cDNA cloning and characterization of a proline (or hydroxyproline)-rich protein from *Santalum album* L.

Anirban Bhattacharya and G. Lakshmi Sita*

Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560 012, India

A proline (or hydroxyproline)-rich cDNA clone, *SaPRP*, was isolated from sandalwood (*Santalum album* L.) somatic embryos pretreated with salicylic acid. The longest open reading frame in *SaPRP* encodes a polypeptide of 326 amino acids. It reveals 48% identity in 233 amino acids overlap with a proline-rich glycoprotein from maize. Southern hybridization with sandalwood genomic DNA digests suggests that *SaPRP* possibly belongs to a small gene family. Northern blot analysis shows that this *SaPRP* is expressed predominantly in leaf tissues without salicylic acid induction. The induction of this *SaPRP* was observed in the somatic embryos treated with salicylic acid.

SANDALWOOD (*Santalum album* L.) is one of the important trees of commercial value known for its highly priced essential oil. Sandalwood production has come down due to severe disease. Breeding for disease resistance in trees in general poses several problems and is not practical. Tree improvement in sandalwood mainly aims at evolving trees that can yield more heartwood and oil. It is imperative that they are resistant to diseases and attack by heartwood borers and other pests. Biotechnology allows precise cloning of disease resistance genes and their transfer into chosen plants. Plants have evolved an array of biochemical response to defend against invading organisms like fungi, bacteria, virus, mycoplasma, etc. The plant defence response is accomplished by activation of a plethora of defence-related genes such as pathogenesis-related (PR) genes, a group of inducible host plant encoded proteins whose synthesis is associated with resistance to plant pathogens as well as various forms of physical and chemical stress[1]. In addition, plants also induce structural changes in the cell wall composition with the accumulation of cell wall proteins as a result of infection. These include extensins, which are hydroxyproline-rich glycoprotein (HRGP)[2], glycine-rich protein (GRP)[3,4] and proline-rich protein (PRP) which are regarded as reactive biopolymers[5,6]. These proteins are characterized by basic repeat motifs that vary among the different classes of cell wall proteins: Ser-$(Hyp)_4$ for extensin, $(Gly\text{-}X)_n$ for GRP and Pro–Pro–Val–X–Y for PRP and are generally specified by plant gene families that are induced during infection as well as by biotic and abiotic elicitors. In contrast to the situation in crop plants, information available in tree species is somewhat limited[7]. There are reports indicating that over-expression of pathogenesis-related genes like chitinase and glucanase in tobacco has resulted in transgenic plants with somewhat more tolerance against fungal diseases[8]. Hence we have undertaken the present study to clone defence-related genes and over-express in transgenic sandalwood plants. We report here the successful cloning and characterization of a proline-rich protein, cDNA (SaPRP) from sandalwood.

Materials and methods

Plant material

Somatic embryos used for this study were obtained by direct somatic embryogenesis (Lakshmi Sita *et al.*, unpublished work) from internodal segments of young shoots. Briefly, explants were inoculated in MS medium supplemented with thidiazuran (TDZ) and benzyl amino purine (BAP) for direct somatic embryogenesis. Globular embryos thus obtained were transferred to MS medium supplemented with gibberellic acid (GA). 3–4-week-old somatic embryos were used for induction with salicylic acid (SA) in subsequent experiments. Leaf and stem tissues used for expression studies are collected from mature trees.

