The EMBO Journal vol.5 no.1 pp.37–40, 1986

Molecular characterization of messenger RNAs for 'pathogenesis-related' proteins 1a, 1b and 1c, induced by TMV infection of tobacco

Ben J.C.Cornelissen, Rob A.M.Hooft van Huijsduijnen, Leendert C.Van Loon[1] and John F.Bol

Department of Biochemistry, State University of Leiden, Wassenaarseweg 64, 2333 AL Leiden, and [1]Department of Plant Physiology, Agricultural University, Arboretumlaan 4, 6703 BD Wageningen, The Netherlands

Communicated by P.van de Putte

A cDNA library was made to poly(A)-containing RNA from tobacco mosaic virus (TMV)-infected Samsun NN tobacco plants and clones corresponding to mRNAs for the 'pathogenesis-related' (PR) proteins 1a, 1b and 1c were identified. One clone was found to contain a complete copy of PR-1b mRNA. The structural organization of this RNA is: a leader sequence of 29 nucleotides, an open reading frame of 504 nucleotides encoding a 30 amino acid long signal peptide and a 138 amino acid long mature protein, and a 3'-non-coding region of 235 nucleotides. Two other clones were found to contain partial copies of PR-1a and PR-1c mRNAs. The data indicate an ~90% homology between the amino acid sequences of PR-1a, -1b and -1c. Using one of the clones as probe it was shown that in the TMV-inoculated lower leaves and the non-inoculated upper leaves of a tobacco plant, the PR-1 mRNAs become detectable from 2 and 8 days after inoculation, respectively.

Key words: gene family/pathogenesis-related proteins/TMV infection/tobaccco

Introduction

Infection of tobacco leaves with tobacco mosaic virus (TMV) induces the synthesis of a number of pH 3-soluble and protease-resistant proteins which accumulate in the intercellular leaf space. A similar induction can be obtained by the application of culture filtrates from a pathogenic fungus or bacterium, by various plant hormones or by a variety of chemicals, notably polyacrylic acid and benzoic acid derivatives such as salicylic acid or aspirin (for review, see Van Loon, 1985). These proteins have been named 'pathogenesis-related' proteins, or PRs, and a unified system of nomenclature has been proposed which groups similar proteins from different tobacco cultivars (Antoniw *et al.*, 1980). The appearance of PR proteins is closely associated with the phenomenon of 'systemic acquired resistance' and it has been suggested that they have an anti-viral function (Kassanis *et al.*, 1974; Van Loon, 1975). PR proteins induced in tobacco cultivars Samsun NN or Xanthi nc can be separated in >10 components. The predominant components, PR-1a, -1b and -1c are related by mol. wt., amino acid composition and serology (Antoniw *et al.*, 1980, 1985; Matsuoka and Ohashi, 1984; Hooft van Huijsduijnen *et al.*, 1985).

PR proteins, notably associated with necrotic reactions, have now been detected in 16 plant species (Van Loon, 1985). Recently, Lucas *et al.* (1985) reported the complete amino acid sequence of a 14-kd PR protein (p14) induced by viroids in tomato, and a few partial sequences of Samsun NN tobacco protein PR-1a. These data indicated an extensive homology between p14 from tomato and PR-1a from tobacco. We used synthetic deoxyoligonucleotides complementary to the mRNA sequence encoding the N-terminal region of PR-1a to sequence the 5' end of the mRNA (Hooft van Huijsduijnen *et al.*, 1985). This showed that the PR-1 proteins are synthesized as precursors, containing an N-terminal signal peptide of 30 amino acids. Here we used one of the synthetic deoxyoligonucleotides as a probe to screen a cDNA library of poly(A)-containing RNA from TMV-infected tobacco. Clones corresponding to the mRNAS for PR-1a, -1b and -1c were identified and one of them was used as probe to monitor the induction of PR-1 mRNAs in inoculated and virus-free leaves of TMV-infected tobacco plants.

