

A Novel Subtilisin-like Protease Gene from *Arabidopsis thaliana* is Expressed at Sites of Lateral Root Emergence

Leon W. NEUTEBOOM,[†] Luz M. VETH-TELLO, Olton R. CLIJDESDALE, Paul J. J. HOOYKAAS, and Bert J. van der ZAAL^{*}

Institute of Molecular Plant Sciences, Leiden University, Clusius Laboratory, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands

(Received 10 December 1998)

Abstract

Differential screening of a cDNA library for mRNA species that specifically accumulate during auxin-induced lateral root formation in *Arabidopsis thaliana* led to the isolation of the *AIR3* cDNA clone. The corresponding single copy gene consists of 10 exons which encode a protein that possesses all the characteristics of subtilisin-like proteases. The promoter of the *AIR3* gene was fused to the *gusA* (β -glucuronidase) reporter gene and introduced into *Arabidopsis*. Expression was almost completely restricted to the outer layers of the parental root at sites of lateral root emergence and could be observed even before protrusion of the newly formed root tip. In the presence of external auxin, GUS activity was visible throughout the parts of the root that are competent for lateral root formation. By digesting structural proteins in the extracellular matrix of cells located above sites of lateral root formation, *AIR3* might weaken cell-to-cell connections and thus facilitate lateral root emergence.

Key words: *Arabidopsis thaliana*; auxin; lateral root; subtilisin-like protease

1. Introduction

Plants can extend their root system by the formation of lateral roots, and *Arabidopsis thaliana* provides a relatively simple system to study this phenomenon histologically.¹ Lateral roots are derived from pericycle cells deep within the parental root tissue. Cell division is initiated in these cells, eventuating in the creation of a lateral root primordium which, as it grows, forces its way through the outer cell layers before final emergence takes place. Studies of mutants altered in lateral root formation^{2,3} indicate that the initiation and maturation of lateral roots are controlled by the plant hormone auxin.

We have previously shown that addition of the auxin 1-naphthaleneacetic acid (1-NAA) to *Arabidopsis* root cultures leads to a ca. 17-fold increase in the number of lateral root primordia within 48 hr.⁴ A differential screening

approach has been exploited to isolate cDNA clones of which the corresponding mRNAs accumulate during this process.⁴ One of these cDNA clones, designated *AIR3*, encodes a protein that possesses all the characteristics of serine proteases belonging to the family of subtilisins.⁴ The typical catalytic triad of the amino acids aspartic acid, histidine and serine, together with the substrate binding site, can be found in the deduced amino acid sequence of the *AIR3* cDNA. In mammals, subtilisins are involved in the cleavage of prohormones and proproteins at specific sites composed of basic amino acids or "dibasic sites."⁵ cDNA clones encoding subtilisin-like proteases have been isolated recently from plants,^{6–11} but it is not clear yet which function is fulfilled by these proteins. All plant subtilisin-like proteases are believed to be active extracellularly.

Here we describe the isolation and characterization of the first complete, actively transcribed, plant subtilisin-like protease gene: the *AIR3* gene. The promoter of *AIR3* directed expression of a *gusA* reporter gene at sites of lateral root emergence. The possible role of the *AIR3* protein in facilitation of lateral root emergence is discussed.

Communicated by Michio Oishi

* To whom correspondence should be addressed. Tel. +31-71-5274760, Fax. +31-71-5274999, E-mail: Zaal@Rulbim.LeidenUniv.nl

† Present address: Department of Plant Molecular Physiology, University of Hawaii at Manoa, St. John 503, Honolulu, HI 96822, USA

‡ The nucleotide sequence data reported will appear in the EMBL, GenBank and DDJB Databases under the accession number AF098632