Construction and screening of cDNA library

3–4 week old somatic embryos were treated with 10 mM SA for 24 h. Total RNA was prepared using GITC-acid phenol extraction method[9]. Poly A + RNA was prepared using messenger affinity paper (Amersham). Double stranded cDNA was prepared, adaptor ligated and subsequently cloned in λgt-11 using Amersham's cDNA synthesis plus cloning module kit. The library thus prepared was amplified before screening. It was then screened by duplicate plaque hybridization method using soybean extensin cDNA clone (kindly provided by Dr J. Hong, USA) as a probe. The plaques were transferred to nylon membrane (Hybond-N, Amersham) and prehybridized for 4 h at 55°C in 6XSSC, 5X Denhardt's solution, 0.1% (w/v) SDS and 250 μg/ml salmon sperm

*For correspondence. (e-mail: sitagl@mcbl.iisc.ernet)

DNA[3]. The labelling of soybean extensin cDNA insert with α-^{32}P dATP was done by the megaprime labelling kit (Amersham). Hybridization was carried out for 24 h under same conditions as the prehybridization. Filters were finally washed under moderately stringent conditions (0.5X SSC, 0.1% SDS at 50°C). Positive plaques were selected and subjected to another round of screening at low density, to ensure purity and elimination of false positives. Positive plaques were then subjected to PCR analysis using λ-gt11 primers to check the size of the inserts. The authenticity of the largest sized cDNA (1.8 kb) was confirmed by Southern hybridization using the same probe under the same conditions. Lambda DNA was isolated as described by Ausubel *et al.*[10]. The *Eco*RI insert was finally subcloned in pGem3Zf(–) to yield *pSaPRP*.

Sequencing of pSaPRP

Sequencing was carried out either by Sequenase (ver 2.0) DNA sequencing kit (USB Biochemicals) or by the automated sequencing facility. The analysis of the amino acid sequence was carried out using Swiss Prot data base.

DNA isolation and blot hybridization

Total genomic DNA was extracted from *S. album* tender leaves as described by Dellaporta *et al.*[11]. Aliquots of 10 µg DNA were digested with different restriction enzymes, separated in 0.8% agarose gel and transferred onto a nylon membrane, Hybond-N (Amersham). Prehybridization and hybridization were carried out[12] in 50% formamide, 6XSSC, 5X Denhardt's solution, 0.1% SDS and 250 µg/ml salmon sperm DNA at 42°C. The 1.8 kb *pSaPRP* insert was labelled with α-^{32}P dATP by the megaprime labelling kit (Amersham) and used as probe. The blots were finally washed at high stringency conditions, 0.1XSSC, 0.1% SDS at 65°C for 30 min.

RNA isolation and Northern blot hybridization

Total RNA was isolated, according to Chomczynski and Sacchi[9], from sandalwood leaves, stem pieces, somatic embryos either treated with SA or mock treated with water. Prior to electrophoresis, the RNA was suspended in 50% formamide, 20 mM 3-(*N*-morpholino)-propane-sulphonic acid (MOPS), 5 mM sodium acetate, 1 mM EDTA, 6% formaldehyde, 6% glycerol and incubated at 65°C for 5 min. 10 µg of RNA was applied per lane on a 1.2% agarose formaldehyde (40% v/v) gel. Ethydium bromide was added in the gel to confirm the quality of the samples. After electrophoresis, the RNA was transferred onto a nylon membrane, Hybond-N (Amersham). The hybridization conditions were the same as those described in Southern blotting experiments. The filters were finally washed at 55°C for 20 min in 1X SSC, 0.1% SDS. The same blots were reprobed with rice rRNA gene probe to confirm equal loading of the samples. All these hybridization experiments were carried out at least three times to confirm the observations.

Results

Isolation and sequence analysis of the proline-rich protein cDNA clone

The cDNA library, constructed from SA-treated sandalwood somatic embryos was screened with soybean extensin cDNA as probe under moderately stringent conditions. After multiple round of screening, 21 positive plaques were picked up for PCR analysis using λgt-11 primers. The size of the inserts was determined by agarose gel electrophoresis and the largest one, of the size 1.8 kb after subcloning in pGEM3Zf(–), was taken up for further characterization. Authenticity of the insert at this stage was confirmed by Southern hybridization of the PCR fragment (data not shown) with heterologous probe using the same conditions as those used during screening. The complete sequence of the clone, *pSaPRP*, and the deduced amino acid sequence of the encoded protein was determined. The cDNA contains 1744 bp with a 978 nucleotide open reading frame. The polypeptide contains repeat units of SPTPP and related sequences throughout (Figure 1).