Results

Characterization of PR-1 clones

Poly(A)-containing RNA purified from Samsun NN tobacco 7 days after the plants had been inoculated with TMV, was used as template to construct a cDNA library. This library was screened with a 5' ^{32}P-labeled synthetic deoxyoligonucleotide complementary to PR-1 mRNAs as probe ('primer 2' used by Hooft van Huijsduijnen *et al.*, 1985). A positively hybridizing clone (clone 110) was used as probe to identify two other PR clones (clones 69 and 119). Sequencing of these three clones revealed that they corresponded to three highly homologous but different mRNAs. The inserts of clones 69, 110 and 119 were found to be 768 bp, 413 bp and 454 bp, respectively. Clones 69 and 119 contained stretches of eight and six A residues, respectively, at the 3' end of the insert, probably derived from the poly(A) tail of the messenger; clones 69 and 110 contained sequences homologous to the 5'-terminal sequence of PR-1 mRNA deduced previously (Hooft van Huijsduijnen *et al.*, 1985). This indicates that clone 69 is a complete copy of PR-1 mRNA.

To permit a correlation of clones 69, 110 and 119 with the respective PR-1 proteins, the coding regions of the clones were compared with available amino acid sequence data. Lucas *et al.* (1985) reported the sequence of the N-terminal 40 amino acids of PR-1a in addition to an internal sequence of 29 amino acids and a C-terminal sequence of 12 amino acids. The coding regions of clones 69 and 110 showed differences in nine and six positions, respectively, with the available PR-1a sequences. However, the internal and C-terminal PR-1a sequences matched completely with the coding region of clone 119 with one exception. Amino acid number 7 from the C terminus of PR-1a was identified as tryptophan by Lucas *et al.* (1985) whereas a tyrosine is encoded in this position in clone 119. The amino acid sequence WVNEKQYYAHD has been identified in PR-1c (J.Lucas, personal communication). This sequence is encoded by clone 110, whereas the corresponding sequence in clones 69 and 119 differs in two positions. Cyanogen bromide cleavage of purified PR-1b and amino acid sequencing of the N-terminals of the cleavage products showed the presence in PR-1b of two methionine residues, preceding the amino acid sequences TAAKAV and WVDEKQ,

