The EMBO Journal vol.6 no.12 pp.3571–3577, 1987

Differences in expression between two seed lectin alleles obtained from normal and lectin-deficient beans are maintained in transgenic tobacco

Toni Voelker, Arnd Sturm and Maarten J.Chrispeels

Department of Biology C-016, University of California, San Diego, La Jolla, CA 92093, USA

Communicated by H.Kössel

Using *Agrobacterium*-mediated transformation, two genes for phytohemagglutinin-L (PHA-L), the lectin seed protein of the common bean *Phaseolus vulgaris*, were stably integrated into the tobacco genome. The two alleles for PHA-L, *dlec2* and *Pdlec2*, were obtained from a normal cultivar (Greensleeves) and a lectin-deficient cultivar (Pinto) respectively. In the bean embryos, the expression of *dlec2* is 30 times greater than the expression of *Pdlec2*. In the *dlec2*-transformed tobacco, PHA-L accumulated specifically in the seeds at the same stages as the tobacco seed storage proteins and was degraded after germination. PHA-L was found in the embryo, and at a 5–7 times lower concentration in the endosperm tissue of the mature tobacco seeds. No PHA could be detected in other parts of the plants. We conclude that the signals for temporal and spatial regulation of the *dlec2* gene are present in the DNA fragment used for transformation. Transformation with the second PHA-L allele *Pdlec2* from the cultivar Pinto caused the accumulation of about 50 times less PHA-L in tobacco seeds when compared to *dlec2*. We conclude from analyzing the 5′ sequences of *dlec2* and *Pdlec2* that the low expression phenotype of the *Pdlec2* allele could be due to the absence or mutation of a *cis*-acting element carried by the *dlec2* fragment.

Key words: phytohemagglutinin genes/*Phaseolus vulgaris*/transgenic tobacco/promoter analysis

Introduction

Phytohemagglutinin (PHA), the lectin of the common bean (*Phaseolus vulgaris*), is synthesized during seed development and makes up several percent of the total protein content of the seeds at maturity. PHA accumulates primarily in the large storage parenchyma cells of the cotyledons where it is sequestered in storage protein vacuoles together with the major seed storage protein phaseolin (see Chrispeels, 1984). PHA is a tetrameric glycoprotein consisting of two different subunits called PHA-E (erythroagglutinating) and PHA-L (leucoagglutinating and mitogenic) (Laevitt *et al.*, 1977). These two polypeptides are encoded by two highly homologous and tandemly linked genes *dlec1* (PHA-E) and *dlec2* (PHA-L) (Hoffman and Donaldson, 1985). The PHA genes are coordinately expressed during seed development (Staswick and Chrispeels, 1984; Chappell and Chrispeels, 1986). The genes are expressed primarily in the cotyledons with 10 times lower levels of expression in the axis (Chrispeels *et al.*, 1984) and very low expression in some vegetative tissues (Borrebaeck and Mattiasson, 1983).

Some cultivars of the common bean accumulate no or very low levels of PHA in the cotyledons. For example the Pinto UI111 cultivar (hereafter referred to as Pinto) has no PHA-E and a 40-fold reduced level of PHA-L (Pusztai *et al.*, 1981). The low levels of PHA can be accounted for by low levels of PHA mRNA (Vitale *et al.*, 1985) and a low rate of PHA gene transcription (Chappell and Chrispeels, 1986). A detailed analysis of the PHA genes and their expression in the cultivar Pinto (Voelker *et al.*, 1986a) showed that the gene for PHA-E (*Pdlec1*) has a frameshift in the coding sequence while the gene for PHA-L (*Pdlec2*) is intact in the reading frame but has a deletion in the 5′ flanking region. *Pdlec2* mRNA was found to be significantly reduced relative to *dlec2*, and *Pdlec1* mRNA levels were found to be extremely low. These features of the PHA genes make then an excellent model system for dissecting the signals which govern the expression of developmentally regulated and tissue-specific genes.

It has previously been shown that seed storage protein genes of legumes, when integrated into the genome of tobacco or petunia, carry the signals needed for correct spatial and temporal gene expression in the heterologous system (Sengupta-Gopalan *et al.*, 1985; Beachy *et al.*, 1985; Okamuro *et al.*, 1986). Plant-to-plant variation of gene expression in plants transformed with the same seed storage gene fragment has been found to be low. This demonstrates that this class of genes is remarkably resistant to position effects when integrated in the plant genome (Chen *et al.*, 1986). These properties permit functional analysis of the promoters of storage protein and seed lectin genes in transgenic plants.

We are interested in determining the mechanism(s) responsible for the differential expression of PHA genes in the two bean cultivars mentioned above, and have focused our investigations on two PHA-L alleles. In a previous paper (Voelker *et al.*, 1986a), we reported the nucleotide sequence of *Pdlec2* and compared it to *dlec2*. We speculated that the low expression of *Pdlec2* may be due to a missing enhancer. In this paper we provide a functional analysis of the two genes in transgenic tobacco. We found that the *dlec2* allele is expressed in transgenic tobacco in a manner similar to that observed in the bean. Transgenic tobacco harboring the *Pdlec2* allele produces seeds with about 50 times less PHA-L, consistent with our previous hypothesis that the low expression phenotype of the Pinto PHA-L gene is due to a mutation in a *cis*-acting element. A comparison of upstream sequences of *dlec2* versus *Pdlec2* and other seed protein genes allows us to tentatively define an enhancer-like structure which is missing in *Pdlec2*.