Sequence comparison

The sequence of the polypeptide was aligned with related sequences and compared. A low but significant identity (30–48%) was obtained with various extensin precursors as shown in Table 1.

Genomic organization of SaPRP

Total DNA was digested to completion with *Eco*RI, *Hin*dIII, *Bam*HI, *Bst*NI and *Xho*I and subjected to Southern blot hybridization with SaPRP insert as probe (Figure 2). *Bam*HI has one internal site within the cDNA while the others have none. Along with the major expected bands there seem to be other weak bands in different lanes. In view of earlier reports from other plants this suggests that *SaPRP* possibly belongs to a small gene family[5,6].

Induction of SaPRP transcript

Northern blot analysis with *SaPRP* was used to monitor the expression in somatic embryos treated with SA (Figure 3). The results indicate the clear induction of

SaPRP transcript in the embryos when treated with SA. The probe hybridizes specifically to single mRNA species of the appropriate size. From a separate experiment we observe that longer exposure results in the appearance of weak signal in the uninduced RNA lane (Figure 4). As SA has been implied as a major signaling molecule in local as well as systemic resistance, our data suggest that this *SaPRP* induction may also be a part of defence response (see discussion).

```
tgcttcacaatgataggaagagccgacatcgaaggatcaaaaagcaacgtcgctatgaac

gcttggctgccacaacc aagttatccctgtggtaacttttctgacacctctagcttcaaa

ttccgaagttctaaaggatcgtcgataggccacgctttcacggttcgtattcgtactgaa

aatcagaatcaaacgagcttttacccttttgttccacacccaatttgagttctcgttgag

ctcatcttaggacacctgcgttatcttttaacagatatgccgccccagccaaactcccca
                                       M  P  P  Q  P  N  S  P

cctgacaatgtcttcccgccggatcgatgcacgagcgtgcgtcttggttctaaaaagagg
P  D  N  V  F  F  P  D  R  C  T  S  V  R  L  G  S  K  K  R

ggcgagagccccgctcccgattcacggaataagaataacgttaaaagtagtggtatttca
G  E  S  P  A  P  D  S  R  N  K  N  N  V  K  S  S  G  I  S

cttttgccgccgtctgctcccacttatcccacacctctccaagtcatttcacaagtcgg
L  L  P  P  S  A  P  T  Y  P  T  P  L  Q  V  I  S  Q  S  R

actaaaattccaagcccaacattgggtcttcctttcccccgctggatccaagaaaaccc
T  K  I  P  S  P  T  L  G  L  P  F  P  P  L  D  P  R  K  P

ccaagtgggcctttggtaatccaaattcaccggcgtcactaataccttggcgacgctgg
P  S  G  P  F  G  N  P  N  S  P  A  S  L  I  P  W  R  R  W

gaccatactcccccgttgtttaaccctccatcggatcccccactccgtacggcacccca
D  H  T  P  P  L  F  N  P  P  S  D  P  P  T  P  Y  G  T  P

cccactgttcgacgcagcccgagcccccggaaagtgccagccccagtccgtcccccgga
P  T  V  R  R  S  P  S  P  P  E  S  A  S  P  S  P  S  P  G

ccgcccgcggcgacccccactccccgcccaagcagcccgacgagtaccccaccaacagat
P  P  A  A  T  ?  T  P  R  P  S  S  P  T  S  T  P  P  T  D

ggggaaacacccaggccgccggggcctcgaagccctagcccctctccaggtccgcccagt
G  E  T  P  R  P  P  G  P  R  S  P  S  P  S  P  G  P  P  S

tgttcgcctagcccaaagagccctagcccgccggcaagcagtccaccacgctcccggccc
C  S  P  S  P  K  S  P  S  P  P  A  S  S  P  P  R  S  R  P

ggacccccgactacaacaacgtctcctagcccccaacccctaggagcgtgcctccaact
G  P  P  D  Y  T  T  S  P  S  P  P  T  P  R  S  V  P  P  T

ccaccggctagtccttcccccgccgaccgctaaaccctcccctccgtcacgggggtcttcc
P  P  A  S  P  S  P  P  T  A  K  P  S  P  P  S  R  G  S  S

ccgagcccacctacatccgccgacgcccacgccgcggcccccatcctactcaccaagcccg
P  S  P  P  T  S  P  T  P  T  P  R  P  P  S  Y  S  P  S  P

acgccaccaagcagcccggcccagccctcctctcaggagtcccatactaaccccgccttct
T  P  P  S  S  ?  P  S  P  P  L  R  S  P  I  L  T  P  P  S

ccagcagccgttcctcctatcggacggagccctccgtcgcctatcgacccaccatgcagc
P  A  A  V  P  P  I  G  R  S  P  P  S  P  I  D  P  P  C  S

agccccgagccgtccagcccgccaaccagtccacccactccgtcaagcccctactagcac
S  P  E  P  S  S  P  P  T  S  P  P  T  P  S  S  P  Y  *

ctggggagtgagcccggcggagcgtccgtcggtgcggatcttggtggtagtagcaaata

ttcaaatgaaactttgaaggccgaaaggggaaaggttccatgtgaacggcacttgcaca

ggattagtcgatcctaagagacgggggaagcccgtccgataccgtgcaccacgcgagctt

cgaaagggaatcgggttaaaattcctcaaccgggatgtggcggctgacggcaaccgttag

ggagtccggagacgtctcgagggcctcgggaagagttatcttttctgtttaacagcctgc

ccaccctgaaacggtttaccggaggtagggtccacggctggaacagcaccgcacgtcgcg

tggtgtctggtgcgcc tcgcggcccttgaaaatccggaggaccgagtgccttccacgcc

tggtcgtactcataac gcatcaggtctccaaggtgaacaacctctggtccatggaacat

ttac
```

