of ENOD40-2 cDNA as PCR template; lane M, 1 kb Fermentas marker, 100 ng. (F) Equal RNA amount used for reverse transcription was confirmed by ethidium bromide staining; lanes 1 to 10 as on Fig. 3D and E, with 0.5 μ g of RNA per lane.

cleaved sites are located within domain 2 at positions C188, G255 and C256 (Fig. 6). Such efficient and highly specific cleavages often occur at strong Pb(II) binding sites, such as, for example, in yeast tRNA^{Phe} (Ciesiolka *et al.*, 1998). Other metal ions, including Mg(II), may occupy the same binding sites and are displaced by Pb(II) under experimental conditions (Kirsebom & Ciesiolka, 2005). Thus, we suggest that domain 2 of ENOD40-1 contains the main metal ion binding sites, which also includes structural Mg(II) ions.

The modeling of the ENOD40-1 structure with fixed domains 1, 2, 3, and 6 (Girard *et al.*, 2003) yielded nine models of free energy ranging from -114.95 to -122.54 kcal/mol.

Domain 5 was not present in any of these models. The models were scored according to the following criteria: number of Pb(II)-induced cleavage sites in double stranded regions without cleavage on the complementary strand (exclusively regions outside the domains) and number of paired nucleotides in regions A35–U40, G88–G94, U376–G381 and

U399–A410 corresponding to *G. max* ENOD40-1 single-stranded regions (Girard *et al.*, 2003). The model with the lowest number of nucleotides matching the above criteria had a dG of -118.34 kcal/mol (Fig. 7). The distribution of nucleotides accessible to Pb(II) digestion within domains revealed single stranded regions, bulges or regions of weak structure. Domains 3 and 6 with stable stems and loops or bulges accessible to Pb(II) ion digestion provide the best evidence of the correspondence between Pb(II) ion digestion results and the computational folding of RNA. The stem-loop structure in region 371–420 between domains 3 and 6 is weak, likely due to a large internal loop. The efficient Pb(II)-induced digestion of this region is in agreement with *G. max* ENOD40-1 data, as the structure contains two partially single stranded regions (376–381 and 399–410) corresponding to the single stranded regions of *G. max* ENOD40-1 (332–337 and 355–366) (Girard *et al.*, 2003).

The global folding of the *L. luteus* ENOD40-1 RNA model involves long distance interactions be-

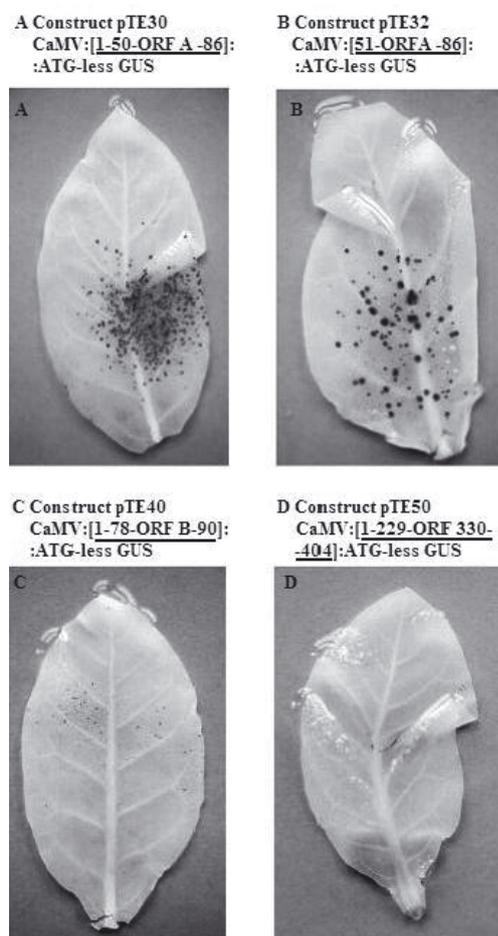


Figure 4. Translational activity of *ENOD40-1* ORFs in *N. tabacum* leaves.