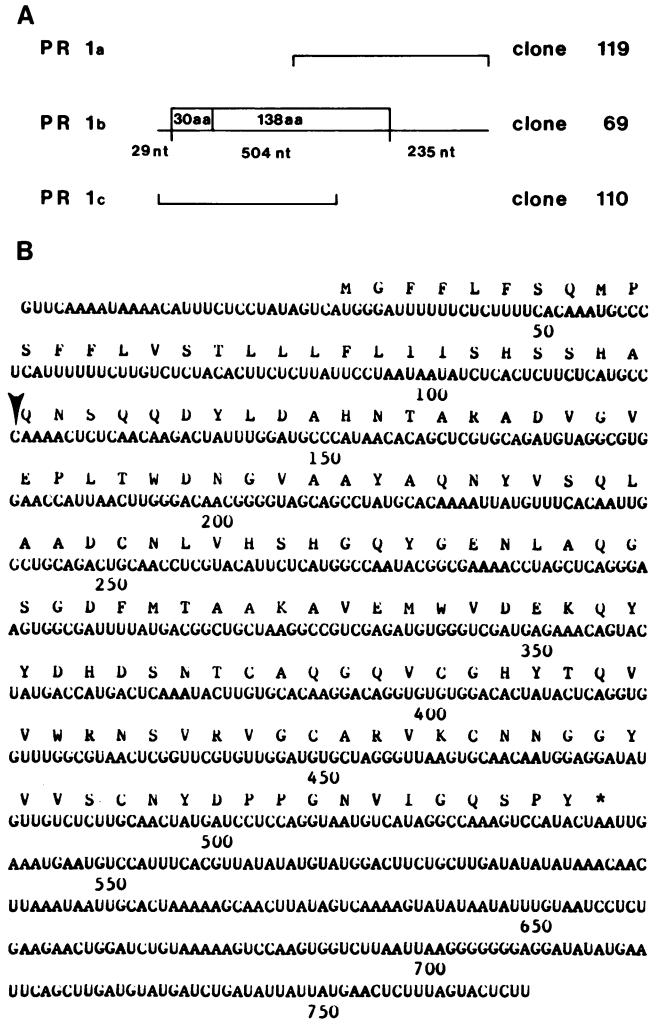

Fig. 1. Schematic representation of the PR-1a, -1b and -1c mRNA sequences present in clones 119, 69 and 110, respectively (**A**), and complete nucleotide sequence of PR-1b mRNA (**B**). The amino acid sequence encoded by the open reading frame in PR-1b mRNA is indicated; the N terminus of the mature PR-1b protein is marked with an arrow (aa = amino acids; nt = nucleotides).

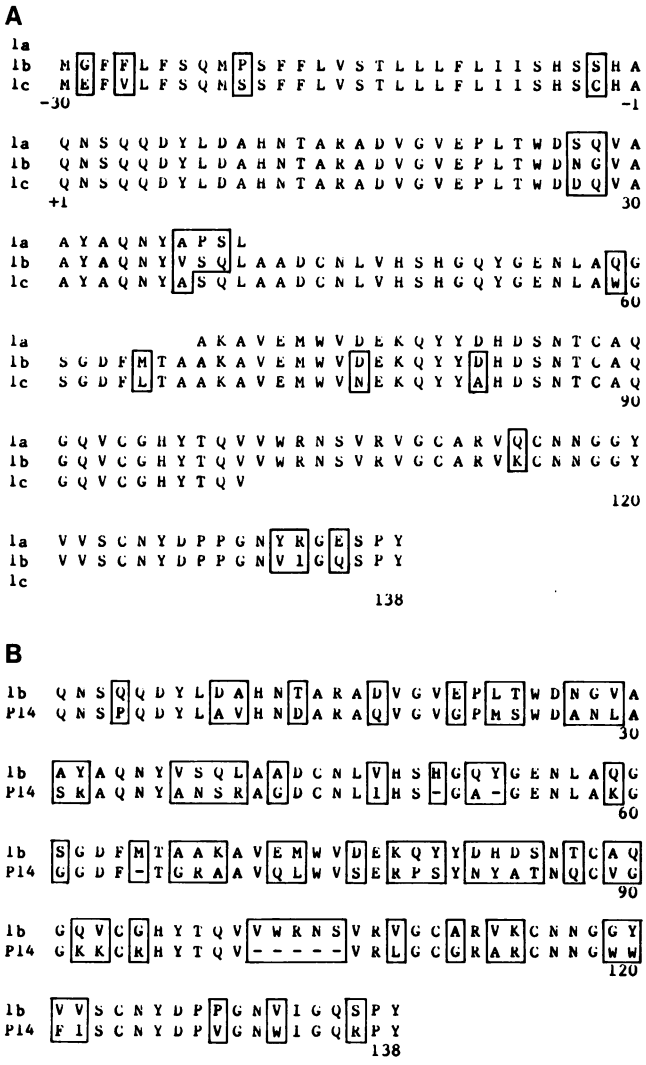

Fig. 2. Comparison of the amino acid sequence of the PR-1b protein with (**A**) the partial sequences of PR-1a and -1c and (**B**) the sequence of p14 from tomato. Differences are indicated by boxes. The sequence of p14 and of amino acids 1–40 of PR-1a is from Lucas *et al.* (1985); other sequences are derived from clones 69, 110 and 119. The numbering refers to the PR-1b sequence; (**A**) and (**B**) show the sequence of the PR-1b precursor and the mature protein, respectively.

respectively (R.A.M.Hooft van Huijsduijnen, unpublished results). The second methionine is present in all three clones. The first methionine, however, is present in clone 69 but not in clone 110 (the corresponding coding region is not represented in clone 119). In summary, the data strongly indicate that clone 69 is a complete copy of PR-1b mRNA and that clones 119 and 110 are partial copies of PR-1a and -1c mRNAs, respectively. A fully unambiguous assignment, however, must await the complete sequencing of all PR-1 mRNAs and proteins. In addition to the three clones described here, we have sequenced four other PR-1 clones. These were found to contain partial sequences of clones 69, 110 and 119, but no evidence for the existence of more than three PR-1 mRNAs was obtained.