Figure 1. Nucleotide and the deduced amino acid sequences of *SaPRP* cDNA. The stop codon is marked by an asterisk. SPSPP/SPTPP repeat sequences are underlined.

Tissue specificity of SaPRP expression

The cell wall proteins being implied in spatial and temporal expressions, Northern blot analysis was further extended using leaf and stem tissues. As mentioned in the previous experiment, RNA from uninduced embryos was loaded in one of the lanes. Evidently from Figure 4, *SaPRP* is predominantly expressed in leaf tissues. The probe here also hybridizes specifically to a single band, however the size of the transcript appears to be different than that seen in case of embryo samples (induced or uninduced).

Discussion

Tree improvement programmes are well in progress in various tropical and temperate trees. As a preliminary step towards developing disease-free plants, tissue culture and genetic transformation in sandalwood are already established[13-15]. Plants were produced via somatic embryo-

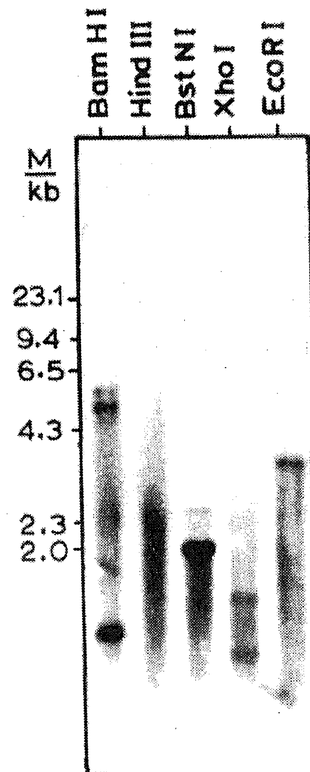

Figure 2. Southern hybridization analysis. Genomic DNA (10 µg per lane) was digested with the indicated restriction enzymes. The migration of the molecular weight standards is indicated.