2. Materials and Methods

2.1. Screening of a genomic library and cloning of the *AIR3* gene

A genomic library of *Arabidopsis thaliana* ecotype Columbia (kindly provided by J. T. Mulligan and R. W. Davis via the EU-BRIDGE *Arabidopsis* DNA Stock Center, Köln, Germany) was screened using [³²P]cDNA inserts of *AIR3*⁴ (accession number AF055848, isolated from ecotype C24). Plaques generating positive signals were purified and phage DNA was isolated. The genomic DNA inserts were analyzed by restriction enzyme mapping and suitable fragments hybridizing to the cDNA probes were cloned in pBluescript SK⁺(II) (Stratagene, La Jolla, CA). All procedures were essentially done according to Sambrook et al.¹² Sequence reactions and sequence analysis were performed as described in Neuteboom et al.⁴

2.2. Construction of a promoter-*gusA* fusion

A plasmid harboring a part of the *AIR3* gene was used for a polymerase chain reaction (PCR) amplification of a 1158-bp promoter fragment introducing a *Nco* I site at the ATG initiation codon. PCR was performed using the M13-20 primer and a primer with the composition 5'-GTGTTAGCTCCATGGTTCTTCTTC-3'. The following cycles were used: cycles 1-4, 1 min 94°C, 1 min 34°C, 2 min 72°C; cycles 5-29, 1 min 94°C, 1 min 44°C, 2 min 72°C; cycle 30, 1 min 94°C, 1 min 44°C, 10 min 72°C. The PCR fragment was digested with *Hind*III and *Nco* I and was cloned into the *Nco* I-*Hind*III sites of *GusXX*.¹³ An additional 6.0-kb *Sac* I-*Hind*III fragment was cloned into the 5' end of the *Hind*III site of the 1158-bp *AIR3* promoter-*gusA* construct resulting in a 7.1-kb *AIR3* promoter-*gusA* fusion. The PCR fragment and fusion with the *gusA* gene were checked by sequencing. The 6.0-kb *Sac* I-*Hind*III fragment contains phage polylinker sequences at the *Sac* I terminus. The *AIR3* promoter-*gusA* construct was cloned into the *Xba* I/*Xho* I sites of the wide host range vector pMOGλCAT.¹³

2.3. Transformation to *Arabidopsis thaliana*

The construct was mobilized to *Agrobacterium tumefaciens* strain MOG101¹⁴ using a triparental mating procedure.¹⁵ Transformation to *Arabidopsis thaliana* ecotype Columbia was performed as described by Vergunst et al.¹⁶

2.4. Histochemical staining of transgenic plants

T₂ transgenic seeds were surface-sterilized as described earlier.⁴ The seeds were sown in a straight line on agar solidified 1/2 MS¹⁷ medium containing 50 mg/l kanamycin. Seeds were pushed into the agar and subsequently vernalized for 4 days at 4°C in the dark to promote synchronized germination. Petri dishes containing the trans-

genic seeds were sealed with parafilm and positioned at an angle of 80° to allow roots to grow downwards. Seedlings were grown for 5 days under a 16-hr illumination cycle (100 μmol photons m⁻² s⁻¹) at 21°C. For induction experiments, seedlings were transferred to six-well macroplates (Greiner) containing 4.0 ml of liquid 1/2 MS medium per well. After the addition of 1-NAA (1 μM or 0.1 μM), the plates were shaken at 90 rpm with a 5-cm displacement for 16 hr at 21°C. Histochemical staining for GUS activities using 5-bromo-4-chloro-3-indolyl glucuronide (X-gluc) was performed as described by Jefferson.¹⁸