Figure 1A shows the structural organization of PR-1b mRNA: a leader sequence of 29 nucleotides, an open reading frame of 504 nucleotides encoding the 30-amino acid signal peptide (Hooft van Huijsduijnen *et al.*, 1985) and the mature PR-1b protein of 138 amino acids, and a 3′-non-coding region of 235 nucleotides excluding the poly(A) stretch. The corresponding parts of the PR-1a and -1c mRNAs present in clones 119 and 110 are indicated. Figure 1B shows the complete nucleotide sequence of PR-1b mRNA and the amino acid sequence of its translation product. The mol. wt. of PR-1b was calculated to be 15 082 daltons, close to the value of 14 200 daltons determined by hydrodynamic methods (Antoniw and Pierpoint, 1978).

Comparison of PR-1 proteins

Figure 2 shows a comparison of the PR-1b amino acid sequence with the partial sequences of PR-1a and -1c. In addition, a comparison is made with the 130 amino acid long p14 protein induced by viroids in tomato (Lucas *et al.*, 1985). The homology of the partial PR-1a and -1c sequences with the PR-1b sequence is >90%. A similar degree of homology is found between the signal peptides of PR-1b and -1c. These homologies correlate well with the overall nucleotide sequence homologies of clones 119 and 110 with clone 69, being 87% and 91%, respectively (data not shown). The internal PR-1a sequence reported by Lucas *et al.* (1985) starts at the methionine residue at position 73; the methionine at position 65 is characteristic of the PR-1b clone.

The homology between PR-1b and p14 is somewhat less, be-

```
PR-1b  GUUCAAAAUAAAACAUUUCUCCUAUAGUC AUG CODING  REGION UAA
PR-1c  AUUCAAAGUACAACAUUUCUCCUAUAGUC
       *      * *            25

         *   *  *** *              *           ***
PR-1a  UUCAAACGACCUACGUCCAUUUCACGUUA-AUAUGUAUGGAUUGUUCUGCUUG
PR-1b  UUGAAAUGA---AUGUCCAUUUCACGUUAUAUAUGUAUGGAC--UUCUGCUUG
                  550                       575

          *****  *              *                        **
PR-1a  AUAU----CAAGAACUUAAAUAAUUGCUCUAAAAAGCAACUUAAAGUCAA--G
PR-1b  AUAUAUAUAAACAACUUAAAUAAUUGCACUAAAAAGCAACUUAUAGUCAAAAG
                  600                       625

            ***********              ***   *       *
PR-1a  UAUAUAGUAAUAGUACUAUAUUUGUAAUCCUCU---GAAGUGGAUCUAUAAAA
PR-1b  UAUAUA-----------AUAUUUGUAAUCCUCUGAAGAACUGGAUCUGUAAAA
                          650                       675

         *          *           ****** *      * ******
PR-1a  AGACCAAGUGGUCAUAAUUAAGGGG----AAAAAUAUGAGUUGAUGAUCAGCU
PR-1b  AGUCCAAGUGGUCUUAAUUAAGGGGGGGAGGAUAUAUGAAU------UCAGCU
                        700                         725

                                   *   *
PR-1a  UGAUGUAUGAUCUGAUAUUAUUAUGAACACUUUUGUACUC
PR-1b  UGAUGUAUGAUCUGAUAUUAUUAUGAACUCUUUAGUACUCUU
                          750
```

Fig. 3. Comparison of the non-coding regions of PR-1 mRNA with the 5′- and 3′-non-coding sequences of PR-1c and PR-1a mRNA, respectively. Differences are indicated by asterisks.

ing ~60%. However, there are a number of conserved amino acid substitutions and a few local reversions of the sequence. For instance, the sequence Asn-Gly at position 27–28 in PR-1b is reversed into Ala-Asn in p14. Similar reversions are seen at position 38–39 (Ser-Gln into Asn-Ser) and position 68–69 (Ala-Lys into Arg-Ala). When these are taken into account the homology between PR-1b and p14 is ~75%.