Results

Isolation of a PHA-L gene of P.vulgaris cv. Greensleeves

A bacteriophage λ EMBL genomic library of *P.vulgaris* cv. Greensleeves was screened with the PHA cDNA clone pSC1 (Staswick and Chrispeels, 1984). A 6-kb *Pst*I subclone (pTV562) hybridizing to pSC1 was further characterized. Restriction enzyme mapping revealed virtual identity to the *dlec2* gene of *P.vulgaris*, cv. Tendergreen, previously described by Hoffman and Donaldson, 1985 (comparison not shown). Partial DNA se-

© IRL Press Limited, Oxford, England

14602075, 1987, 12, Downloaded from https://www.embopress.org/doi/10.1002/j.1460-2075.1987.tb02687.x by Cochrane Canada Provision, Wiley Online Library on [24/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



Fig. 1. DNA sequence comparisons of the 5′ upstream regions of the *P. vulgaris* PHA-L gene alleles of cv. Greensleeves (*dlec2*) and cv. Pinto (*Pdlec2*). Homologous areas between the two genes are indicated by double dots between the sequences. Deletions have been inserted for best alignment. Direct and inverted repeats are indicated by arrows above or below the sequences (letters indicate similarities). The transcription start, as determined by Hoffman and Donaldson (1985) for *dlec2*, is indicated by the broken double line above the end of the *dlec2* sequence. 'Met' indicates the first amino acid of the polypeptide. The Tendergreen *dlec2* sequence begins at bp 658 (Hoffman and Donaldson, 1985). The sequence of *dlec2* was obtained by dideoxy sequencing of the *dlec2* allele cloned from *P. vulgaris*, cv. Greensleeves. The sequence for *Pdlec2* is taken from Voelker *et al.* (1986a).

quencing of the 6-kb *Pst*I fragment from Greensleeves confirmed this interpretation. Figure 1 (*dlec2*) contains the complete 1188 bp sequence between the *Pst*I site of cloning and the beginning of the *dlec2* reading frame (coding area only partially determined, not shown). The gene, *dlec2*, codes for PHA-L (Hoffman and Donaldson, 1985, 1987; Voelker *et al.*, 1986b) which represents 1.25% of the total protein of the mature beans in the cultivar Greensleeves (see below).

Introduction of the PHA alleles into the tobacco genome

For tobacco transformation we used the binary Ti vector system of Bevan (1984), with Bin19 as a cloning vector, and followed the leaf disk method of Horsch *et al.* (1985). A 3.5-kb *Sal*I fragment (Figure 2C) from the genomic clone pTV562 (containing the *dlec2* gene from Greensleeves) was selected as a highly expressed PHA allele. A 2.7-kb *Bam*HI/*Sst*I fragment (Figure 2C), carrying a second PHA allele, *Pdlec2*, from Pinto (Voelker *et al.*, 1986a), was also used. Its gene product, Pinto PHA-L, accumulates to only 0.04% of total cotyledon protein in Pinto beans (see below, and Pusztai *et al.*, 1981). *Pdlec2* is about twenty times less active than *dlec2* as judged by the level of mRNA in developing embryos (Voelker *et al.*, 1986a).

Following co-cultivation, kanamycin-resistant shoots were rooted on kanamycin-containing medium, and for each construction we obtained and analyzed five plants. Total DNA of these plants was analyzed by Southern blotting (Figure 2A, B). A nick-translated DNA probe specific for the PHA gene reading frame does not give rise to any signal with the DNA from either untransformed (lane 0) or Bin19-transformed (lane 19) tobacco plants. Digestion of DNA of the *dlec2*-transformed plants (Figure 2A, lanes *dlec2*A−E) with *Hin*dIII shows in each case a single band co-migrating with the 2.9-kb band obtained by *Hin*dIII digestion of the Bin19−*dlec2* plasmid itself, lane M (see Figure 2C for expected sizes of restriction fragments). A comparison of the band intensities from the transformed plants (lanes *dlec2*A−E) with a genomic reconstruction (lanes 3, 1 and 0.3 in panel A) shows a copy number of less than 0.3 for the plants A, C and E, and slightly more for plants B and D. The allelic frequency of a single insertion into the genome of the allotetraploid *Nicotiana tabacum* (Narayan and Rees, 1974) is 0.25. Therefore we can conclude from the Southern analysis of our *dlec2*-transformed plants that in all five isolates the *dlec2* allele was inserted without obvious rearrangements. The semi-quantitative nature of the reconstruction experiment allows the conclusion that the copy number is low, probably one or several.

The results obtained by DNA analysis of the *Pdlec2* transformants were comparable to those described above (Figure 2B). *Bam*HI/*Sst*I digests of total DNA of all five plants (lanes *Pdlec2* A−E) show a band co-migrating with the respective 2.7-kb fragment of *Pdlec2* plasmid DNA (lanes 3, 1 and 0.3 in panel B). In addition higher molecular weight bands are present in all digests of the *Pdlec2*-transformed plants. A likely explanation of this pattern is that it represents partial digestion due to a higher sensitivity to DNA methylation of these restriction enzymes. We cannot exclude the possibility that the higher bands represent rearranged copies. The copy number determination (compare band intensities with the lanes 3, 1 and 0.3 in the genomic reconstitution) again shows a low copy number per genome in all five plants.

Flowers of all the plants were selfed and the seeds harvested after maturation. When germinated on medium containing kanamycin, about 25% of the seeds of five or six chosen transformants

14602075, 1987, 12, Downloaded from https://www.embopress.org/doi/10.1002/j.1460-2075.1987.tb02687.x by Cochrane Canada Provision, Wiley Online Library on [24/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