3. Results and Discussion

3.1. Isolation and sequence analysis of the *AIR3* gene

The *AIR3* cDNA insert was used as a probe to screen a genomic library. Two phages were isolated which both contained a large portion of the same *AIR3* gene. The gene contains ten exons which encode a protein of 772 amino acid residues (Fig. 1). The predicted protein and the 3' untranslated region show minor differences with the protein predicted from the original *AIR3* cDNA clone and its 3' untranslated region as a result of a difference in ecotype (the cDNA clone was isolated from ecotype C24, while the gene was isolated from ecotype Columbia). *AIR3* was found to be a single copy gene by Southern blot analysis (data not shown), despite the presence of other *Arabidopsis* subtilisin-like protease sequences in the databases. Cross-hybridization with the other genes does not take place as high sequence similarity is restricted to the characteristic regions of the corresponding proteins. Remarkably, the exon/intron arrangement of the *AIR3* gene is not correlated with the characteristic regions of the protein (Fig. 1A). Moreover, the nucleotides which encode the highly conserved region around the D residue, which is part of the catalytic triad, are separated by an intron (Fig. 1B). Between exons IV and V and between exons V and VI (Fig. 1A), the *AIR3* gene possesses two unusually large introns (1431 and 1761 bp, respectively). The latter intron, which can be found between position 3893 and 5654 in Fig. 1B, contains a region, between position 4725 and 5424, that shows resemblance with a putative non-autonomous transposon that has been identified previously in the sequences with the accession numbers AC002354, Z97335. Similar transposon-like sequences seem to be spread over all chromosomes (data not shown). The remaining introns do not share any significant homology with sequences in the database.

Genomic sequencing projects predict the existence of more than eight possible members of the subtilisin-like protease family in *Arabidopsis*. Some of these entries are genes composed of 10 exons, like *AIR3*, while others are apparently intronless. The *ARA12* subtilisin-like

