Volume 15 Number 11 1987 Nucleic Acids Research

Isolation and characterization of cDNA clones encoding pathogenesis-related proteins from tobacco mosaic virus infected tobacco plants

Ursula M.Pfitzner and Howard M.Goodman

Department of Molecular Biology, Massachusetts General Hospital and Department of Genetics, Harvard Medical School, Massachusetts General Hospital, Boston, MA 02114, USA

Received February 18, 1987; Revised and Accepted March 30, 1987 Accession no. Y00335

ABSTRACT

Infection of the tobacco cultivar Samsun NN by tobacco mosaic virus (TMV) results in a hypersensitive response. During this defense reaction several host encoded proteins, known as pathogenesis-related proteins (PR-proteins), are induced. Poly(A)$^+$ RNA from TMV infected tobacco plants was used to construct a cDNA library. Thirty two cDNA clones were isolated and after digestion with different restriction endonucleases, twenty clones were found to code for PR-1a, six clones for PR-1b, and four clones for PR-1c. Two independent cDNA clones of each class were further characterized by DNA sequence analysis. All clones analyzed contained the 138 amino acid coding regions of their respective mature proteins, but only partial sequences of the signal peptides. Minor differences between the nucleotide sequences for clones belonging to the same class were detected. Comparison of the amino acid sequence for PR-1a deduced from its nucleotide sequence with published data obtained by Edman degradation of the protein showed four differences. Analysis of the 3' ends of the cDNA clones indicates that various alternate poly(dA) addition sites are used. Southern blot analysis using these cDNAs as probes suggests the presence of multiple PR-protein genes in the genomes of tobacco and tomato plants.

INTRODUCTION

Infection of tobacco plants with tobacco mosaic virus (TMV) results in two possible distinct host responses depending on the genetic constitution of the plant cultivar and the virus strain. In an incompatible reaction, various biochemical changes are observed in the plants concomitant with the expression of disease resistance. These include increases in the activities of the enzymes involved in the phenylpropanoid pathway and the *de novo* synthesis of pathogenesis-related proteins (PR-proteins; for review see 1).

PR-proteins have several features in common. They are soluble in acidic buffers, have low molecular weights ranging from 10 kD - 20 kD, occur in the intercellular spaces of the leaves, and are

resistant to proteases (for review see 2). In TMV infected tobacco plants, ten different PR-proteins have been described. The best characterized of these is the PR-1 group which consists of three members. Biochemical, serological, and genetic evidence suggested that these proteins, although closely related, are encoded by separate genes (3,4,5,6,7). Recently, partial amino acid sequences of PR-1a were reported (8) and a full-length cDNA clone for PR-1b and partial clones for PR-1a and PR-1c obtained (9).

Here, we report the cloning of cDNAs encoding PR-proteins from TMV infected tobacco leaves cv. Samsun NN, describe nucleotide sequences in the coding regions of PR-1a and PR-1c and compare the deduced amino acid sequences over the full length of the mature proteins. By analyzing 32 cDNA clones, we have found that PR-protein mRNAs accumulate to different levels and utilize different polyadenylation sites. Furthermore, investigation of the organization of PR-protein genes in the genomes of tobacco and tomato suggests the presence of more than three PR-1 related genes in tobacco plants.

MATERIALS AND METHODS

Plant material

Plants (*Nicotiana tabacum* cv. Samsun NN and *Lycopersicon esculentum*) were grown under normal greenhouse conditions at 20°C. Two to three months old tobacco plants were infected by rubbing a suspension of purified TMV (10; common strain, 30 µg/ml) onto the leaves and the plants transferred to a growth chamber. Control plants were obtained by mock inoculation of leaves with 0.1 M sodium phosphate buffer, pH 7. Leaves were harvested 7 days after the inoculation procedure, frozen in liquid nitrogen, and stored at -70°C.

RNA and DNA isolation

Total cellular RNA was isolated from the frozen leaves by the guanidinium/cesium chloride method (11) except that the extraction buffer was 4 M guanidinium isothiocyanate, 50 mM Tris-HCl pH 7.6, 10 mM EDTA, 2% sodium lauryl sarkosinate, and 0.1% β-mercaptoethanol. Poly(A)$^+$ RNA was selected by oligo(dT)-cellulose chromatography (12). Total DNA was extracted from three months old tobacco plants and six months old tomato plants according to

the method in reference (13). Plasmid DNA was isolated by the alkaline lysis method (14).

Construction and screening of the cDNA library

Double stranded cDNA was synthesized from 5 μg poly(A)$^+$ RNA isolated from TMV infected Samsun NN leaves (15). After homopolymeric tailing with dC-residues, the cDNA was annealed to PstI cut, dG-tailed pUC12 and transformed into *E.coli* strain DK1. Plating and screening of the transformants was as described (16). A 30 nucleotide long synthetic oligodeoxynucleotide, 5'-GTCTTGTTGAGAGTTTTGGGCACGACAAGA-3', complementary to PR-protein mRNA (17) was end-labeled with [γ-^{32}P]ATP by T4 polynucleotide kinase (specific activity 8x10^{8}cpm/μg) and used as probe. Hybridizations were performed at 42^{o}C in 6xSSC (1xSSC = 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0), 10xDenhardt's solution (1xDenhardt's = 0.2 g/l each of Ficoll, polyvinylpyrrolidone, and bovine serum albumin), 50mM HEPES pH 7.0, and 162.5 μg/ml denatured salmon sperm DNA. Filters were washed at room temperature, at 42^{o}C, and at 60^{o}C in 6xSSC for 3 x 15 min each.