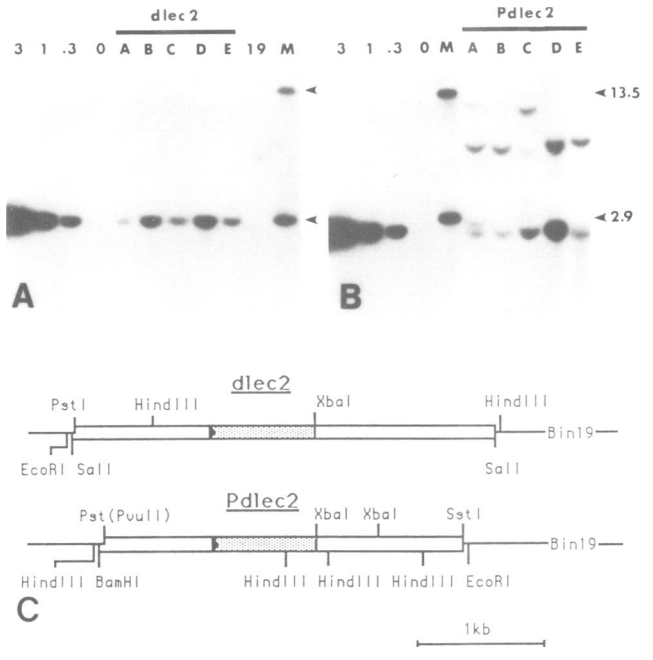


Fig. 2. Southern blot analysis of DNA from transformed tobacco plants (A,B) and restriction maps of the integrated PHA-L alleles (C). **A,B**: 5 μg of DNA isolated from tobacco plants were digested with the appropriate restriction enzymes, fractionated on 0.6% agarose gels, transferred to nitrocellulose membrane and probed with a ^{32}P-labeled nick-translated DNA fragment containing the coding sequence of *dlec2*. The specific activity of the probe was 6×10^8 c.p.m. The final washing stringency was performed at 65°C in 0.1 × SSC. The size marker, **lanes M**, is a partial *Hin*dIII digest of the *dlec2*-Bin19 clone (Figure 2C). Bands represent 13.5 kb and 2.9 kb (arrows). **Panel A**: The four lanes on the left show copy number reconstructions of 3, 1, 0.3 and 0 copies of *dlec2* per haploid tobacco genome (1.5×10^9 bp, see Zimmerman and Goldberg, 1977). Lanes *dlec2*A–E represent the DNA from the five independently obtained tobacco plants transformed with *dlec2* cut with *Hin*dIII. **Lane 19** is a control and represents DNA from Bin19-transformed tobacco. **Panel B**: The four lanes on the left show copy number reconstructions of 3, 1, 0.3 and 0 copies of *Pdlec2* per haploid tobacco genome. Lanes *Pdlec2*A–E represent the DNA from five independently transformed tobacco plants. The DNA was cut with *Bam*HI and *Sst*I. **C**: Restriction enzyme map of the PHA clones used in the tobacco transformations. Both fragments are shown as inserted in the vector Bin19. Shaded areas indicate the coding regions for the respective lectin genes with the orientation of transcription from left to right. *dlec2*: the 3.5-kb *Sal*I restriction fragment carrying the PHA allele *dlec2* of *P. vulgaris*, cv. Greensleeves. This gene encodes PHA-L. *Pdlec2*: the 2.7-kb *Bam*HI/*Sst*I restriction fragment carrying the *Pdlec2* PHA gene of *P. vulgaris*, cv. Pinto. This gene encodes Pinto PHA-L. The complete sequences 5′ of the reading frames are shown in Figure 1.

died after several weeks, i.e. the kanamycin resistance behaved as a single Mendelian trait. One plant (*Pdlec2*E) produced only 6% kanamycin-sensitive seeds, indicating two unlinked insertions into the tobacco genome.

In summary the genetic and molecular analyses of our transformed tobacco plants showed that all plants grown from independent shoots have at least one copy of a PHA allele without major rearrangements. In most cases, the kanamycin resistance marker indicated a single Mendelian unit.

Development of an immunoprecipitation/immunoblot assay to detect low levels of PHA in tobacco seeds

We used immunoblot analysis to detect PHA-L in extracts of tobacco seeds. When we loaded 500 μg of total tobacco seed protein extract (Bin19-transformed control, Figure 3, lane 2), we obtained no PHA signal on an immunoblot. Lane 1 contains the same amount of protein from the *dlec2*B-transformed plant. The signal consists of a deformed band in the region of PHA-L (lane

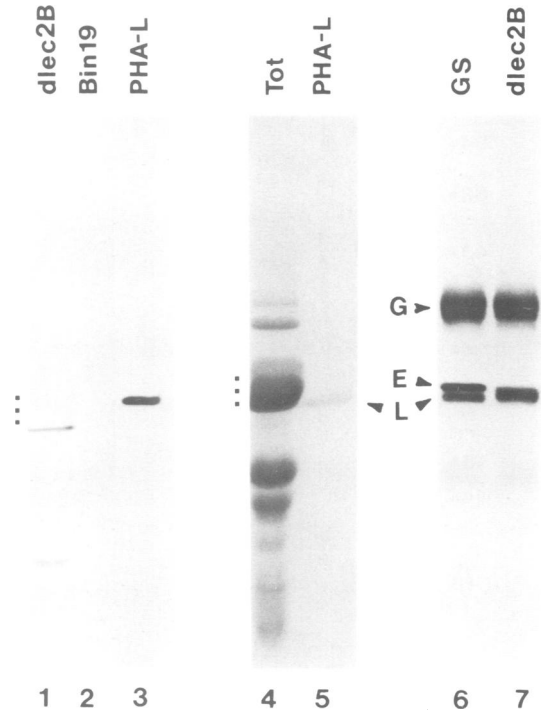


Fig. 3. Immunoblot analysis of PHA from seeds of transgenic tobacco after separation by SDS–PAGE. PHA was visualized with antibodies against PHA and peroxidase coupled to goat-anti-rabbit IgG. **Lanes 1–3:** direct immunoblot of seed proteins. In **lanes 1** and **2**, 500 μg of protein extracted from mature tobacco seeds of plant *dlec2*B, and a Bin19-transformed tobacco plant were loaded on the gel. **Lane 3** represents 25 ng of PHA-L. **Lane 4**: Coomassie stain of 100 μg of mature tobacco seed proteins. **Lane 5**: Coomassie stain of 5 μg PHA-L. **Lanes 6** and **7**: immunoblot analysis of PHA precipitated by antibodies against PHA and protein A Sepharose. **Lane 6**: PHA from 6 μg total protein of mature cotyledons of *P. vulgaris*, cv. Greensleeves. **Lane 7**: PHA from 1 mg of tobacco seed protein (plant *dlec2*B). E and L indicate PHA-E and PHA-L respectively. G indicates the large subunit of the IgG molecules (anti-PHA) which are solubilized when the PHA is released from the Sepharose beads.

3) and a band with smaller M_r which we interpret as a breakdown product. We assumed that the signal from PHA-L was partially masked and displaced on the gel by the large amount of tobacco protein, because PHA-L has a mobility similar to the most abundant tobacco polypeptides (Figure 3, lanes 1 and 4, dotted area). Indeed, when adding small amounts of PHA-L to extracts of tobacco seeds, we found that the PHA signal was substantially weaker than expected and/or displaced on the immunoblot.