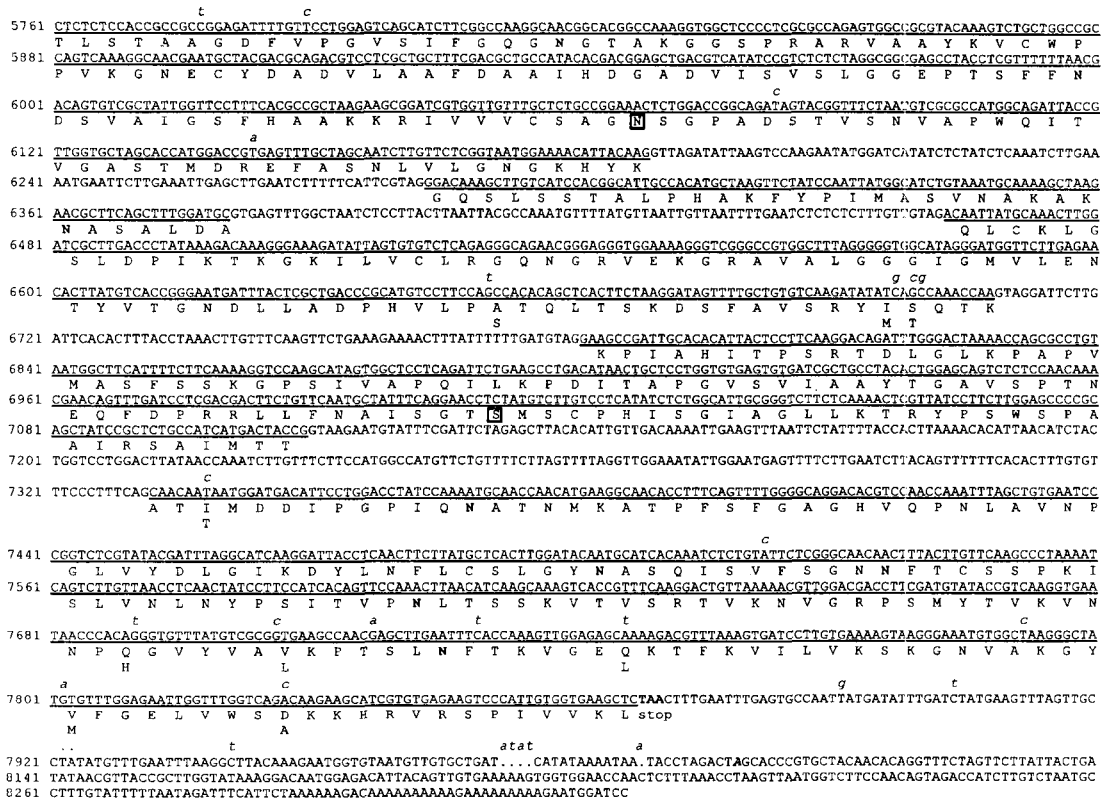


Figure 1. The *AIR3* gene. A, Genomic organization of the *AIR3* gene (above) and distribution of the exons over the protein (below). B, Nucleotide and deduced amino acid sequence of *AIR3* and flanking regions. Coding regions are underlined. The CAAT box, TATA box, translation start and stop are shown in boldface. Potential N-linked glycosylation sites are shown in boldface in the amino acid sequence. The putative signal peptide splicing site and pro-region splicing site are indicated by triangles. The reactive D, H and S region as well as the substrate binding site are boxed. The putative transcription start and the 5' coding end of the cDNA clone are indicated by arrows. The polyadenylation site is shown in bold italics. Differences at the nucleotide level between the cDNA clone (ecotype C24, accession number AF055848) and the genomic clone (ecotype Columbia) are indicated in italics above the sequence. The cDNA clone lacks two bases at position 7922 indicated by two dots above the sequence. At position 7972 and 7988 the cDNA clone possesses extra bases. Differences at the amino acid level are indicated below the amino acid sequence.

protease from *Arabidopsis*⁸ seems to be encoded by such an intronless gene (AF065639). A variety of *Arabidopsis* ESTs have been submitted, but due to their unfaithful and partial sequencing these cannot easily be designated to sequenced genes. It could thus be that several of the intronless genes are pseudogenes created by integration events of mRNA-derived sequences. Also in tomato, it has been shown that many subtilisin-like genes are present. Subtilisin-like proteases are apparently widely distributed in many different plant species and their activity seems to be required in an array of processes, such as microsporogenesis (*LIM9*¹⁹), nodule formation (Ag12⁸) and defense against pathogen infection (PR-P69A⁹ and PR-P69B¹⁰).

3.2. Plant subtilisin-like proteases consequently share the same basic structure

The deduced amino acid sequence of the *AIR3* gene has been analyzed before.⁴ Here, we have used the predicted *AIR3* peptide to compare the protein with all the pre-

dicted, complete plant subtilisin-like proteases encoded by cDNA clones (Fig. 2), including the latest entries in the database (the *SBT* genes from *Lycopersicon esculentum*). The structural similarity of plant subtilisin-like proteases, which has been acknowledged previously,^{4,9,10} is not invalidated: like other eukaryotic subtilisin-like proteases, all proteins possess a putative signal peptide, a pro-region and four well-conserved domains (in terms of both in amino acid composition and relative position within the protein) that form the active site of the mature protein. Unique for the plant members is the large spacing (ca. 230 amino acids) between the substrate binding site and the active serine residue compared to other eukaryotic subtilisin-like proteases. Also, the pro-region splicing site is different. In other eukaryotic subtilisin-like proteases this cleavage is believed to be achieved by autolysis at dibasic sites.²⁰ For PR-P69A,²¹ cucumisin⁶ and LIM9¹⁹ the mature protein was determined to start with two threonine residues. At comparable positions, all other plant subtilisin-like proteins also possess two threonine residues, making this the most likely site for