Table 1. Percentage of identity of aligned amino acid sequences of proline (hydroxyproline)-rich glycoprotein

	SaPRP	Maize extensin	Tobacco extensin	Sorghum extensin	Arabidopsis anther specific extensin	Tobacco pistil-specific extensin	
SaPRP	100	48.1	34	33.3	37.6	30.6	Percent identity
SaPRP	326	233	314	207	149	216	amino acid overlap

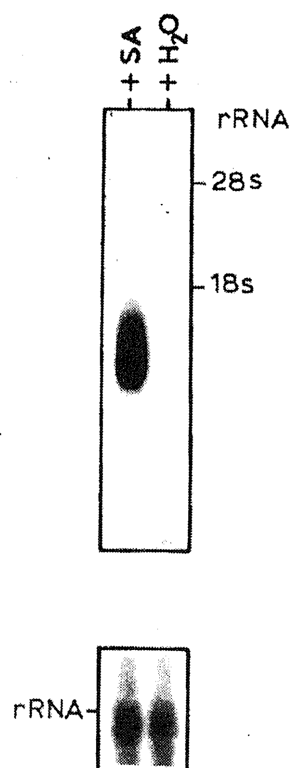

Figure 3. Expression of *SaPRP* transcript on SA-treated somatic embryos. 10 μg of total RNA was loaded in each lane. Upper panel shows the induction of *SaPRP* mRNA in somatic embryos of sandalwood when treated with salicylic acid. The same blot was then washed and rehybridized with *rRNA* gene probe, shown in the bottom panel.

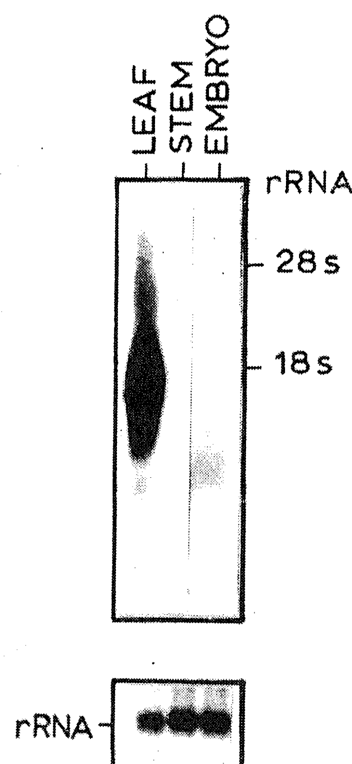

Figure 4. Tissue specificity of *SaPRP* transcript. 10 μg of total RNA from leaf, stem and embryo was loaded on each lane as indicated. The upper panel shows the expression of *SaPRP* transcript in leaf tissue. Embryo RNA lane was over exposed. Faint appearance of transcript in shorter length can be observed (see discussion). The same blot was reprobed with *rRNA* gene as shown in the bottom panel.

genesis from callus cultures, derived from mature phenotypically disease-free plants. Though it was possible to produce hundreds of plants which could be established successfully, it was not certain whether these are genetically resistant. There are reports of transgenic plants constitutively expressing high levels of chitinase and glucanase with enhanced resistance[8,16] to various pathogens. This prompted us to study the defence responses in sandalwood which could be used in developing strategies for protection through genetic manipulation. Towards this end, cloning of defence-related genes was attempted.