To obtain an interference-free and reproducible signal, we precipitated PHA-L out of the seed protein extract with purified anti-PHA IgG and protein A–Sepharose. The immunocomplex, now free from tobacco seed proteins, was released from protein A by boiling in SDS-buffer and then subjected to SDS–PAGE. The proteins were transferred to nitrocellulose paper, decorated with rabbit anti-PHA IgG, and visualized wtih goat-anti-rabbit IgG coupled to horseradish peroxidase. Two such analyses are shown in Figure 3, lanes 6 and 7. Lane 6 shows the signal obtained when 6 μg of total protein from beans of the cultivar Greensleeves was used. There are strong signals for PHA-E and PHA-L (labeled E and L, respectively) in the bean extract. Lane 7 contains the material from 1 mg seed protein of tobacco transformant *dlec2*B; we found a band in the same position as PHA-L. The smaller PHA breakdown product, see lane 1, was not recovered by this method. The additional broad band (labeled G) represents rabbit large subunit IgG polypeptides which were released from

14602075, 1987, 12, Downloaded from https://www.embopress.org/doi/10.1002/j.1460-2075.1987.tb02687.x by Cochrane Canada Provision, Wiley Online Library on [24/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

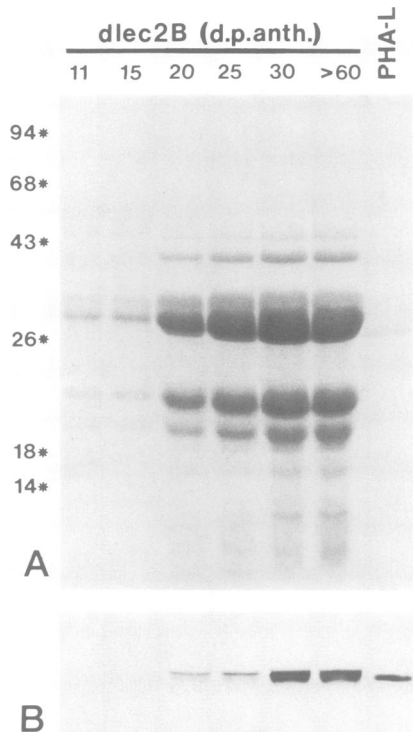


Fig. 4. Time course of PHA accumulation in developing seeds of transgenic tobacco. Total protein extracts of the developing seeds of the transgenic tobacco plant *dlec2*B were either used directly for SDS−PAGE and stained by Coomassie blue (**panel A**), or subjected to the two-step PHA immunodetection procedure as described (**panel B**). Lane marked PHA-L represents 5 μg PHA-L in **panel A** and 25 ng in **panel B**. Numbers at the left indicate $M_r \times 10^{-3}$. Numbers at the top refer to days post-anthesis.

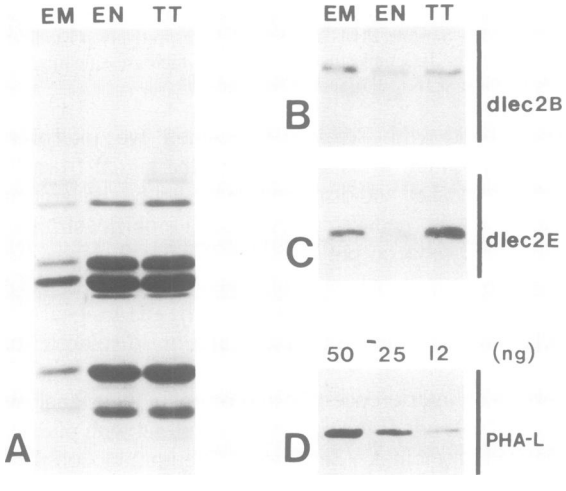


Fig. 5. Analysis of PHA in embryo and endosperm of transgenic tobacco seeds were dissected as described. **Panel A**: Coomassie stain of SDS−PAGE of total proteins from mature seeds of transgenic tobacco plant *dlec2*B. Each lane represents protein from 3 seeds. **EM**, embryo; **EN** endosperm with seedcoat; and **TT**, total seeds. **Panels B−D**: immunoblot detection of PHA. **Panels B** and **C** show PHA from 25 seeds each from the transgenic plants *dlec2*B and *dlec2*E, respectively. **Panel D** contains a dilution series of PHA-L for a reference of quantity (originally on the same blot as B and C).

the protein A−Sepharose beads when they were boiled with the SDS-containing buffer. This band is present in all samples, whether they contain PHA or not. This immunological method for determining the levels of PHA was reproducible, sensitive, and quantitative (see below).

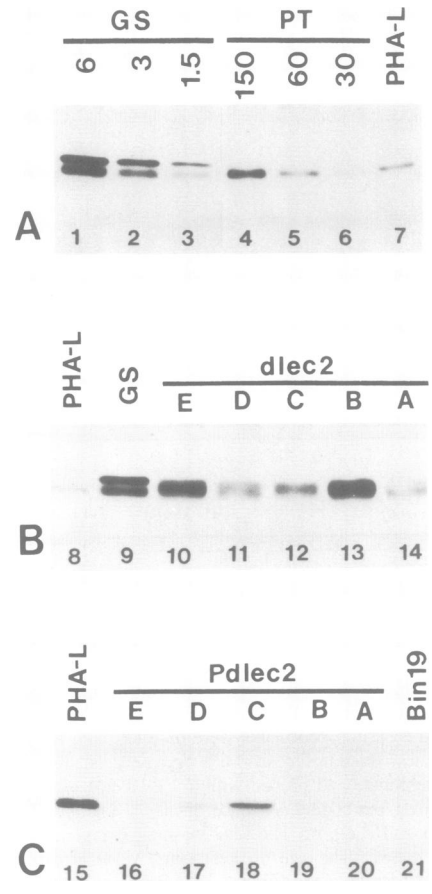


Fig. 6. Quantitation of PHA accumulation in the seeds of transgenic tobacco. All lanes represent antibody precipitation/immunoblot analysis as described in Materials and methods and shown in lanes 6 and 7 of Figure 3. **Panel A**: Quantitation of PHA in cotyledons of mature beans of *P.vulgaris*. PHA was immunoprecipitated from 6, 3 and 1.5 μg of total cotyledon protein of the cv. Greensleeves (**GS**) and from 150, 60 and 30 μg total cotyledon protein of the cv. Pinto (**PT**). **Lane 7** contains 25 ng of PHA-L. **Panel B, lane 8**: PHA-L (25 ng). **Lane 9**: PHA from 6 μg of bean cotyledon protein (cv. Greensleeves). **Lanes 10−14** each represent the PHA-L isolated from 1 mg of protein from mature seeds of the independently derived *dlec2* tobacco transformants. **Panel C, lane 15**: PHA-L (50 ng). **Lanes 16−20**: PHA obtained from 5 mg total seed protein from mature seeds of the *Pdlec2*-transformed tobacco plants. **Lane 21**: 5 mg of protein of the mature seeds of Bin19-transformed tobacco plants were used as negative control to show levels of nonspecific background stain.