DNA sequence analysis

Nucleotide sequences were determined using the dideoxynucleotide chain termination method (18) with [α-^{35}S]dATP (19) and the universal external M13 primer. Restriction enzyme fragments were subcloned into M13mp18 and mp19 vectors. In order to sequence the 3' terminus of pNt^{SNN}cPR1a/8 on both DNA strands, a synthetic 15 base long oligodeoxynucleotide (5'-TCAAGTAGCTAGACC-3') complementary to the very end of the clone was used as primer.

Genomic Southern blots

Ten μg of plant DNA were digested to completion with BamHI or HindIII and separated by electrophoresis on 0.8% agarose gels. The DNA was transferred onto nitrocellulose filters (11) and hybridized with nick translated pNt^{SNN}cPR1a/35 probe (specific activity 5x10^{8}cpm/μg). In order to evaluate nonspecific hybridization with vector sequences, parallel blots were hybridized with nick translated pUC12 probe. Hybridizations were performed according to the method in reference (11). All filters were washed for 15 min at room temperature in 2xSSC, 0.5% SDS. Tobacco DNA blots were further washed for 1 hr at 68^{o}C in 0.3xSSC, 0.5% SDS. Filters with tomato DNA were washed for 1 hr

at 68°C in either 1xSSC, 0.5% SDS or 0.5xSSC, 0.5% SDS. Copy number calculations are based on a haploid genome size of 3.9 pg for *Nicotiana tabacum*.

Northern blots

Thirty μg of total cellular RNA were denatured in glyoxal (20), and separated by electrophoresis on 1.5% agarose gels. The RNAs were transferred onto nitrocellulose filters, probed with the 30 base synthetic oligodeoxynucleotide, hybridized, and washed using the conditions described in "Construction and screening of the cDNA library" above. The positions in the gel of the 28S and 18S rRNA bands were used as standards for size estimation of the hybridizing band.

Antibodies directed against PR-proteins

PR-proteins were extracted from frozen leaves with two volumes of citrate phosphate buffer (32 mM Na_2HPO_4, 84 mM citric acid, pH 2.8) containing 0.1% β-mercaptoethanol. The crude protein extract was concentrated by the addition of ammonium sulfate (95% final concentration) and separated by chromatofocusing. PR-proteins were eluted from the column at pH 3.9 (PR-1b) and pH 3.5 (PR-1a). These fractions were further purified by electrophoresis on 11.25% native polyacrylamide gels prepared as described in reference (21). For immunizations, two New Zealand White Rabbits were each injected with an emulsion of crushed polyacrylamide gel slices containing 30 μg PR-1a or PR-1b in complete Freund's adjuvant. Booster injections were administered 2 weeks after the initial injections, and the rabbits were bled 1 week and 2 weeks later. The antisera were used without further purification. Complete immunological identity of PR-1a and PR-1b was shown by the Ouchterlony double diffusion method.

Immunoblots

Crude protein extracts were concentrated by the addition of acetone (80% final concentration). Precipitated proteins were redissolved in SDS gel loading buffer, subjected to electrophoresis on 15% polyacrylamide gels (22), and blotted onto nitrocellulose membranes (23). The filters were incubated in a 1:5 dilution of the specific antiserum directed against PR-1a. Immunodetection was achieved by Protein-A coupled horseradish peroxidase according to the instructions of the manufacturer (Bio-Rad). The positions in the gel of protein standards

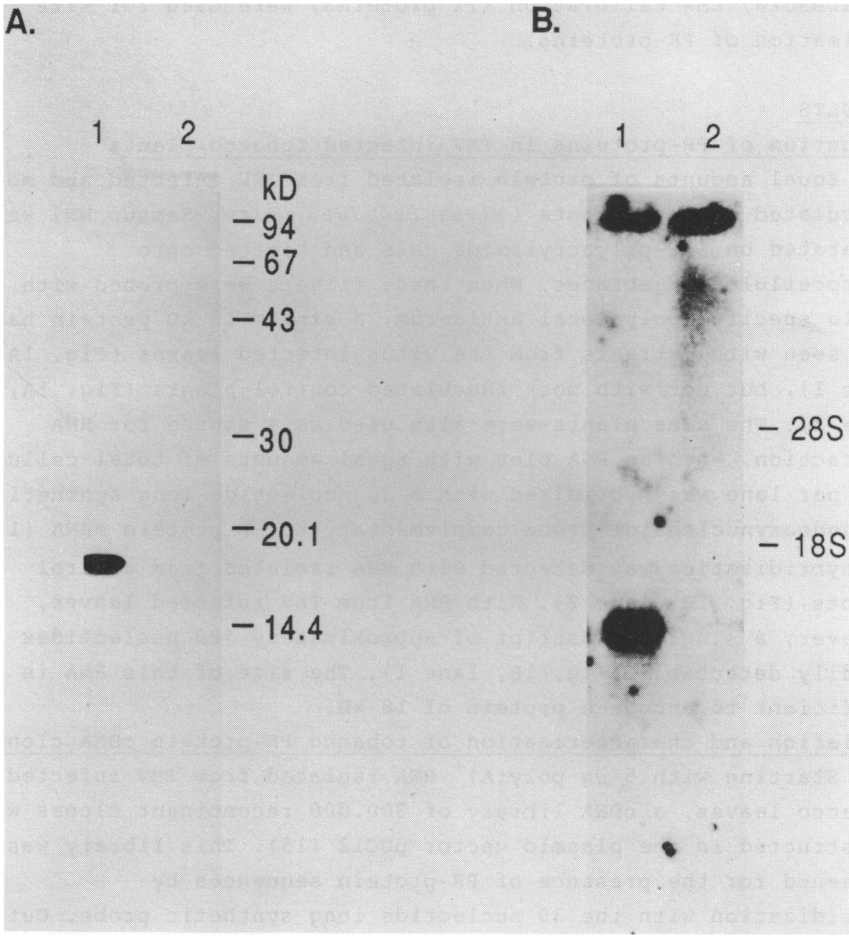

Fig. 1
Analysis of protein (A.) and RNA (B.) levels of PR-proteins in TMV infected and mock inoculated *Nicotiana tabacum* cv. Samsun NN leaves.