SA has been proposed to have a central role as a signaling molecule leading to systemic acquired resistance (SAR), as its concentration rises dramatically after pathogen infection[17–25]. Furthermore, exogenously applied SA leads to typical SAR response such as increased resistance to viral infection[24–26]. Salicylic acid being implied as a major signaling molecule in local as well as systemic defence response[27,28], somatic embryos were pretreated with SA for the preparation of the cDNA library. Subsequently, a proline (or hydroxyproline)-rich protein clone could be isolated from cDNA library. From the genomic hybridization studies and also from other reports[29], *SaPRP* possibly belongs to a small gene family. Sequence analysis of the same reveals its relatedness with other proteins isolated from different systems. As it turns out, though sandalwood *SaPRP* definitely shares certain identity with other related proteins, however, it has some unique features. It has the repeat motif of

SPTPP instead of SPPPP as described in extensins. Interestingly, it shows identity with certain tissue-specific extensin precursors also (Table 1). Tissue specificity of these groups of proteins has been reported in many crop plants[30,31]. To check whether *SaPRP* also has some tissue specificity Northern blot analysis was done. It was clear that *SaPRP* is expressed abundantly in leaf tissues compared to stems and barely in somatic embryos. Significant induction of the same was observed in somatic embryos when treated with SA. To the best of our knowledge, all the reported extensins are rather expressed less in leaf tissues. However, one of the proline-rich proteins from soybean, SbPRP3 was reported to be abundant in leaves[30]. Thus, in functional terms *SaPRP* is closer to proline (or hydroxyproline)-rich protein group rather than extensins. However, at present it is difficult to explain the functional significance of the absence of SaPRP transcript in stems. Expression studies reveal that this SaPRP is highly inducible by SA, justifying our ability to clone the same from the SA-induced embryos. This also reflects that the induction of this gene may occur during wounding or infection through SA-mediated signal transduction pathway. It was observed that the size of the transcript present in leaf tissue is different from that present in the induced or uninduced embryos. It is not clear whether these two transcripts, come from different genes or from the same gene. It is possible that these originate from the same gene and are processed differently. Moreover, the gene seems to be developmentally regulated while expression of it becomes significant only at the later stages of development in a tissue-specific manner. The expression as a consequence of SA treatment at the early embryo stage is probably a part of defence response. However, this does not explain the apparent difference in size of the transcripts. If these transcript populations are derived from different genes one would expect to detect an additional, SA-induced, transcript of different sizes in the leaf or stem tissue when treated with elicitors. Experiments are now directed to answer these questions. The present report of cloning of defence-related gene is the first of its kind in the forest trees. For disease control of mycoplasma and other diseases in sandalwood this could well be a beginning. Genetic transformation with marker genes is already established[15]. The protocols developed will be used for the over-expression of defence-related genes.