PHA-L accumulates specifically in the developing seeds of the transformants

To determine whether PHA is expressed specifically in the seeds, extracts were made of leaves, stems, roots, flowers, and seeds of a *dlec2* transformant and analyzed for their PHA content by the immunoblotting procedure. We obtained no signal for any organs other than seeds (data not shown), indicating that PHA is at least 50 times less abundant (limit of immunoblot detection) in these other tissues.

To find out if the expression of PHA is developmentally regulated in the tobacco seeds, we made extracts of seeds obtained from the *dlec2*-transformed plant at different times after pollination. Figure 4A shows the analysis of these extracts by SDS−PAGE. It is obvious that the accumulation of the abundant storage proteins begins between 15 and 20 days post-anthesis (see also Santo and Kawashima, 1983). Figure 4B shows the immunoblot analysis of these extracts, and it appears that the accumulation of PHA-L occurs at the same time as the accumulation of the tobacco storage proteins. When seeds from the selfed

14602075, 1987, 12, Downloaded from https://www.embopress.org/doi/10.1002/j.1460-2075.1987.tb02687.x by Cochrane Canada Provision, Wiley Online Library on [24/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14602075, 1987, 12, Downloaded from https://www.embopress.org/doi/10.1002/j.1460-2075.1987.tb02687.x by Cochrane Canada Provision, Wiley Online Library on [24/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

*dlec2*B plant were allowed to germinate, there was a rapid disappearance of PHA-L. No PHA-L was detectable after 6 days of growth (data not shown). PHA is degraded in the cotyledons of young bean seedlings; PHA-L in tobacco also follows this pattern.

PHA-L accumulates in both the embryo and the endosperm of tobacco seeds

To determine whether the accumulation of PHA-L is embryo-specific, we dissected mature seeds of *dlec2* transformants into embryo and endosperm (plus seedcoat) fractions and analyzed each for total protein and PHA-L content. Total protein determinations showed that the mature seeds contained an average of 10.5 μg protein. The endosperm plus seedcoat fraction contained 7.5 μg protein per seed and 3.0 μg of protein was found per embryo. Analysis by SDS−PAGE (Figure 5A) shows that the polypeptides of the 11.7 S major storage protein (Sano and Kawashima, 1983) are located largely in the endosperm. Analysis of PHA-L in the embryo and endosperm (plus seedcoat) fractions of 25 seeds obtained from 2 different *dlec2*-transformed plants is shown in Figure 5, panels B and C. It appears that PHA-L accumulated in both the embryo and the endosperm. We estimate that the PHA-L accumulation in embryos is 2−3 times higher than in the endosperm. Taking into account that embryos contain 2.5 times less total protein, PHA-L is 5−7.5 times more abundant in the embryo compared to the endosperm. Quantitation with a dilution series of purified PHA-L (Figure 5D) reveals about 1.5 ng PHA-L per tobacco embryo. With 3 μg total protein per embryo, PHA makes up 0.05% of the protein in tobacco embryos.

Levels of PHA accumulation in seeds of transformed tobacco are allele-specific

Transformed tobacco plants were used to determine whether the expression of the PHA genes and the accumulation of PHA-L in the seeds differ for tobacco transformed with *dlec2* and *Pdlec2* in the same way as they do for the Greensleeves and Pinto cultivars of *P. vulgaris*. A quantitative measure of the amount of PHA-L in the cotyledons of the cultivars Greensleeves and Pinto is shown in Figure 6A. Lanes 1−3 show the PHA present in 6, 3 and 1.5 μg respectively of Greensleeves cotyledon protein; lanes 4−6 contain the PHA present in 150, 60 and 30 μg of Pinto cotyledon protein. Note that Greensleeves contains PHA-E and PHA-L, whereas Pinto contains only PHA-L. Lane 7 represents 25 ng of purified PHA-L. By visual comparison of signal intensities with the reference, lane 7, we calculated that Greensleeves contains about 1.25% of its total cotyledon protein as PHA-L, while Pinto contains about 0.04% of its total protein as PHA-L. Thus the accumulation of PHA-L is reduced about 30-fold in Pinto compared to Greensleeves. These values obtained using the immunological method are in good agreement with the biochemical determinations of Pusztai *et al.* (1981) and show that our immunoprecipitation/immunodetection assay allows the quantification of PHA-L in complex protein mixtures.

Figure 6B shows the level of PHA-L in seeds of five tobacco plants independently transformed with *dlec2*. Lanes 10−13 contain the PHA present in 1 mg of tobacco seed protein compared with the PHA in 6 μg of Greensleeves protein (lane 9) and 25 ng of PHA-L, lane 8. The five transformants accumulated similar amounts of PHA, and the average amount of PHA-L is about 100−200 ng per mg of protein or 0.01−0.02% of the total seed protein. These values are 50−100 times lower than the levels found in Greensleeves cotyledons.

Figure 6C shows the amount of PHA-L in five tobacco plants independently transformed with *Pdlec2*. Lanes 16−20 contain the PHA present in 5 mg of tobacco seed protein compared to 50 ng of PHA-L (lane 15). All transformants produced PHA-L, although there was more variability than in the *dlec2*-transformed plants. The amount of PHA-L in these seeds is equivalent to 1−5 ng per mg of total seed protein or, on average, 50 times less than the amount observed in the seeds of *dlec2* transformants. This allele-specific difference in the level of gene expression parallels that observed in the beans. A plant transformed with the vector alone (lane 21) did not give any signal for PHA.