Figure 3 shows a comparison of the leader sequences of the PR-1b and -1c mRNAs, and an alignment of the 3′-non-coding regions of PR-1a and -1b mRNA. The leader sequence is highly conserved, is rich in A and U residues and contains only a few G residues. In the two mRNAs the first AUG from the 5′ end is the beginning of the open reading frame. The 3′-non-coding sequence seems to be more variable. In this region PR-1a and -1b differ by 13 base substitutions and nine deletions/insertions. Polyadenylation signals in plants seem to be somewhat more variable compared with the strict AAUAAA consensus sequence used in animal systems (Messing *et al.*, 1983). However, no clear polyadenylation signal can be recognized in the 3′-non-coding PR-1 mRNA sequences. Nevertheless, the RNAs are polyadenylated as can be concluded from their selective binding to poly(U)-Sepharose, their template activity in oligo(dT)-primed reverse transcription and the presence of short poly(A) sequences in clones 119 and 69.

PR-1 mRNA induction

When a Northern blot of poly(A) RNA from TMV-infected tobacco was probed with clone 110, a single band of PR-1 mRNA was seen in a previous study (Hooft van Huijsduijnen *et al.*, 1985) whereas the signal obtained with RNA from healthy plants was at least 100-fold lower. Here we used clone 110 as a probe to follow the time course of the appearance of PR-1 mRNAs in the lower and upper leaves after the lower leaves had been inoculated with TMV. Lanes A–I of Figure 4 show that, in the lower leaves, the synthesis of PR-1 mRNAs became detectable 2 days after inoculation. In the upper leaves, which contain no detectable amount of virus, the appearance of PR-1 mRNAs occurred much later. Only after 8 days did an increasing amount of PR-1 mRNA become detectable (Figure 4, lanes J–M). Hence, the synthesis and/or transport of the putative 'mobile compound'

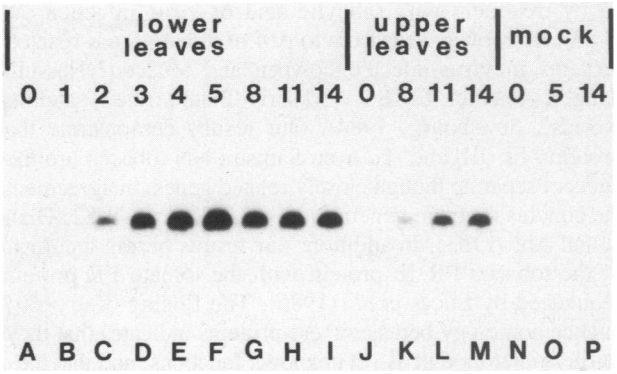

Fig. 4. Induction of PR-1 mRNA in Samsun NN tobacco plants after inoculation of the lower leaves with TMV. At the indicated time after inoculation (0–14 days), RNA was extracted from the lower leaves (**lanes A–I**), the upper leaves (**lanes J–M**) and mock-infected plants (**lanes N–P**), electrophoresed in an agarose gel, blotted onto nitrocellulose and hybridized to ^{32}P-labeled clone 110 DNA.

Table I. Amino acid composition of tobacco PR-1b and tomato p14

	PR-1b	p14
Ala	14	14
Val	16	9
Leu	5	6
Ile	1	3
Pro	4	5
Phe	1	2
Trp	3	5
Tyr	10	7
Met	2	1
Gly	13	16
Ser	8	8
Thr	5	3
Cys	6	6
Asn	12	13
Gln	12	8
Asp	10	6
Glu	4	2
Lys	3	3
Arg	4	10
His	5	3
	138	130

Amino acid composition of p14 is taken from Lucas *et al.* (1985).

that acts as the transmitting signal for the induction of PR proteins (Van Loon, 1985) is a rather slow process. The exposure time used to produce the autoradiogram of Figure 4 does not reveal the low amounts of PR-1 mRNAs that are known to be present (Hooft van Huijsduijnen *et al.*, 1985) in non-inoculated (Figure 4, lanes A and J) or mock-inoculated (lanes N–P) tobacco plants.