1. Carr, J. P. and Klessig, D. F., in *Genetic Engineering: Principles and Methods* (ed. Setlow, J. K.), Plenum Press, New York, 1989, pp. 65–109.
2. Chen, J. and Varner, J. E., *Proc. Natl. Acad. Sci. USA*, 1985, 4399–4403.
3. Condit, C. M. and Meagher R. B., *Nature*, 1986, **323**, 178–181.
4. Keller, B., Sauer, N. and Lamb, C. J., *EMBO J.*, 1988, **7**, 3625–3633.
5. Hong, J. C., Nagao, R. T. and Key, J. L., *J. Biol. Chem.*, 1987, **262**, 3867–3876.
6. Hong, J. C., Nagao, R. T. and Key, J. L., *J. Biol. Chem.*, 1990, **265**, 2470–2475.
7. Bao, W., O'Malley, D. M. and Sederoff, R. R., *Proc. Natl. Acad. Sci. USA*, 1992, **39**, 6604–6608.
8. Zhu, D., Maher, E. A., Masound, S., Dixon, R. A. and Lamb, C. J., *Biotechnology*, 1994, **12**, 807–812.
9. Chomczynski, P. and Sacchi, N., *Anal. Biochem.*, 1987, **162**, 156–159.
10. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., *Current Protocols in Molecular Biology*, Greene Publishing Associates/Wiley Interscience, 1992.
11. Dellaporta, S. L., Wood, J. and Hicks, J. B., in *Molecular Biology of Plants: A Laboratory Course Manual* (eds Malmberg, R., Messing, J. and Sussex, I.), Cold Spring Harbor Press, New York, 1983, pp. 36–37.
12. Sambrook, J., Fritsch, E. F. and Maniatis, T., in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, Cold Spring Harbor Laboratory Press, 1989.
13. Lakshmi Sita, G., Raghava Ram, N. V. and Vaidyanathan, C. S., *Plant Sci. Lett.*, 1979, **20**, 63–90.
14. Lakshmi Sita, G., in *Biotechnology of Agriculture and Forestry, vol. 2, Trees* (ed. Bajaj, Y. P. S.), Springer Verlag, 1986, pp. 363–374.
15. Lakshmi Sita, G., Srinivas, G. L. S. and Anirban Bhattacharya, *Plant Tissue Culture Biotechnol.*, 1998, **4** (in press).
16. Broglie, K., Chet, I., Holliday, M., Cressman, R., Biddle, P., Knowlton, S., Manvais, C. J. and Broglie, R., *Science*, 1991, **254**, 1194–1197.
17. Delaney, T., Uknes, S. and Vernooij, B., *Science*, 1994, **266**, 1247–1250.
18. Enyedy, A. J., Yalpani, N., Silverman, P. and Raskin, I., *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 2480–2484.
19. Malamy, J., Carr, J. P., Klessig, D. F. and Raskin, I., *Science*, 1990, **250**, 1002–1004.
20. Metraux, J. P., Signer, H., Ryals, J., Ward, E., Wyness-Benz, M., Gaudin, J., Raschdorf, K., Schmid, E., Blum, W. and Inverardi, B., *Science*, 1990, **250**, 1004–1006.
21. Rasmussen, J. B., Hammerschmidt, R. and Zook, M. N., *Plant Physiol.*, 1991, **97**, 1342–1347.
22. Uknes, S., Winter, A. M., Delany, T., Vernooij, B., Morse, A., Friedrich, L., Nye, G., Potter, S., Ward, E. and Ryals, J., *Mol. Plant–Microbe Interact.*, 1993, **6**, 692–698.
23. Yalpani, N., Silverman, P., Wilson, T. M. A., Klier, D. A. and Raskin, I., *Plant Cell*, 1991, **3**, 809–818.
24. Malamy, J. and Klessig, D. F., *Plant J.*, 1992, **2**, 643–654.
25. White, R. F., *Virology*, 1979, **99**, 410–412.
26. Ye, X. S., Pan, S. Q. and Kuc, J., *Physiol. Mol. Plant. Pathol.*, 1989, **35**, 161–175.
27. Davis, K. R. and Ausubel, F. M., *Mol. Plant–Microbe Interact.*, 1989, **2**, 363–368.
28. Ryals, J. A., Neuenschwander, Urs, H. and Willits, M. G., *Plant Cell*, 1996, **8**, 1809–1819.
29. Showalter, A. M. and Varner, J. E., in *The Biochemistry of Plants* (eds Stumpf, P. K. and Conn, E. E.), Academic Press, New York, 1989, vol. 15, pp. 247–281.
30. Hong, J. C., Nagao, R. T. and Key, J. L., *Plant Cell*, 1989, **1** 937–943.
31. Wyatt, R. E., Nagao, R. T. and Key, J. L., *Plant Cell*, 1992, **4**, 99–110.

ACKNOWLEDGEMENTS. We thank Dr J. Hong for kindly providing the soybean extensin clone used in this study. We also thank Department of Biotechnology (DBT) sponsored, automated sequencing facility in the Department of Biochemistry, IISc. G.L.S. thanks DBT for financial assistance for this study. A.B. is the recipient of fellowship from the University Grants Commission. The nucleotide sequence data reported will appear in GenBank under the accession number AF020261.

Received 25 May 1998; revised accepted 3 August 1998