Discussion

Using a binary T-DNA vector system, PHA-L genes of two different cultivars of *P. vulgaris* were stably integrated into the tobacco genome. PHA-L was detected in all transformed plants by immunoprecipitation/immunoblotting. Comparisons of the M_r values on immunoblots show that *Pdlec2* gives rise to a polypeptide with the same mobilty as authentic PHA-L isolated from Pinto beans. In seeds obtained from tobacco plants transformed with *dlec2*, we usually observe a doublet of PHA-L polypeptides. Synthesis of PHA in beans involves removal of a signal peptide, the attachment of two high mannose glycans and modification of one of the glycans in the Golgi (see Chrispeels, 1984). We conclude that with high, but not complete, efficiency the heterologous system recognizes the signals for the modifications of the primary PHA translation product, and heterologous PHA-L is produced.

The PHA high expression allele *dlec2* from the cultivar Greensleeves is active in tobacco in a pattern similar to that observed in the bean. PHA-L accumulates in developing seeds and is not detectable in other plant organs. Soon after germination the PHA-L stored in the tobacco seeds is degraded. The major part of the information necessary for the regulation of spatial and temporal expression of this PHA gene is present on the DNA fragment transferred and is properly interpreted in tobacco. The amount of PHA-L in the embryos of tobacco seeds averages 0.05% of total protein. Mature cotyledons of Greensleeves beans contain 25 times more PHA-L (1.25% of total protein). This difference in PHA accumulation may be in part due to degradation of PHA in the tobacco protein bodies and to a lower expression of *dlec2* in the heterologous system.

In *P. vulgaris* PHA genes are most highly expressed in the developing cotyledons and to a lesser extent in the embryonic axis (Chrispeels *et al.*, 1984; Greenwood *et al.*, 1984) and vegetative tissues. In tobacco the PHA allele *dlec2* causes the accumulation of PHA-L not only in the developing embryo but also in the endosperm. Whether seed storage protein genes are expressed in endosperm of *P. vulgaris* has not been examined, because the liquid endosperm degenerates before the onset of storage protein and lectin accumulation (Öpik, 1968). A phaseolin gene, which is expressed in the developing bean in a pattern comparable to the PHA genes, caused the accumulation of phaseolin in the endosperm of transgenic tobacco (Sengupta-Gopalan *et al.*, 1985). Immunocytochemical analysis unequivocally established its presence in the protein bodies of embryo and endosperm cells in seeds of transgenic tobacco (Greenwood and Chrispeels, 1985). Excluding the remote possibility of protein transfer from embryo to endosperm, the data show that promoters of seed storage protein and lectin genes direct the expression of these genes not only in embryos, but are also active in the endosperm tissue when it functions as a protein storage compartment.

It has been shown that T-DNA-mediated integration into the tobacco genome is essentially random (Czernilofsky *et al.*, 1986). Since the five independently transformed *dlec2* plants show very

similar amounts of PHA-L in their seeds, gene expression must be independent of chromosomal location. This property of the bean lectin gene makes a quantitative comparison of *dlec2* with the *Pdlec2* alelle of Pinto possible. In the bean PHA-L of Pinto accumulates 30 times less than PHA-L in the cultivar Greensleeves. We find a similar ratio in transgenic tobacco and conclude that the low expression of *Pdlec2* may be caused by the absence or lack of function of a *cis*-acting element carried on the transferred DNA fragment.

The reading frame of *Pdlec2* and more than 100 bp 3′ including the poly(A) site are greater than 90% homologous to *dlec2*, the high expression gene of PHA. However the situation found 5′ to the reading frame is substantially different. In Figure 1 the complete upstream sequences of the transferred *dlec2* and *Pdlec2* fragments are compared with each other, with the translational start codon at the right and bottom of the sequence. After a short stretch of almost complete homology between the initiation codon and the TATA box (left end of 'BR'), *dlec2* has a large deletion relative to *Pdlec2* (and the other PHA alelles). This deletion does not cause attenuation of the expression of this gene since *dlec2* is highly expressed. Further upstream a 114-bp stretch of DNA is missing in *Pdlec2*. This area contains a 63-bp direct repeat structure ('A'), with a 16-bp inverted repeat in the center (found also in *dlec1* and *Pdlec1*, see Hoffman and Donaldson, 1985; Voelker *et al.*, 1986a). After an additional 18-bp matching sequence defining the 5′ border of the 114-bp deletion of *Pdlec2*, the two sequences completely diverge (from this point on, the sequences of all four PHA alleles diverge). We suggest that the 114-bp deletion in *Pdlec2* is responsible for the low expression phenotype of the PHA-L gene of Pinto.

The pattern of expression observed with the PHA *dlec2* allele is in agreement, both qualitatively and quantitatively, with the expression of the related soybean seed lectin gene *le1* in tobacco (Okamuro *et al.*, 1986). The DNA sequences of *dlec2* and *le1* show about 60% homology in the reading frames (see Hoffman and Donaldson, 1985; Vodkin *et al.*, 1983, for sequences). A further sequence comparison indicated two significant homologies between *dlec2* and *le1* in the transcribed but untranslated sequences. Transcriptional starts are mapped to a conserved inverted repeat sequence TGCA which is followed by a second TGCA motif after several bases (RY in Figure 1). A foldback at the very 5′ end of the lectin messages is possible. Downstream of the reading frame, a 16-bp inverted repeat (repeat 3a, b for *le1*, see Okamuro *et al.*, 1986) in the vicinity of the poly(A) addition signal is almost completely conserved. Again a stem loop structure at the 3′ end of the lectin mRNA is possible. All other PHA genes share these features. These conserved structures may play a role in the transcriptional activity and/or stability of the lectin mRNA (Walling *et al.*, 1986). No significant matches were found in the sequences up to about 1 kb 5′ of the transcriptional start sites.