Discussion

Upon exposure to a variety of pathogens or chemicals, numerous plant species are now known to produce a whole series of PR proteins, which can accumulate to levels of up to 10% of the soluble leaf protein (Van Loon, 1985). There is growing evidence that many of these proteins are structurally related. An antiserum to PR-1a of Xanthi nc tobacco was found to react specifically with proteins in cowpea, potato and *Gomphrena globosa* induc-

ed by treatment with salicylic acid or virus infection (White, 1983). Moreover, antiserum to p14 of tomato cross-reacted with proteins in virus-infected cowpea and tobacco (Nassuth and Sänger, Abstract, EMBO workshop 'Plant viruses, satellites and viroids', Strasbourg, 1984). Our results demonstrate that PR proteins-1a, -1b and -1c from Samsun NN tobacco are the products of separate though closely related genes, in agreement with the conclusion from genetic studies (Ahl *et al.*, 1982; Gianinazzi and Ahl, 1983). In addition, our results permit a comparison of the tobacco PR-1b protein with the tomato PR protein p14 sequenced by Lucas *et al.* (1985). The finding of a >60% sequence homology between these proteins indicates that they may fulfill similar though as yet unknown functions, and that the genes for these proteins are well conserved in the plant kingdom. To facilitate a comparison, the amino acid compositions of PR-1b and p14 are listed in Table I. Both proteins contain a rather high percentage (~10%) of aromatic amino acids. Although the total number of charged residues is the same, a major difference between the two proteins is the relative proportion of basic and acidic amino acids. In PR-1b there is an excess of acidic residues whereas in p14 there is a considerable excess of basic residues, which is in agreement with their respective isoelectric points of 4.0 (Pierpoint, 1983) and 10.7 (Camacho Henriquez and Sänger, 1984). It seems that the isoelectric point of these proteins is not critical to their function and does not explain their selective solubility at low pH (Van Loon, 1985). Moreover, knowledge of their primary sequence does not help explain the protease resistance of these proteins (Van Loon, 1985) as potential cleavage sites for trypsin- or chymotrypsin-like enzymes are abundantly present. Possibly, these cleavage sites are not accessible due to the compact structure of the proteins.

In addition to the PR-1 proteins, TMV infection in tobacco induces the synthesis of at least seven other components of higher mol. wt. which, like the PR-1 proteins, are pH 3 soluble, protease resistant and located in the extracellular space of the infected plant tissue. Our PR-1 clones do not hybridize to the mRNAs for these proteins (Hooft van Huijsduijnen *et al.*, 1985). This indicates that under stress conditions different classes of PR genes are induced.

Insight into the primary sequence of PR proteins and the structural relationship between the sets of PR proteins induced in different plant species is a necessary step towards an ultimate understanding of their enigmatic role. The availability of a complete PR-1 cDNA clone makes it possible to study the putative anti-viral activity of these proteins. Plants can be transformed with the clones linked to an efficient promoter, in order to construct healthy tobacco plants producing high levels of PR proteins. The response to virus infection of such plants can be tested. Moreover, the clones described in this paper can be used to characterize the virus-inducible promoter of PR genes. Such a promoter may be useful to express foreign genes in plant cells upon virus infection or treatment with chemical inducers, e.g., acetylsalicylic acid (White, 1979).