The translational activity of *ENOD40-1* ORF A, ORF B, A and ORF 330 (see Fig. 6) was analyzed with a transient expression assay; the constructs pTE30–pTE50 are composed of *ENOD40-1* fragment with the analyzed ORF fused in frame with ATG-less GUS under the control of the 35S CaMV promoter. **A.** Expression of the pTE30 construct with *ENOD40-1* fragment 1–86 bp (ORF A with 5' UTR). **B.** The pTE32 construct with *ENOD40-1* fragment 51–86 bp (ORF A without 5' UTR). **C.** The pTE40 construct with *ENOD40-1* fragment 1–90 bp (ORF B with the upstream region); and **D.** the pTE50 construct with the *ENOD40-1* fragment 1–404 (ORF 330 with the upstream region); the *ENOD40* fragments are underlined in brackets.

tween the conserved region I and region situated between domains 3 and 6. This interface between conserved region I and the distal part of the *L. luteus ENOD40-1* RNA molecule is rich in nucleotides resistant to Pb(II)-induced cleavage and self-degradation: G64-G, U80-G-G, A93-G, G306-C, A343-U-G-G, and C425-C. Such nucleotides are characteristic for double stranded, stable regions of strong structure, such as stem base. The 330–335 region is a unique single stranded region composed of nucleotides resistant to digestion, which suggests that the motif might be involved in long distance interactions responsible for the transcript tertiary structure forma-

tion. The high conservancy of motif 330–335 reinforces such hypothesis. Also, pairing between conserved region I and one of the motifs conserved in legume *ENOD40* genes (motif 424–432) supports the *L. luteus ENOD40-1* model, which shows the conserved region I interacting with a region upstream of domain 6.

DISCUSSION

Here we present an analysis of the *L. luteus ENOD40* genes including expression patterns, translational capability and RNA secondary structures. Detailed analysis of *L. luteus ENOD40* genes has revealed regions within the promoters to be structurally similar to the *G. max ENOD40-2* promoter, repeats and 89 bp conserved fragment just upstream of the transcription start site. The conservation of this region upstream of the transcription start site correlates with the phylogeny of the *ENOD40* transcriptional units; even evolutionarily distant *ENOD40* genes, like the genes from *L. luteus* and *M. truncatula* or *M. sativa*, showed 85% identity within the 41 bp regions upstream of the ACTA element. The ACTA element and upstream conserved region are present in the 185-bp *G. max ENOD40-2* minimal promoter, which is sufficient for expression under symbiotic conditions (Gronlund *et al.*, 2005), supporting the hypothesis regarding the functional importance of these elements. The function of the 10 nt ACTA element, which is present in the promoters of *Papilionideae* and *Poaceae ENOD40* genes, may be associated with transcription start recognition.

The pattern and expression level of both genes are similar during symbiosis, with the highest expression detected in mature nodules. The decrease of expression between 3 and 8 DPI might be associated with abortion of nodule primordia observed at approx. 5 DPI (Lotocka *et al.*, 2000). The low level of *LIENOD40-2* transcripts detected in uninfected roots and root tips is similar to the expression patterns of *M. sativa ENOD40-1* and *ENOD40-2* (Fang & Hirsch, 1998) and *S. rostrata* (Corich *et al.*, 1998).

The high efficiency of *ENOD40-1* ORF A translation indicates that the ORF A starting codon is in a context (sequence composition, position regard mRNA 5' end, structural elements) promoting initiation of peptide synthesis. The ORF B fusion with reporter gene exhibited lower levels of expression, and in the native *ENOD40* transcript the ORF B mRNA may not be translated due to the mRNA structure. The start codon of ORF 330, the longest *LIENOD40-1* ORF, is located within the highly conserved motif 330–335 and is not active for translation initiation. The conservation of the motif might

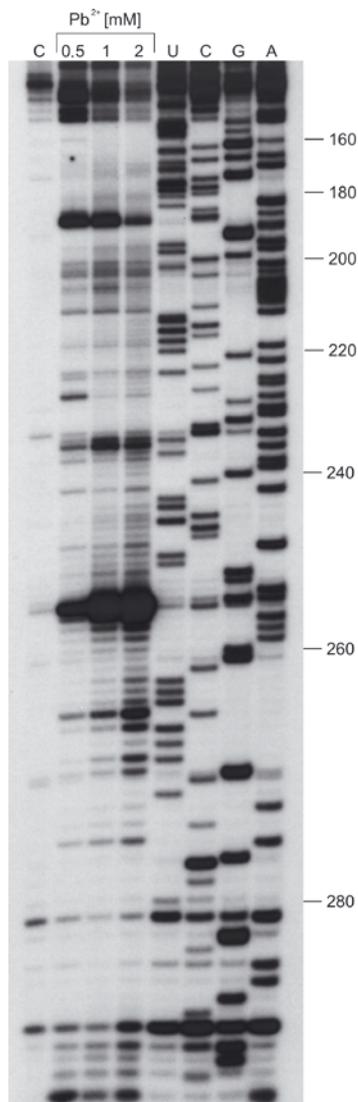


Figure 6. Autoradiogram of Pb(II)-induced cleavage sites in the *ENOD40-1* transcript, region 160–293.