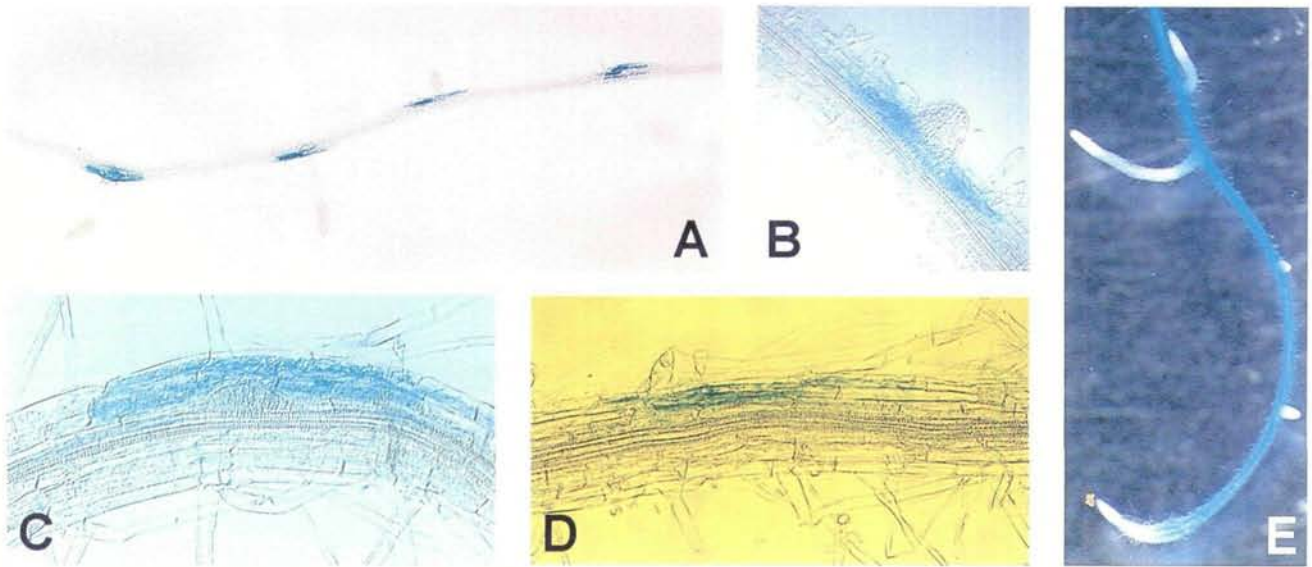


Figure 3. Histochemical localization of *AIR3* promoter-GUS activity. A, staining around lateral roots. B and C, details of staining at sites where lateral roots protrude or will protrude, viewed with Nomarski optics. D, GUS expression in a single isle at a position where activity in the pericycle is not visible. E, GUS expression in roots after 16 hr of 1 μ M 1-NAA treatment.

splicing of the pro-region. The putative pro-region splicing sites are not preceded by pairs of basic amino acids. This might indicate that, in contrast with other eukaryotic subtilisin-like proteases, the plant subtilisin-like proteases studied so far have a broad substrate specificity. This hypothesis is supported by *in vitro* activity measurements of cucumisin.²²

3.3. Study of transgenic *Arabidopsis* harboring an *AIR3* promoter-gusA fusion

Despite the structural similarities, plant subtilisin-like proteases are expressed in response to entirely different developmental and environmental stimuli. We investigated the expression pattern of *AIR3* to learn more about the specific function of the *AIR3* protein during lateral root formation. The sequence flanking the ATG initiation codon at the 5' end has been determined. A consensus CAAT box and TATA box can be found respectively 126 bp and 77 bp upstream from the ATG. The promoter area lacks clear homology to any known (promoter) sequence. Elements which are known to be important in the promoters of diverse auxin-inducible genes²³ cannot be found in the promoter sequence shown in Fig. 1B. An *AIR3* promoter-gusA construct was introduced into *Arabidopsis* by *Agrobacterium tumefaciens*-mediated transformation. Seedlings derived from three independent transformants with a 3:1 segregation for the kanamycin resistance gene in the T₂ and from six additional independent transformants with multiple copies of the T-DNA were tested for GUS activity. In the absence of exogenous auxin, staining was visible in the outer layers of the parental root around lateral roots and at sites where lat-

eral roots were about to emerge (Fig. 3A, 3B and 3C). Even at very early stages of lateral root development, when cell division activity in the pericycle was barely or not detectable, GUS activity could be observed in the outer layers of the root, suggesting that expression precedes the actual formation of the lateral root primordium (Fig. 3D). Upon addition of auxin, GUS expression increased along the length of the root except for the parts that are not competent for lateral root formation; i.e., the root tips, the meristematic regions and the zones directly behind these regions (Fig. 3E).