Materials and methods

Cloning and sequencing of PR-cDNA

Poly(A) RNA was isolated from the leaves of 9–10 week old tobacco plants (*Nicotiana tabacum* cv. Samsun NN), 7 days after inoculation with TMV (strain WU1) as described by Hooft van Huijsduijnen *et al.* (1985). This RNA was copied into DNA by the method of Gubler and Hoffman (1983): the first strand was synthesized with reverse transcriptase using oligo(dT) as a primer, the second strand was synthesized with a combination of the enzymes Pol I, RNase H and ligase. Double-stranded cDNA was tailed with dCTP and annealed to *Pst*I-cut, G-tailed pUC9 as described (Cornelissen *et al.*, 1983). Tailed cDNA (25 ng) was cloned to yield a library of ~15 000 recombinants. The library was screened with the 5′ ^{32}P-labeled deoxyoligonucleotide 5′ GAGAGTTTTGGGCACG 3′, complementary to nucleotides 129–114 of PR-1 mRNA (Hooft van Huijsduijnen *et al.*, 1985). Re-screening of the library with the positively responding clone 110 yielded clones 119 and 69. cDNA inserts were subcloned into M13 and sequenced by the dideoxy chain terminating method (Sanger *et al.*, 1977) using [α-^{35}S]dATP (Biggin *et al.*, 1983).

Assay of PR-1 mRNA

Total RNA was isolated from tobacco leaves (Hooft van Huijsduijnen *et al.*, 1985), electrophoresed in agarose gels (70 μg per slot), blotted onto nitrocellulose filters and hybridized to ^{32}P-labeled clone 110 DNA (Sarachu *et al.*, 1985).

Acknowledgements

We thank Drs. J.Lucas, A.Henschen and H.L.Sänger (Max Planck Institute of Biochemistry, 8033 Martinsfied bei München, FRG) for sending us their manuscript prior to publication and Dr. John F.Antoniw (Rothamsted Experimental Station, Harpenden, UK) for fruitful discussions. This work was sponsored in part by the Netherlands Foundation for Chemical Research (S.O.N.) with financial aid from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

References

Ahl,P., Cornu,A. and Gianinazzi,S. (1982) *Phytopathology*, **72**, 80-85.
Antoniw,J.F. and Pierpoint,W.S. (1978) *J. Gen. Virol.*, **39**, 343-350.
Antoniw,J.F., Ritter,C.E., Pierpoint,W.S. and Van Loon,L.C. (1980) *J. Gen. Virol.*, **47**, 79-87.
Antoniw,J.F., White,R.F., Barbara,D.J., Jones,P. and Longley,A. (1985) *Plant Mol. Biol.*, **4**, 55-60.
Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963-3965.
Camacho Henriquez,A. and Sänger,H.L. (1984) *Arch. Virol.*, **81**, 263-284.
Cornelissen,B.J.C., Brederode,F.Th., Moormann,R.J.M. and Bol,J.F. (1983) *Nucleic Acids Res.*, **11**, 1253-1265.
Gianinazzi,S. and Ahl,P. (1983) *Neth. J. Plant Pathol.*, **89**, 275-281.
Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 263-269.
Hooft van Huijsduijnen,R.A.M., Cornelissen,B.J.C., Van Loon,L.C., Van Boom,J.H., Tromp,M. and Bol,J.F. (1985) *EMBO J.*, **4**, 2167-2171.
Kassanis,B., Gianinazzi,S. and White,R.F. (1974) *J. Gen. Virol.*, **23**, 11-16.
Lucas,J., Camacho Henriquez,A., Lottspeich,F., Henschen,A. and Sänger,H.L. (1985) *EMBO J.*, **4**, in press.
Matsuoka,M. and Ohashi,Y. (1984) *J. Gen. Virol.*, **65**, 2209-2215.
Messing,J., Geraghty,D., Heidecker,G., Hu,N.-T., Kridl,J. and Rubinstein,I. (1983) in Kusuge,T., Meredith,C.P. and Hollaender,A. (eds.), *Genetic Engineering in Plants: An Agricultural Perspective*, Plenum Press, NY, pp. 211-227.
Pierpoint,W.S. (1983) *Phytochemistry*, **22**, 2691-2697.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Sarachu,A.N., Huisman,M.J., Van Vloten-Doting,L. and Bol,J.F. (1985) *Virology*, **141**, 14-22.
Van Loon,L.C. (1975) *Virology*, **67**, 566-575.
Van Loon,L.C. (1985) *Plant Mol. Biol.*, **4**, 111-116.
White,R.F. (1979) *Virology*, **99**, 410-412.
White,R.F. (1983) *Neth. J. Plant Pathol.*, **89**, 305-323.

Received on 9 September 1985; revised on 4 November 1985