The autoradiogram shows products of Pb(II) cleavage reaction copied on cDNA with reverse transcription primer RT-3 and run along with dideoxy sequencing reactions on 12% polyacrylamide gel; the region 160–293 spanning almost the entire domain 2 contains C188, G255 and C256, the most efficiently cleaved nucleotides. The numbers of positions given on the left of the gel are according to the sequence presented in Figs. 5 and 7. Lane C shows the control without treatment with Pb(II). Lanes marked 0.5, 1, and 2 mM indicate increasing concentration of Pb(II) ions; lanes U, C, G, and A indicate sequencing lanes of thymidine, cytosine, guanine and adenine, respectively.

on well-defined single-stranded regions of *G. max ENOD40-1* (Girard *et al.*, 2003). Two of these regions, *LIENOD40-1* regions 35–40 and 376–381, are single-stranded and the other two, 88–95 and 399–410 are partially single-stranded with almost all nucleotides, even within the double-stranded structures, accessible to Pb(II)-induced cleavage, indicating that the structures are weak. The localization of conserved motifs in legume *ENOD40* transcripts also supports the pro-

posed structure: the 150–157 motif (ACAGUUUG) is positioned opposite a highly conserved motif from the conserved region II at the basis of domain 2, motifs 447–464 and 480–493 are located on the opposite strands of domain 6, and the pairing between motif 58–65 and motif 424–433 is one of the long distance interactions crucial for the *LIENOD40-1* transcript folding (Fig. 7). These long distance interactions responsible for organizing the domains within the *ENOD40-1* mRNA molecule depend on pairing between regions 52–95 and 306–441. Analysis of 24 nucleotide pairs formed between regions 52–95 and 306–441 revealed ten substitutions between *LIENOD40-1* and *LIENOD40-2* that do not disrupt the model, and only slightly decrease the number of paired nucleotides from 24 for *LIENOD40-1* to 21 for *LIENOD40-2*.

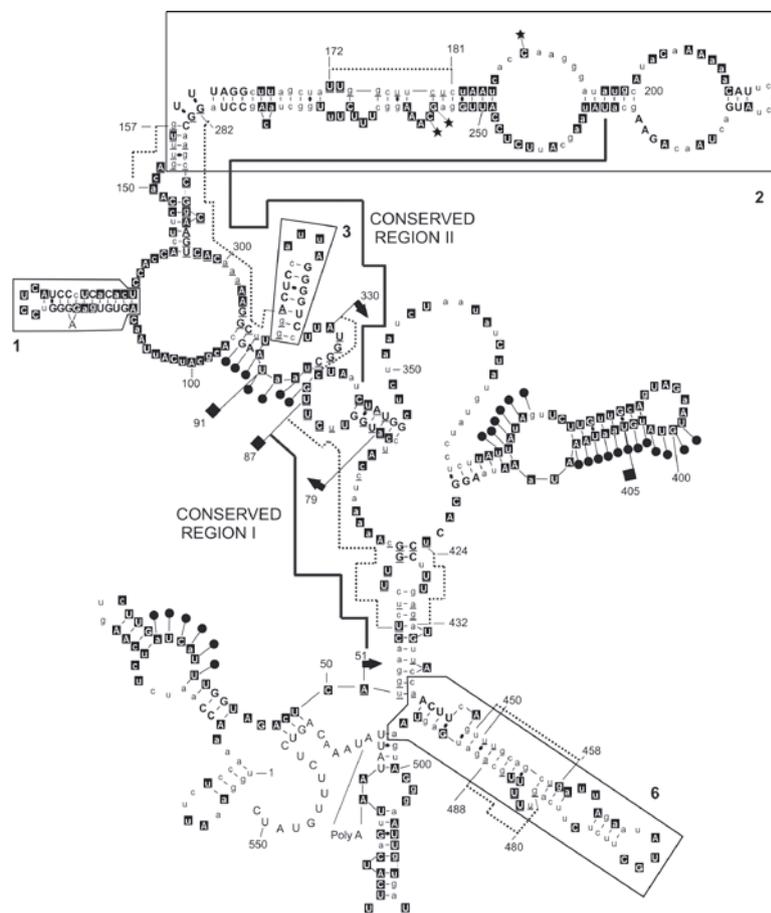
The most important difference between the *LIENOD40-1* and *G. max ENOD40-1* models concerns the interactions of conserved region I, especially at its 5' end. In the case of the *G. max ENOD40-1* model, this region forms hairpins, whereas the *LIENOD40-1* model proposes pairing of the 5' end of the conserved region I with region 425–441 located upstream of domain 6. The *G. max ENOD40-1* regions corresponding to paired *L. luteus ENOD40-1* regions 51–65 and 425–441 are well conserved and might form a 13-nucleotide double-stranded structure (one more than for *L. luteus ENOD40-1*), but in *G. max ENOD40-1* the region corresponding to *L. luteus* region 425–441 is involved in the formation of domain 5.