A: Samples of 40 μg of protein isolated from TMV infected (lane 1) or mock inoculated (lane 2) leaves were analyzed for PR-protein content by SDS polyacrylamide gel electrophoresis followed by immunoblotting. The positions in the gel of a standard set of protein size markers are shown.

B: Samples of 30 μg of total cellular RNA isolated from TMV infected (lane 1) or mock inoculated (lane 2) leaves were denatured in glyoxal, separated on a 1.5% agarose gel, blotted onto a nitrocellulose filter, and hybridized with an end-labeled 30 base long oligodeoxynucleotide complementary to PR-protein mRNA. The positions in the gel of the 28S and 18S rRNA bands were used as standards for size estimation of the hybridizing band. The nature of the counts at the origin of the gel is unknown.

(Pharmacia, LMW calibration kit proteins) were used for size estimation of PR-proteins.

RESULTS

Induction of PR-proteins in TMV infected tobacco plants

Equal amounts of protein isolated from TMV infected and mock inoculated tobacco plants (*Nicotiana tabacum* cv. Samsun NN) were separated on SDS-polyacrylamide gels and blotted onto nitrocellulose membranes. When these filters were probed with a PR-1a specific polyclonal antiserum, a single 18 kD protein band was seen with extracts from the virus infected leaves (Fig. 1A, lane 1), but not with mock inoculated control plants (Fig. 1A, lane 2). The same plants were also used as a source for RNA extraction. When an RNA blot with equal amounts of total cellular RNA per lane was hybridized with a 30 nucleotide long synthetic oligodeoxynucleotide probe complementary to PR-protein mRNA (17), no hybridization was detected with RNA isolated from control plants (Fig. 1B, lane 2). With RNA from TMV infected leaves, however, a single transcript of approximately 700 nucleotides is readily detectable (Fig. 1B, lane 1). The size of this RNA is sufficient to encode a protein of 18 kD.

Isolation and characterization of tobacco PR-protein cDNA clones

Starting with 5 μg poly(A)$^+$ RNA isolated from TMV infected tobacco leaves, a cDNA library of 300,000 recombinant clones was constructed in the plasmid vector pUC12 (15). This library was screened for the presence of PR-protein sequences by hybridization with the 30 nucleotide long synthetic probe. Out of 15,000 transformants, a total of 176 PR-protein specific clones were identified. This corresponds to a level of PR-protein mRNAs of ~1%, assuming an equal conversion of each mRNA species into cDNA. Subsequently, 32 clones were purified and characterized. Thirty clones varied in insert size from 700 bp to 800 bp suggesting that they might represent nearly complete copies of their respective mRNAs, a view supported by the fact that the synthetic probe is complementary to sequences encoding amino acids at the N-terminus of PR-proteins.

By digestion with different restriction endonucleases, three classes of PR-protein cDNA clones were identified. All clones share a conserved PstI site in the protein coding regions (Fig.

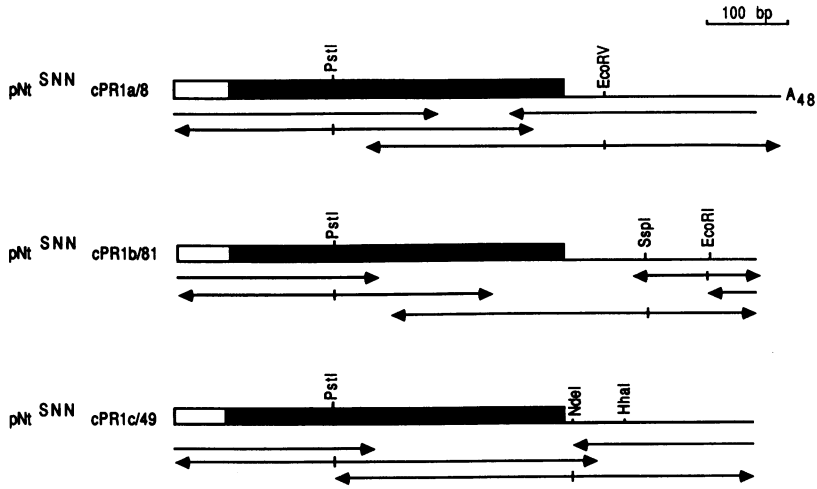

Fig. 2
Restriction maps of PR-protein cDNA clones and DNA sequencing strategies. The restriction maps show only those restriction endonuclease sites relevant in distinguishing the three classes of PR-protein cDNA clones and the conserved PstI site in the protein coding regions. Thick black bars, protein coding regions of mature proteins; thick open bars, protein coding regions of signal peptides; thin bars, 3' untranslated regions. Horizontal arrows under each clone indicate direction and extent of sequence determinations.

2), but differ in the unique restriction sites in their 3' untranslated regions. Class I clones (20 members) are characterized by an EcoRV site, class II clones (6 members) by SspI and EcoRI sites, and class III clones (4 members) by NdeI and HhaI sites (Fig. 2). Members of each group, the two longest clones of their class, were further characterized by DNA sequence analysis.