3.4. A possible role for *AIR3* subtilisin-like protease gene

So far, biochemical evidence for protease activity of the *AIR3* protein is lacking. By comparison with other plant subtilisin-like proteases, one is able to speculate about the role of *AIR3*. The plant subtilisin-like proteases that have been identified all possess a putative signal peptide for translocation over the endoplasmic reticulum and a pro-region, probably necessary to keep the protein inactive, while none of the proteins contains possible transmembrane regions or any other known targeting signals. This suggests that all known plant subtilisin-like proteases, including *AIR3*, are active outside the cell. For some of these proteins it has been shown that the mature, active forms are present extracellularly. The processed LIM9 protein for instance, has been identified in the locules of anthers by immunocytochemistry experiments and could be responsible for the degradation of cell wall matrices in order to release microspores.¹⁹ The mature PR-P69A protein has been isolated from the inter-

cellular spaces of viroid-infected leaves²¹ and was shown to interact with a cell wall matrix protein, named LRP.²⁴ In analogy with other plant subtilisin-like proteins, the most likely substrates for *AIR3* are structural proteins in the cell wall.

3.5. Conclusion

We have isolated and characterized the first complete plant subtilisin-like protease gene, the *AIR3* gene, and established its expression pattern by construction of a promoter-*gusA* fusion which was analyzed in *Arabidopsis* seedlings. The *AIR3* protein has the same structure as all plant subtilisin-like proteins identified so far. The promoter activity at sites of lateral root emergence suggests that *AIR3* digests structural proteins in the extracellular matrix in order to weaken the tissue and facilitate lateral root emergence.

Acknowledgments: The authors would like to thank Johan Pinas for assistance in preparation of Fig. 3.