The differences between the *L. luteus ENOD40-1* and *G. max ENOD40-1* models might result from transcript length, as the *G. max ENOD40-1* transcript is 107 nt (19%) longer than *L. luteus ENOD40-1*, the shortest legume *ENOD40* gene. Another possibility is that *ENOD40* transcripts may switch between two structures depending on cellular conditions, and the two models correspond to these two different states. Furthermore, such two states of the *ENOD40* mRNA molecule might be associated with different activities of the transcript, either functioning as a template for short peptide synthesis or as an RNA of biological function independent of translation.

Acknowledgements

We thank B. Dlugaszewska and G. Nimmagada for help with genomic clone isolation and analysis, I. Femiak for assistance with transient expression, and Dr. Z. Michalski for assistance with preparation of the figures.

This work was funded by the State Committee for Scientific Research, grant P04B 020 23, the Ministry of Science and Higher Education – network “Genomics and transgenesis of crop plants” and grant N 519314635.



REFERENCES

- Asad S, Fang Y, Wycoff KL, Hirsch AM (1994) Isolation and characterization of cDNA and genomic clones of MsENOD40; transcripts are detected in meristematic cells of alfalfa. *Protoplasts* **183**: 10–23.
- Barciszewski J, Legocki A (1997) Two plant signaling peptides: systemin and ENOD 40. *Acta Biochim Polon* **44**: 795–802.
- Campalans A, Kondorosi A, Crespi M (2004) Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell* **16**: 1047–1059.
- Charon C, Johansson C, Kondorosi E, Kondorosi A, Crespi M (1997) enod40 induces dedifferentiation and division of root cortical cells in legumes. *Proc Natl Acad Sci USA* **94**: 8901–8906.
- Charon C, Sousa C, Crespi M, Kondorosi A (1999) Alteration of enod40 expression modifies *Medicago truncatula* root nodule development induced by sinorhizobium meliloti. *Plant Cell* **11**: 1953–1966.
- Ciesiolka J, Michalowski D, Wrzesinski J, Krajewski J, Krzyzosiak WJ (1998) Patterns of cleavages induced by lead ions in defined RNA secondary structure motifs. *J Mol Biol* **275**: 211–220.
- Compaan B, Ruttink T, Albrecht C, Meeley R, Bisseling T, Franssen H (2003) Identification and characterization of a *Zea mays* line carrying a transposon-tagged ENOD40. *Biochim Biophys Acta* **1629**: 84–91.
- Corich V, Goormachtig S, Lievens S, Van Montagu M, Holsters M (1998) Patterns of ENOD40 gene expression in stem-borne nodules of *Sesbania rostrata*. *Plant Mol Biol* **37**: 67–76.

Figure 7. Model of the secondary structure of ENOD40-1 transcript.

The *L. luteus* ENOD40-1 model is one of nine structures generated with the mfold program and selected as described in the text; the structures of domains 1, 2, 3, and 6 are according to Girard *et al.* (2003). The features important for the model validation are marked with an asterisk (the most efficiently Pb²⁺ cleaved nucleotides), capital letters in black squares (Pb²⁺ strongly cleaved nucleotide), small letters in black squares (Pb²⁺ weakly cleaved nucleotide), small letters (uncleaved nucleotide), bold capital letters (nucleotide resistant to Pb²⁺ cleavage and self degradation (regions of highly stable structure)), black circles (regions corresponding to single stranded regions of *G. max* ENOD40-1 mRNA (Girard *et al.*, 2003)), black arrows (starting codon of ORFs discussed in the text), black squares (stop codon of the ORFs discussed in the text), rectangles 1, 2, 3, and 6 (domains of conserved structure (Girard *et al.*, 2003)), solid lines (conserved regions I and II), dotted lines (motifs conserved in legume ENOD40 genes (see text)), underlined letters (conserved position), and poly(A) (putative polyadenylation signal).