Only minor differences in the DNA sequences were found between cDNA clones of the same class (Fig. 3). The clones are therefore derived from the same gene or from allelic genes. The clone with the longest open reading frame (499 nucleotides) was pNt^{SNN}cPR1a/35. Two clones, pNt^{SNN}cPR1a/8 and pNt^{SNN}cPR1c/83, had long poly(dA) tails with 48 and 35 residues, respectively, while another clone, pNt^{SNN}cPR1a/35, had only seven dA-residues. Assuming these represent the *in vivo* polyadenylation sites, the

pNt^{SNN}cPR1a/8

```
  1  ATTTGTTCTC TTTTCACAAT TGCCTTCATT TCTTCTTGTC TCTACACTTC TCTTATTCCT
 61  AGTAATATCC CACTCTTGCC GTGCCCAAAA TTCTCAACAA GACTATTTGG ATGCCCATAA
121  CACAGCTCGT GCAGATGTAG GTGTAGAACC TTTGACCTGG GACGACCAGG TAGCAGCCTA
181  TGCGCAAAAT TATGCTTCCC AATTGGCTGC AGATTGTAAC CTCGTACATT CTCATGGTCA
241  ATACGGCGAA AACCTAGCTG AGGGAAGTGG CGATTTCATG ACGGCTGCTA AGGCCGTTGA
301  GATGTGGGTC GATGAGAAAC AGTATTATGA CCATGACTCA AATACTTGTG CACAAGGACA
361  GGTGTGTGGA CACTATACTC AGGTGGTTTG GCGTAACTCG GTTCGTGTTG GATGTGCTAG
421  AGGTTCAGTGT AACAATGGAG GATATGTTGT CTCTTGCAAC TATGATCCTC CAGGTAATTA
481  TAGAGGCGAA AGTCCATACT AAttgaaacg acctacgtcc atttcacgtt aatatgtatg
541  gattgttctg cttgatatca agaacttaaa taattgctct aaaaagcaac ttaaagtcaa
601  gtatatagta atagtactat atttgtaatc ctctgaagtg gatctataaa aagaccaagt
661  ggtcataatt aaggggaaaa tatgagttga tgatcagctt gatgtatgat ctgatattat
721  tatgaacact tttgtactca tacgaatcat gtgttgatgg tctagctact tg-dA48
```

pNt^{SNN}cPR1b/81

```
  1  TTCTCTTTTC ACAAATGCCC TCATTTTTTC TTGTCTCTAC ACTTCTCTTA TTCCTAATAA
 61  TATCTCACTC TTCTCATGCC CAAAACTCTC AACAAGACTA TTTGGATGCC CATAACACAG
121  CTCGTGCAGA TGTAGGCGTG GAACCATTAA CTTGGGACAA CGGGGTAGCA GCCTATGCAC
181  AAAATTATGT TTCTCAATTG GCTGCAGACT GCAACCTCGT ACATTCTCAT GGCCAATACG
241  GCGAAAACCT AGCTCAGGGA AGTGGCGATT TTATGACGGC TGCTAAGGCC GTCGAGATGT
301  GGGTCGATGA GAAACAGTAC TATGACCATG ACTCAAATAC TTGTGCACAA GGACAGGTGT
361  GTGGACACTA TACTCAGGTG GTTTGGCGTA ACTCGGTTCG TGTTGGATGT GCTAGGGTTA
421  AGTGCAACAA TGGAGGATAT GTTGTCTCTT GCAACTATGA TCCTCCAGGT AATGTCATAG
481  GCCAAAGTCC ATACTAAttg aaatgaatgt ccatttcacg ttatatatgt atggacttct
541  gcttgatata tataaacaac ttaaataatt gcactaaaaa gcaacttata gttaaaagta
601  tataatattt gtaatcctct gaagaactgg atctgtaaaa agtccaagtg gtcttaatta
661  agggggggag gatatatgaa ttcagcttga tgtatgatct gatattatta tgaactcttt
721  agtactctta cggaa
```

pNt^{SNN}cPR1c/49

```
  1  CACAAATGTC TTCATTTTTT CTTGTCTCTA CGCTTCTCTT ATTCCTAATA ATATCCCACT
 61  CTTGTCATGC TCAAAACTCT CAACAAGACT ATTTGGATGC CCATAACACA GCTCGTGCAG
121  ATGTAGGTGT AGAACCTTTG ACCTGGGACG ACCAGGTAGC AGCCTATGCA CAAAATTATG
181  CTTCCCAATT GGCTGCAGAT TGTAACCTCG TACATTCTCA TGGTCAATAC GGCGAAAACC
241  TAGCTTGGGG AAGTGGCGAT TTCTTGACGG CCGCTAAGGC CGTCGAGATG TGGGTCAATG
301  AGAAACAGTA TTATGCCCAC GACTCAAACA CTTGTGCCCA AGGACAGGTG TGTGGACACT
361  ATACTCAGGT GGTTTGGCGT AACTCGGTTC GTGTTGGATG TGCTAGGGTT CAGTGTAACA
421  ATGGAGGATA TATTGTCTCT TGCAACTATG ATCCTCCAGG TAATGTTATA GGCAAAAGCC
481  CATACTAAtt gaaaacatat gtccatttca cgttatatat gtgtggactt ctgcttgata
541  tatatcaaga acttaaataa ttgcgctaaa aagcaactta tagttaagta tatagtacta
601  tatttgtaat tctctgaagt ggatatataa taagacctag tgctcttgat tatggggaaa
661  aaatatgaat tccgcttgat gtatgatatg atattattat gaaatcttta gtacttttac
721  g dA29
```