References

1. Malamy, J. E. and Benfey, P. N. 1997, Organization and cell differentiation in lateral roots of *Arabidopsis thaliana*, *Development*, **124**, 33–44.
2. Celenza, J. L. Jr., Grisafi, F. L., and Fink, G. R. 1995, A pathway for lateral root formation in *Arabidopsis thaliana*, *Genes Devel.*, **9**, 2131–2142.
3. Romano, C. P., Robson, P. R., Smith, H., Estelle, M., and Klee, H. 1995, Transgene-mediated auxin overproduction in *Arabidopsis*: hypocotyl elongation phenotype and interactions with the *hy6-1* hypocotyl elongation and *arr1* auxin-resistant mutants. *Plant Mol. Biol.*, **27**, 1071–1083.
4. Neuteboom, L. W., Ng, J. M. Y., Kuyper, M., Clijdesdale, O. R., Hooykaas, P. J. J., and van der Zaal, B. J. 1999, Isolation and characterization of cDNA clones corresponding with mRNAs that accumulate during auxin-induced lateral root formation, *Plant Mol. Biol.*, **39**, 273–287.
5. Barr, P. J. 1991, Mammalian subtilisins: the long-sought dibasic processing endoproteases, *Cell*, **66**, 1–3.
6. Yamagata, H., Masuzawa, T., Nagaoka, Y., Ohnishi, T., and Iwasaki, T. 1994, Cucumisin, a serine protease from melon fruits, shares structural homology with subtilisin and is generated from a large precursor, *J. Biol. Chem.*, **269**, 32725–32731.
7. Kobayashi, T., Kobayashi, E., Sato, S. et al. 1994, Characterization of cDNAs induced in meiotic prophase in lily microsporocytes, *DNA Res.*, **1**, 15–26.
8. Ribeiro, A., Akkermans, A. D., van Kammen, A., Bisseling, T., and Pawlowski, K. 1995, A nodule-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development, *Plant Cell*, **7**, 785–794.
9. Tornero, P., Conejero, V., and Vera, P. 1996, Primary structure and expression of a pathogen-induced protease (PR-P69) in tomato plants: similarity of functional domains to subtilisin-like endoproteases, *Proc. Natl. Acad. Sci. USA*, **93**, 6332–6337.
10. Tornero, P., Conejero, V., and Vera, P. 1997, Identification of a new pathogen-induced member of the subtilisin-like processing protease family from plants, *J. Biol. Chem.*, **272**, 14412–14419.
11. Sabala, I., Franzén, H., and von Arnold, S. 1997, A spruce gene, *af70*, constitutively expressed in somatic embryos and induced by ABA and low temperature in seedlings, *Physiol. Plant.*, **99**, 316–322.
12. Sambrook, L., Fritsch, E. F., and Maniatis, T. 1989, *Molecular cloning: A laboratory manual*, second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
13. Pasquali, G., Ouwerkerk, P. B., and Memelink, J. 1994, Versatile transformation vectors to assay the promoter activity of DNA elements in plants, *Gene*, **149**, 373–374.
14. Hood, E. E., Gelvin, S. B., Melchers, L. S., and Hoekema, A. 1993, New *Agrobacterium* helper plasmids for gene transfer to plants, *Transgenic Res.*, **2**, 208–218.
15. Ditta, G., Stanfield, S., Corbin, D., and Helinski, D. R. 1980, Broad host range DNA cloning system for gram-negative bacteria: construction of gene bank of *Rhizobium meliloti*, *Proc. Natl. Acad. Sci. USA*, **77**, 7347–7351.
16. Vergunst, A. C., de Waal, E. C., and Hooykaas, P. J. J. 1998, Root transformation by *Agrobacterium tumefaciens*. In: *Methods in Molecular Biology*, vol 82: *Arabidopsis* Protocols. Eds: J. Martínez Zapaler and J. Salinas. Humana Press Inc, Totowa NJ 227–244.
17. Murashige, T. and Skoog, F. 1962, A revised medium for rapid growth and bioassays with tobacco tissue culture, *Physiol. Plant.*, **15**, 473–497.
18. Jefferson, R. A. 1987, Assaying chimeric genes in plants: the GUS gene fusion system, *Plant Mol. Biol. Rep.*, **5**, 387–405.
19. Taylor, A. A., Horsch, A., Rzepczyk, A., Hasenkampf, C. A., and Riggs, C. D. 1997, Maturation and secretion of a serine proteinase is associated with events of late microsporogenesis, *Plant J.*, **12**, 1261–1271.
20. Leduc, R., Molloy, S. S., Thorne, B. A., and Thomas, G. 1992, Activation of human furin precursor processing endoprotease occurs by an intramolecular autoproteolytic cleavage, *J. Biol. Chem.*, **267**, 14304–14308.
21. Vera, P., Hernández-Yago, J., and Conejero, V. 1989, Immunogold localization of the citrus exocortis viroid-induced pathogenesis-related proteinase P69 in tomato leaves, *Plant Physiol.*, **91**, 119–123.
22. Kaneda, M. and Tominaga, N. 1975, Isolation and characterization of a protease from the sarcocarp of melon fruit, *J. Biochem.*, **78**, 1287–1296.
23. Abel, S., Ballas, N., Wong, L. M., and Theologis, A. 1996, DNA elements responsive to auxin, *Bioessays*, **18**, 647–654.
24. Tornero, P., Mayda, E., Dolores Gómez, M., Cañas, L., Conejero, V., and Vera, P. 1996, Characterization of LRP, a leucine-rich repeat (LRR) protein from tomato plants that is processed during pathogenesis, *Plant J.*, **10**, 315–330.