- Crespi MD, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, Kondorosi A (1994) enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. *EMBO J* **13**: 5099–112.
- Dey M, Complainville A, Charon C, Torrizo L, Kondorosi A, Crespi M, Datta S (2004) Phytohormonal responses in enod40-overexpressing plants of *Medicago truncatula* and rice. *Physiol Plant* **120**: 132–139.
- Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiol* **131**: 900–910.
- Fang Y, Hirsch AM (1998) Studying early nodulin gene ENOD40 expression and induction by nodulation factor and cytokinin in transgenic alfalfa. *Plant Physiol* **116**: 53–68.
- Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**: 368–376.
- Girard G, Roussis A, Gulyaev AP, Pleij CW, Spaink HP (2003) Structural motifs in the RNA encoded by the early nodulation gene enod40 of soybean. *Nucleic Acids Res* **31**: 5003–5015.
- Golinowski W, Kopcinska J, Borucki W (1987) The morphogenesis of lupine root nodule during infection by *Rhizobium lupinii*. *Acta Soc Bot Pol* **56**: 687–703.
- Gonzalez-Sama A, Lucas MM, de Felipe MR, Pueyo JJ (2004) An unusual infection mechanism and nodule morphogenesis in white lupin (*Lupinus albus*). *New Phytologist* **163**: 371–380.
- Goormachtig S, Alves-Ferreira M, Van Montagu M, Engler G, Holsters M (1997) Expression of cell cycle genes