PR-protein cDNA clones differ in the lengths of their 3' untranslated regions; 194 nucleotides for pNt^{SNN}cPR1a/35 and 270 nucleotides for pNt^{SNN}cPR1a/8 (both class I) and 169 nucleotides for pNt^{SNN}cPR1c/83 and >233 nucleotides for pNt^{SNN}cPR1c/49 (both class III). The two class II clones analyzed (pNt^{SNN}cPR1b/1 & /81) terminated prior to the poly(dA) tail. The PR-protein genes therefore use alternative polyadenylation sites, a finding in agreement with observations made for other plant genes (24). Furthermore, the highly conserved polyadenylation signal AATAAA, which is located 10-30 bp upstream of the poly(dA) addition site in most mammalian genes (25), is not present in the PR-protein

Fig. 3
Nucleotide sequences of six independent PR-protein cDNA clones. The strands homologous to the respective mRNAs are shown. Coding sequences are in upper case letters and 3' untranslated regions in lower case letters. Only the sequence of the longest clone of each class is given in its entirety. Nucleotides are numbered in the 5' to 3' direction starting with the first nucleotide of the clone with the longest 5' end. Differences in sequence of the shorter clone of each class are shown in bold. Putative polyadenylation signals are underlined. PR1a: The sequence of pNt^{SNN}cPR1a/8 is shown starting at nucleotide (nt) #14. The 30 nucleotide long probe is complementary to (nt) #74 to 103 with two mismatches at nt #79 and 91. The sequence of the other PR-1a clone, pNt^{SNN}cPR1a/35 (shown in bold on the line under pNt^{SNN}cPR1a/8), was identical to that shown except that its 5' terminus was at position #1 (i.e. it has 13 additional nucleotides at its 5' end, nt #1-13 shown in bold), it has an A (rather than a G) at nt #361, it has an A residue added at nt #680 (shown by a ↑+a), and has a 77 nt shorter 3' untranslated region. The poly(dA) is attached to a C at nt #695 in pNt^{SNN}cPR1a/35 (seven A's; dA_7) and to a G at nt #772 in pNt^{SNN}cPR1a/8 (48 A's; dA_{48}). PR1b: The sequence of pNt^{SNN}cPR1b/81 is shown. The 30 nucleotide long probe is complementary to nucleotides (nt) #69 to 98 with two mismatches at nt #73 and 76. The sequence of the other PR-1b clone, pNt^{SNN}cPR1b/1, was identical to that shown except that its 5' terminus was at position #1 (i.e. it has 13 additional nucleotides at its 5' end, nt #1-13 shown in bold) and it has a 50 nt shorter 3' untranslated region. Neither clone has a poly(dA) tract. PR1c: The sequence of pNt^{SNN}cPR1c/49 is shown. The 30 nucleotide long probe is complementary to nucleotides (nt) #60 to 89 with two mismatches at nt #67 and 71. The sequence of the other PR-1c clone, pNt^{SNN}cPR1c/83, was identical to that shown except that it has a 58 nt shorter 3' untranslated region. The poly(dA) is attached to an A at nt #663 in pNt^{SNN}cPR1c/83 (35 A's; dA_{29}) and no poly(dA) tract is apparent in pNt^{SNN}cPR1c/49. The sequence of both strands was determined for the entire clone of pNt^{SNN}cPR1c/49 (class III), greater than 60% for pNt^{SNN}cPR1b/81 (class II), and greater than 80% for pNt^{SNN}cPR1a/8 (class I). For the three other clones the sequences were only determined on one strand.

```
                                                    -1
PR-1a     -28  FVLFSQLPSFLLVSTLLLFLVISHSCRAQNSQQDYLDAHNTARADVGVEPLTWDDQVAAY
                                    *     *
PR-1b     -26    ↑   MP  F          I    SH                              NG
PR-1c     -23        ↑MS  F          I    CH                              DQ

PR-1a      33  AQNYASQLAADCNLVHSHGQYGENLAEGSGDFMTAAKAVEMWVDEKQYYDHDSNTCAQGQ
                *
PR-1b      33      V                        Q      M           D      D
                *
PR-1c      33      A                        W      L           N      A

PR-1a      93  VCGHYTQVVWRNSVRVGCARVQCNNGGYVVSCNYDPPGNYRGESPY
PR-1b      93                           K       V           VI Q
                                                *
PR-1c      93                           Q       I           VI K
```

Fig. 4
Comparison of the amino acid sequences of PR-proteins deduced from the respective cDNA clones. The PR-1a protein sequence as deduced from the nucleotide sequence of pNt^{SNN}cPR1a/35 is shown. Only amino acids which are different from PR-1a are shown for PR-1b and PR-1c as deduced from the nucleotide sequences of pNt^{SNN}cPR1b/1 and pNt^{SNN}cPR1c/83, respectively. Identical amino acids are indicated by blanks, differences by the changed amino acid, and conservative changes (V→I, R→H, and A→V) are indicated by (*). Residues are numbered beginning with 1 for the first amino acid of the mature proteins (Q, shown in bold). Negative numbers indicate amino acids of the signal peptides. The processing site is between the A at -1 and the Q at +1. Every tenth residue in each direction has a dot over it. Empty spaces at the N-terminus of PR-1b and PR-1c to the left of the arrows denote lack of sequence information due to shorter cDNA clones, i.e. PR-1b and PR-1c terminate at the arrows at -26 and -23, respectively. The single letter amino acid code is used.

cDNA clones. However, the related sequences ATAATT and AAAAAT (pNt^{SNN}cPR1a/35), AATCAT (pNt^{SNN}cPR1a/8), and ATAATA (pNt^{SNN}cPR1c/83) occur at the correct spacing with respect to the positions of the poly(dA) tracts and therefore might function as polyadenylation signals (24). The overall homology of PR-protein cDNA clones at the nucleotide level is 90% in the protein coding regions and 80% in the 3′ untranslated regions.