In a second search we included in our sequence comparison seed storage protein genes. The short conserved upstream sequence motif CATGCAT was reported by N.Nielsen (EMBO Workshop, Plant Storage Protein Genes, Breisach, FRG, 1986) to be present in all seed storage protein genes examined. It is found in *dlec2* (center of repeat A, Figure 1). Gatehouse *et al.* (1986) defined a 'vicilin box'; a conserved sequence at position −100 relative to the transcriptional start of several vicilin genes. The 5′ border sequence GCCACCTCA of this vicilin box is also found in legumin genes, and is completely contained in repeat A of *dlec2*. Both consensus sequences are absent in *Pdlec2* due to the 114-bp deletion. Further functional testing in transgenic plants will constitute additional proof of the significance of these structures.

Materials and methods

Bacterial strains and vectors

The helper strain *Escherichia coli* HB101 harboring pRK2013, the *Agrobacterium tumefaciens* strain LBA4404 and the vector Bin19 were originally obtained from M.Bevan (see Bevan, 1984).

Plant genomic DNA, PHA gene clones

The PHA gene, *dlec2*, was isolated and subcloned from a genomic bacteriophage λ EMBL library obtained from total DNA of *P.vulgaris*, cv. Greensleeves originally prepared by A.Vitale. The subclone pTV562 containing the *dlec2* gene represents a 6-kb *Pst*I DNA fragment cloned into pUC12. pTV562 was isolated using the PHA cDNA pSC1 (Staswick and Chrispeels, 1984) as a probe. The PHA *Pdlec2* gene of *P.vulgaris*, cultivar Pinto UI111, was isolated from a separate λ library prepared by P.Staswick (see Voelker *et al.*, 1986a). Figures 1 and 2C show restriction enzyme maps and partial sequence data for *dlec2* and *Pdlec2*.

Mobilization of Bin19 constructs, transformation of tobacco

The PHA alelles were cloned into the vector Bin19 (see Figure 1C) and maintained in *E.coli* HB101. For the triparental mating, equal volumes of overnight cultures of *E.coli* HB101 harboring the respective recombinant Bin19 plasmid, *E.coli* HB101 with the conjugative plasmid pRK2013, and the recipient *A.tumefaciens* strain LBA4404 were mixed and incubated on LB plates overnight at 30°C. Subsequently, the bacteria were resuspended in M9 minimal medium containing sucrose and shaken for several hours before the transconjugant *A.tumefaciens* were selected by plating on M9 plates containing sucrose and 50 μg kanamycin/ml and grown at 30°C (all *E.coli* strains used cannot utilize sucrose). Overnight cultures from kanamycin-resistant colonies were prepared in liquid and used for co-cultivaton of leaf discs of *N.tabacum*, var. Xanthii (see Horsch *et al.*, 1983) on MS shooting medium (Linsmaier and Skoog, 1965). After 2 days of cocultivation discs were transferred to MS shooting medium containing 250 μg/ml carbenicillin and 100 μg/ml kanamycin. Kanamycin-resistant shoots were transferred to hormone-free MS medium containing 250 μg/ml carbenicillin and 100 μg/ml kanamycin for rooting. The kanamycin-resistant plants were later transferred to soil and grown in the greenhouse. All plants flowered about 6 months after the start of the transformation procedure. Selected flowers were selfed on the day of anthesis. Mature seeds were harvested after dehydration (>30 days after anthesis) and kept dry at room temperature. Developing seeds were frozen in liquid nitrogen and kept at −70°C.

Genomic analysis of the transformed plants

To investigate the mode of transmission of the introduced kanamycin resistance marker, mature seeds of selfed plants were surface sterilized and plated on sucrose-free MS medium containing 100 μg/ml kanamycin. After several weeks, the ratio of kanamycin-resistant to sensitive plantlets was determined.

Total DNA isolation followed essentially the CTAB procedure of Murray and Thompson (1980). DNA was digested with a 5-fold excess of the respective restriction enzyme and subjected to electrophoresis in 0.6% Tris−Borate−EDTA agarose gels. DNA was transferred by standard Southern procedures and blotted to nitrocellulose filters (Maniatis *et al.*, 1982). Hybridization to the nick-translated probes in 50% formamide was carried out essentially as described by Meinkoth and Wahl (1984).

DNA sequencing

Partial DNA sequencing of the *P.vulgaris*, cv. Greensleves, *dlec2* gene by the method of Sanger *et al.* (1977) was as described in Voelker *et al.* (1986a).

Preparation of specific anti-PHA IgG

Four milligrams of PHA, isolated from mature beans by affinity chromatography (Felsted *et al.*, 1975), were chemically deglycosylated by trifluormethanesulfonic acid (Edge *et al.*, 1981). The deglycosylated lectin was dissolved in 20 mM Tris−HCl, pH 7.5, containing 0.4% SDS. 100 μl of protein solution and 300 μl phosphate-buffered saline (10 mM Na-phosphate, pH 7.5, and 150 mM NaCl) were emulsified with 0.4 ml Freund's complete adjuvant (Behring Diagnostics) and injected subcutaneously in 5−6 portions into a New Zealand white rabbit obtained from a local breeder. Two more booster injections were given at intervals of 4 weeks. After 2 additional weeks, blood was taken from the ear vein. Serum was prepared and IgG specific for PHA isolated by affinity chromatography, using PHA-Sepharose (5 mg PHA/ml Sepharose) as an affinity gel, as described by Smith *et al.* (1978). Anti-PHA IgG was dissolved in phosphate-buffered saline (final concentration 2.7 mg/ml).

Western immunoblot analysis

Proteins were extracted from tobacco seeds by grinding and subsequent boiling in sample buffer (20 mM Tris−HCl, pH 8.6, 1% SDS, 0.3% β-mercaptoethanol and 10% glycerol; 100 mg seeds were homogenized in 1 ml buffer). A volume

14602075, 1987, 12, Downloaded from https://www.embopress.org/doi/10.1002/j.1460-2075.1987.tb02687.x by Cochrane Canada Provision, Wiley Online Library on [24/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

equivalent to 1 mg seed protein was diluted with sample buffer to 0.2 ml; 0.8 ml 20 mM Tris−HCl, pH 8.6, containing 1.25% Triton X-100 were added (final concentration 1% Triton X-100 and 0.2% SDS). The sample was boiled and particulate matter removed by centrifugation. Four micrograms of specific anti-PHA IgG were added and the sample tumbled for 90 min at 4°C. Six milligrams of protein A−Sepharose (Pharmacia) were added and the tumbling continued for another 60 min at 4°C. The Protein A−Sepharose beads were carefully washed and the bound protein eluted by boiling with 30 μl of 2× sample buffer. Immunoblot analysis was performed according to the procedure detailed in the Bio-Rad manual. Anti-PHA serum and goat-anti-rabbit IgG coupled to horseradish peroxidase were both diluted 1:1000.