- during *Sesbania rostrata* stem nodule development. *Mol Plant Microbe Interact* **10**: 316–325.
- Gronlund M, Roussis A, Flemenakis E, Quaevlieg NE, Schlaman HR, Umehara Y, Katinakis P, Stougaard J, Spaink HP (2005) Analysis of promoter activity of the early nodulin Enod40 in *Lotus japonicus*. *Mol Plant Microbe Interact* **18**: 414–427.
- Gulyaev AP, Roussis A (2007) Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants. *Nucleic Acids Res* **35**: 3144–3152.
- Guzzo F, Portaluppi P, Grisi R, Barone S, Zampieri S, Franssen H, Levi M (2005) Reduction of cell size induced by enod40 in *Arabidopsis thaliana*. *J Exp Bot* **56**: 507–13.
- Hardin SC, Tang GQ, Scholz A, Holtgraewe D, Winter H, Huber SC (2003) Phosphorylation of sucrose synthase at serine 170: occurrence and possible role as a signal for proteolysis. *Plant J* **35**: 588–603.
- Kirsebom LA, Ciesiolka J (2005) Pb(II)-induced cleavage of RNA. In: *Handbook of RNA Biochemistry*. Hartmann RK, Bindereif A, Schon A, Westhof E, eds, pp 214–228. Wiley VCHGmbH& Co, KgaA, Weinheim, Germany.
- Klimecka M, Muszynska G (2007) Structure and functions of plant calcium-dependent protein kinases. *Acta Biochim Polon* **54**: 219–233.
- Kouchi H, Hata S (1993) Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. *Mol Gen Genet* **238**: 106–119.
- Kouchi H, Takane K, So RB, Ladha JK, Reddy PM (1999) Rice ENOD40: isolation and expression analysis in rice and transgenic soybean root nodules. *Plant J* **18**: 121–129.
- Kumagai H, Kinoshita E, Ridge RW, Kouchi H (2006) RNAi knock-down of ENOD40s leads to significant suppression of nodule formation in *Lotus japonicus*. *Plant Cell Physiol* **47**: 1102–1111.
- Larsen K (2003) Molecular cloning and characterization of a cDNA encoding a ryegrass (*Lolium perenne*) ENOD40 homologue. *J Plant Physiol* **160**: 675–687.
- Linsey K (2001) Plant peptide hormones: The long and the short of it. *Curr Biol* **11**: R741–R743.
- Logemann J, Schell J, Willmitzer L (1987) Improved method for the isolation of RNA from plant tissues. *Anal Biochem* **163**: 16–20.
- Lotocka B, Kopcińska J, Górecka M, Golinowski W (2000) Formation and abortion of root nodule primordia in *Lupinus luteus* L. *Acta Biol Crac Ser Bot* **42**: 87–100.
- Matvienko M, Van de Sande K, Yang WC, van Kammen A, Bisseling T, Franssen H (1994) Comparison of soybean and pea ENOD40 cDNA clones representing genes expressed during both early and late stages of nodule development. *Plant Mol Biol* **26**: 487–493.
- Minami E, Kouchi H, Cohn JR, Ogawa T, Stacey G (1996) Expression of the early nodulin, ENOD40, in soybean roots in response to various lipo-chitin signal molecules. *Plant J* **10**: 23–32.
- Murakami Y, Miwa H, Imaizumi-Anraku H, Kouchi H, Downie JA, Kawaguchi M, Kawasaki S (2006) Positional cloning identifies *Lotus japonicus* NSP2, a putative transcription factor of the GRAS family, required for NIN and ENOD40 gene expression in nodule initiation. *DNA Res* **13**: 255–265.
- Ndoye I, de Billy F, Vasse J, Dreyfus B, Truchet G (1994) Root nodulation of *Sesbania rostrata*. *J Bacteriol* **176**: 1060–1068.
- Papadopoulou K, Roussis A, Katinakis P (1996) Phaseolus ENOD40 is involved in symbiotic and non-symbiotic organogenetic processes: expression during nodule and lateral root development. *Plant Mol Biol* **30**: 403–417.
- Rohrig H, Schmidt J, Miklashevichs E, Schell J, John M (2002) Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc Natl Acad Sci USA* **99**: 1915–1920.
- Rohrig H, John M, Schmidt J (2004) Modification of soybean sucrose synthase by S-thiolation with ENOD40 peptide A. *Biochem Biophys Res Commun* **325**: 864–867.
- Ruttink T, Boot K, Kijne J, Bisseling T, Franssen H (2006) ENOD40 affects elongation growth in tobacco Bright Yellow-2 cells by alteration of ethylene biosynthesis kinetics. *J Exp Bot* **57**: 3271–3282.
- Saitou N, Nei M (1987) The neighbor-joining method. A new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.
- Sambrook J, Russell DW (2002) *Molecular Cloning, a laboratory manual*; pp 7.31–7.34. Cold Spring Harbor, New York.
- Sousa C, Johansson C, Charon C, Manyani H, Sautter C, Kondorosi A, Crespi M (2001) Translational and structural requirements of the early nodulin gene enod40, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. *Mol Cell Biol* **21**: 354–366.
- Strozycki PM, Skapska A, Szczesniak K, Sobieszczuk E, Briat J-F, Legocki AB (2003) Differential expression and evolutionary analysis of the three ferritin genes in the legume plant *Lupinus luteus*. *Physiol Plantarum* **118**: 380–389.
- Strozycki PM, Szczurek A, Lotocka B, Figlerowicz M, Legocki AB (2007) Ferritins and nodulation in *Lupinus luteus*: iron management in indeterminate type nodules. *J Exp Bot* **58**: 3145–3153.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.
- Varkonyi-Gasic E, White DW (2002) The white clover enod40 gene family. Expression patterns of two types of genes indicate a role in vascular function. *Plant Physiol* **129**: 1107–1118.
- Vijn I, Yang WC, Pallisgard N, Ostergaard Jensen E, van Kammen A, Bisseling T (1995) VsENOD5, VsENOD12 and VsENOD40 expression during Rhizobium-induced nodule formation on *Vicia sativa* roots. *Plant Mol Biol* **28**: 1111–1119.
- Vincent JM (1970) *A Manual for the Practical Study of Root-Nodule Bacteria*. Blackwell Scientific Publications, London.
- Vleghels I, Hontelez J, Ribeiro A, Fransz P, Bisseling T, Franssen H (2003) Expression of ENOD40 during tomato plant development. *Planta* **218**: 42–49.
- Wan X, Hontelez J, Lillo A, Guarnerio C, van de Peut D, Fedorova E, Bisseling T, Franssen H (2007) *Medicago truncatula* ENOD40-1 and ENOD40-2 are both involved in nodule initiation and bacteroid development. *J Exp Bot* **58**: 2033–2041.
- Wen J, Lease KA, Walker J (2004) DVL, a novel class of small polypeptides: overexpression alters *Arabidopsis* development. *Plant J* **37**: 668–677.
- Yang WC, Katinakis P, Hendriks P, Smolders A, de Vries F, Spee J, van Kammen A, Bisseling T, Franssen H (1993) Characterization of *GmENOD40*, a gene showing novel patterns of cell-specific expression during soybean nodule development. *Plant J* **3**: 573–585.
- Zucker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.