Amino acid sequences of tobacco PR-proteins

Comparison of the amino acid sequences of the PR-proteins as deduced from the cDNA clones indicates that the sequences are identical for the same clone class and very similar between classes (Fig. 4). The assignment of the cDNA clones to the respective proteins was based on comparison with the amino acid sequence of protein PR-1a (8) and nucleotide sequences of PR-1a, PR-1b, and PR-1c (9). Class I clones correspond to PR-1a and

represent the mRNA expressed at the highest level, class II clones correspond to PR-1b, and class III cDNA clones, which are found at the lowest level, correspond to PR-1c. The abundance level of the mRNA was assumed to be directly related to the number of cDNA clones isolated of each type. Four amino acid differences are observed, when the predicted amino acid sequence for PR-1a is compared to data obtained by Edman degradation of purified PR-1a (8). These are Asp→Ser at position 27, Ser→Pro at position 38, Gln→Ser at position 39, and Tyr→Trp at position 132.

The longest amino acid sequence as deduced from the cDNA clones (166 residues) was obtained for PR-1a, however, it is still missing the initiation methionine at position -30 (known to occur at this position from sequence data on a λ clone from a genomic tobacco library with essentially an identical sequence with this cDNA (A.J.P.Pfitzner, U.M.Pfitzner, and H.M.Goodman, unpublished results). For PR-1b and PR-1c, amino acid sequences of 164 and 161 residues, respectively, were obtained. Both of these sequences contain a methionine in the correct reading frame at position -22. However, it is not likely to be the start of translation since another in-frame methionine located eight residues upstream has been previously described for PR-1b and PR-1c (9) and the sequences surrounding Met -22 in PR-1b and PR-1c do not agree with the favoured nucleotide sequences flanking functional initiation codons of eukaryotic mRNAs (26). Clone $pNt^{SNN}cPR1a/35$ (PR-1a) is therefore presumably missing five nucleotides, $pNt^{SNN}cPR1b/1$ (PR-1b) ten nucleotides, and $pNt^{SNN}cPR1c/49$ (PR-1c) nineteen nucleotides of their signal peptide coding regions.

PR-proteins are synthesized as precursor molecules with N-terminal extensions (17,27). The cleavage site (indicated in Fig. 4) of the signal peptides is predicted from the amino acid sequence established for the mature protein PR-1a (8) and the pre-protein sequence as deduced from the nucleotide sequence. PR-protein signal peptides share common features with other eukaryotic signal sequences (28). The mature proteins each contain 138 amino acids. This corresponds to a calculated molecular weight of 15,225 for PR-1a, 15,075 for PR-1b, and 15,129 for PR-1c and is in reasonable agreement with the M_r of 18 kD estimated by SDS polyacrylamide gel electrophoresis. The homology of PR-proteins at the amino acid level is 90% since some

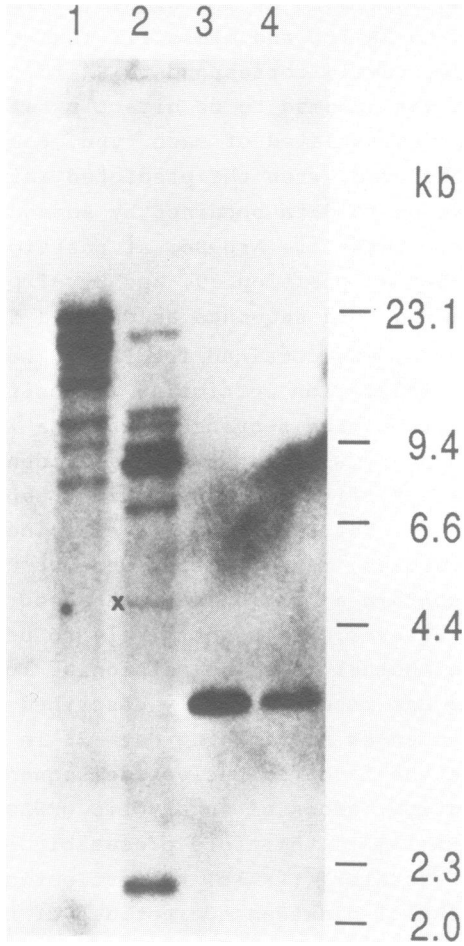

Fig. 5
Southern blot analysis of PR-protein genes in *Nicotiana tabacum* cv. Samsun NN. Ten μg of total DNA were digested to completion with endonucleases BamHI (lane 1) or HindIII (lane 2). The digests were separated on a 0.8% agarose gel, and blotted onto a nitrocellulose filter. The filter was hybridized with nick translated pNt^{SNN}cPR1a/35 probe. Hybridization due to nonspecific binding to vector sequences alone is marked by (x). Lanes 3 and 4 show hybridization signals from linearized pNt^{SNN}cPR1a/35 corresponding to 10 gene copies and 1 gene copy, respectively. The positions in the gel of a HindIII digest of wild type bacteriophage λ DNA were used as standards for size estimation of the hybridizing bands.