Protein assay

Protein was assayed by the method of Lowry *et al.* (1951).

Dissection of tobacco seeds

Single, mature tobacco seeds sticking to a masking-tape surface were cut under the dissecting microscope (×25) with two scalpel blades and the embryo removed from the seed coat plus endosperm. After visual inspection to ensure the complete integrity of the embryo, the two seed fractions were collected for further analysis. We estimate that cross-contamination between endosperm and embryo is of the order of ≤10%.

Acknowledgements

We thank Bobbie Bradford for excellent technical assistance. This work was supported by grants from the United States Department of Agriculture (Competitive Research Grants Program/Genetic Mechanisms for crop improvement) and the National Science Foundation (Metabolic Biology).

References

Beachy,R.N., Chen,Z.-L., Horsch,R.B., Rogers,S.G., Hoffmann,N.J. and Fraley,R.T. (1985) *EMBO J.*, **4**, 3047−3053.
Bevan,M. (1984) *Nucleic Acids Res.*, **12**, 8711−8721.
Borrebaeck,C.A.K. and Mattiasson,B. (1983) *Physiol. Plant.*, **58**, 29−32.
Chappell,J. and Chrispeels,M.J. (1986) *Plant Physiol.*, **81**, 50−54.
Chen,Z.-L., Schuler,M.A. and Beachy,R.N. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8560−8564.
Chrispeels,M.J. (1984) *Phil. Trans. R. Soc. Lond. B*, **304**, 309−322.
Chrispeels,M.J., Vitale,A. and Staswick,P. (1984) *Plant Physiol.*, **76**, 791−796.
Czernilofsky,A.P., Hain,R., Herrera-Estrella,L., Lörz,H., Goyvaerts,E., Baker,B.J. and Schell,J. (1986) *DNA*, **5**, 101−113.
Edge,A.S.B., Faltynek,C.R., Hof,L., Reichert,L.E.,Jr and Weber,P. (1981) *Anal. Biochem.*, **118**, 131−137.
Felsted,R.L., Laevitt,R.D. and Bachur,N.R. (1975) *Biochim. Biophys. Acta*, **450**, 72−81.
Gatehouse,J.A., Evans,I.M., Croy,R.R.D. and Boulter,D. (1986) *Phil. Trans. R. Soc. Lond. B*, **314**, 367−384.
Greenwood,J.S. and Chrispeels,M.J. (1985) *Plant Physiol.*, **79**, 65−71.
Greenwood,J.S., Keller,G.A. and Chrispeels,M.J. (1984) *Planta*, **162**, 548−555.
Hoffman,L.M. and Donaldson,D.D. (1985) *EMBO J.*, **4**, 883−889.
Hoffman,L.M. and Donaldson,D.D. (1987) *Biotechnology*, **5**, 157−160.
Horsch,R.B., Fry,J.E., Hoffmann,N.L., Eichholtz,D., Rogers,S.G. and Fraley, R.T. (1985) *Science*, **227**, 1229−1231.
Laevitt,R.D., Felsted,R.L. and Bachur,N.R. (1977) *J. Biol. Chem.*, **252** 2961−2966.
Linsmaier,E.M. and Skoog,F. (1965) *Physiol. Plant.*, **18**, 100−127.
Lowry,O.H., Rosebrough,N.J., Farr,A.L. and Randall,R.J. (1951) *J. Biol. Chem.*, **193**, 265−275.
Maniatis,T., Fritsch,E. and Sambrook,J. (1982) *Molecular Cloning*. Cold Spring Harbor Laboratory, NY.
Meinkoth,J. and Wahl,G. (1984) *Anal. Biochem.*, **138**, 267−284.
Murray,M.G. and Thompson,W.F. (1980) *Nucleic Acids Res.*, **8**, 4321−4325.
Narayan,R.K. and Rees,H. (1974) *Chromosoma*, **47**, 75−83.
Okamuro,J.K., Jofuku,K.D. and Goldberg,R.B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8240−8244.
Öpik,H. (1986) *J. Exp. Bot.*, **19**, 64−76.
Pusztai,A., Grant,G. and Stewart,J.C. (1981) *Biochim. Biophys. Acta*, **671**, 146−154.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463−5467.
Sano,M. and Kawashima,N. (1983) *Agric. Biol. Chem.*, **47**, 1305−1310.
Sengupta-Gopalan,C., Reichert,N.A., Barker,R.F., Hall,T.C. and Kemp,J.D. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3320−3324.
Smith,J.A., Hurrell,J.G.R. and Leach,S.J. (1978) *Anal. Biochem.*, **87**, 299−305.
Staswick,P. and Chrispeels,M.J. (1984) *J. Mol. Appl. Genet.*, **2**, 525−535.
Vitale,A., Ceriotti,A. and Bollini,R. (1985) *Planta*, **166**, 201−207.
Vodkin,L.O., Rhodes,P.R. and Goldberg,R.B. (1983) *Cell*, **34**, 1023−1031.
Voelker,T.A., Staswick,P. and Chrispeels,M.J. (1986a) *EMBO J.*, **5**, 3075−3082.
Voelker,T.A., Florkiewicz,R.Z. and Chrispeels,M.J. (1986b) *Eur. J. Cell Biol.*, **42**, 218−223.
Walling,L., Drews,G.N. and Goldberg,R.B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2123−2127.
Zimmerman,J.L. and Goldberg,R.B. (1977) *Chromosoma*, **59**, 227−252.

Received on July 21, 1987; revised on August 28, 1987

14602075, 1987, 12, Downloaded from https://www.embopress.org/doi/10.1002/j.1460-2075.1987.tb02687.x by Cochrane Canada Provision, Wiley Online Library on [24/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License