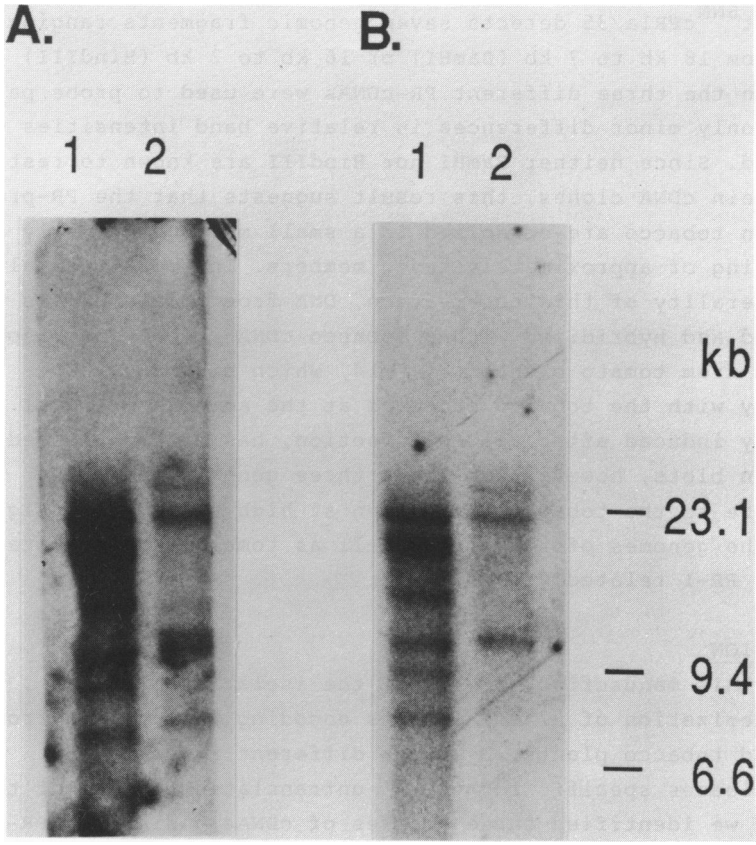

Fig. 6
Southern blot analysis of PR-protein genes in *Lycopersicon esculentum*. Ten µg of total DNA isolated from tobacco plants (lane 1) or tomato plants (lane 2) were digested to completion with endonuclease BamHI. The DNA transfer blots were washed either in 1xSSC, 0.5% SDS (panel A.) or in 0.5xSSC, 0.5% SDS (panel B.).

nucleotide changes are silent. Variations in the amino acid sequences are predominantly found in the signal peptide and at the C-terminus.

Identification of PR-protein genes in tobacco and tomato plants

The organization of PR-protein genes was investigated by Southern blot analysis. Total DNA from *Nicotiana tabacum* cv. Samsun NN digested with endonucleases BamHI or HindIII and probed

with pNt^{SNN}cPR1a/35 detects seven genomic fragments ranging in size from 18 kb to 7 kb (BamHI) or 16 kb to 2 kb (HindIII) (Fig. 5). When the three different PR-cDNAs were used to probe parallel blots, only minor differences in relative band intensities were observed. Since neither BamHI nor HindIII are known to restrict PR-protein cDNA clones, this result suggests that the PR-protein genes in tobacco are organized in a small multigene family consisting of approximately seven members. In order to evaluate the generality of this observation, DNA from tomato plants was isolated and hybridized with a tobacco cDNA probe. One major PR-protein from tomato plants (8), p14, which has about 60% homology with the tobacco proteins at the amino acid level and is markedly induced after viroid infection, has been described. On Southern blots, however, at least three genomic fragments hybridize to the tobacco probe even at high stringency (Fig. 6). Thus, the genomes of tobacco as well as tomato plants contain several PR-1 related genes.

DISCUSSION

In this manuscript, we report the isolation and characterization of 32 cDNA clones encoding PR-proteins from TMV infected tobacco plants. By using different restriction endonucleases specific for the 3' untranslated regions of these clones, we identified three classes of cDNAs coding for PR-1a, PR-1b, and PR-1c. Only two cDNA clones could not be assigned unambiguously to one of these classes because they did not extend far enough into the 3' regions. We characterized two clones of each class by DNA sequence analysis. All clones analyzed were independent and clones belonging to the same class differed at most by two nucleotides. These data suggest that only three PR-protein genes are actively transcribed in tobacco plants after TMV infection. On the other hand, hybridization of cDNA probes with BamHI or HindIII digested genomic DNA revealed the presence of about seven fragments homologous to PR-protein sequences. Southern blotting experiments therefore suggest that the PR-protein genes are organized in a small multigene family consisting of at least three and more likely about seven members. This interpretation is consistent with our observation that three independent PR-protein related genes which we isolated as λ

clones from a genomic tobacco library do not have internal BamHI or HindIII sites (A.J.P.Pfitzner, U.M.Pfitzner, and H.M.Goodman, unpublished results). These observations can be accounted for in several ways: (1) The detection of more than three genomic fragments might be due to restriction fragment length polymorphisms in allelic genes. (2) Some of the fragments may encode inactive pseudogenes. (3) The presence of multiple PR-protein genes might reflect a differential activation of these genes by different environmental stimuli. (4) Some of the fragments detected by Southern blotting might carry genes, which are related in their nucleotide sequences to PR-1a, PR-1b, or PR-1c, but which do not give rise to proteins that can be detected by cross-reaction with antibodies raised against PR-1a.

PR-protein cDNA clones are closely related to each other with a homology at the nucleotide level of 90% in the protein coding regions and 80% in the 3' untranslated regions. The homology at the amino acid level is 90%. The number of amino acids in the coding regions of the mature proteins is identical and the calculated molecular weights range from 15.1 kD - 15.2 kD. These properties of PR-protein cDNAs are consistent with earlier conclusions from biochemical and serological studies predicting that the members of the PR-1 group have similar primary structures (3,4,5). Thus, PR-protein genes probably evolved by duplication from a common ancestral gene. In the tobacco genome, however, the genes are not very tightly linked. Lambda genomic clones with inserts of ~15 kb contain only single copies of PR-protein genes (A.J.P.Pfitzner, U.M.Pfitzner, and H.M.Goodman, unpublished results).

After TMV infection, PR-proteins are markedly induced. The increase in protein synthesis is due to highly elevated mRNA levels. This is in agreement with reports from other groups (17,27). We estimate from our cDNA cloning experiment that the PR-protein mRNAs accumulate to >1% of the total mRNA population in the diseased plants. PR-1a mRNA represents the most prominent species (20/30 cDNA clones identified) and PR-1b (6/30 cDNA clones identified) and PR-1c mRNA (4/30 cDNA clones identified) are present at about the same level. This is consistent with our observation that the PR-proteins themselves are also found at levels around 1% of the total soluble proteins in the TMV

infected plants. PR-1a is most abundant and PR-1c is the least abundant, although at the protein level the ratio of PR-1a : PR-1c is on the order of 10:1. Thus, assuming an equal conversion of each mRNA into cDNA, PR-protein mRNAs are present in the total mRNA population at approximately the same level as are PR-proteins in the total protein population. This and the fact that PR-protein mRNAs accumulate to such high amounts in the virus infected plants argue against an exclusive translational control of PR-protein synthesis as was proposed previously (29,30).

Our experiments cannot differentiate between the accumulation of PR-protein mRNAs from increased rates of gene transcription or post-transcriptional RNA stabilization. Recently, however, it has been shown that treatment of parsley cells with elicitor results in a marked transcriptional activation of PR-protein genes and highly increased steady state mRNA levels (31). Transcriptional gene activation can be attributed to 5′ upstream as well as 3′ downstream regulatory sequences (32,33). A detailed study of the structure of PR-protein genes and their immediate flanking sequences may provide information as to whether cis-regulatory elements are associated with these genes and therefore might be involved in gene induction processes controlled by pathogens. Furthermore, comparison of the primary structures of these three closely related genes may allow us to recognize general features that are responsible for high level and low level expression of plant genes.

ACKNOWLEDGEMENTS

We thank Lerzan Kizilay and John A.Smith for the synthesis of oligodeoxynucleotides and Artur J.P.Pfitzner for helpful discussions. This work was supported by a grant from Hoechst AG.

REFERENCES

1. Matthews, R.E.F. (1980) In "Comprehensive Virology 16" (H. Fraenkel-Conrat and R.R. Wagner, eds.) Plenum Press, New York, 297-359.
2. van Loon, L.C. (1985) Plant Mol. Biol. **4**, 111-116.
3. Antoniw, J.F., Ritter, C.E., Pierpoint, W.S. and van Loon, L.C. (1980) J. Gen. Virol. **47**, 79-87.
4. Matsuoka, M. and Ohashi, Y. (1984) J. Gen. Virol. **65**, 2209-2215.
5. Antoniw, J.F., White, R.F., Barbara, D.J., Jones, P. and Longley, A.(1985) Plant Mol. Biol. **4**, 55-60.
6. Ahl, P., Cornu, A. and Gianinazzi, S. (1982) Phytopathology 72, 80-85.

7. Gianinazzi, S. and Ahl, P. (1983) Neth. J. Plant Pathol. **89**, 275-281.
8. Lucas, J., Camacho Henriquez, A., Lottspeich, F., Henschen, A. and Saenger, H.L. (1985) EMBO J. **4**, 2745-2749.
9. Cornelissen, B.J.C., Hooft van Huijsduijnen, R.A.M., van Loon, L.C. and Bol, J.F. (1986) EMBO J. **5**, 37-40.
10. Leberman, R. (1966) Virology **30**, 341-347.
11. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
12. Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA **69**, 264-268.
13. Murray, M.G. and Thompson, W.F. (1980) Nucl. Acids Res. **8**, 4321-4325.
14. Birnboim, H.C. and Doly, J. (1979) Nucl. Acids Res. **7**, 1513-1523.
15. Gubler, U. and Hoffman, B.J. (1983) Gene **25**, 263-269.
16. Hanahan, D. and Meselson, M. (1980) Gene **10**, 63-67.
17. 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.
18. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
19. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA **80**, 3963-3965.
20. McMaster, G.K. and Carmichael, G.G. (1977) Proc. Natl. Acad. Sci. USA **74**, 4835-4838.
21. Davis, B.J. (1964) Ann. N.Y. Acad. Sci. **121**, 404-427.
22. Laemmli, U.K. (1970) Nature **227**, 680-685.
23. Burnette, W.N. (1981) Anal. Biochem. **112**, 195-203.
24. Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H. and Bedbrook, J. (1986) Nucl. Acids Res. **14**, 2229-2240.
25. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature **263**, 211-214.
26. Kozak, M. (1984) Nucl. Acids Res. **12**, 857-872.
27. Carr, J.P., Dixon, D.C. and Klessig, D.F. (1985) Proc. Natl. Acad. Sci. USA **82**, 7999-8003.
28. von Heijne, G. (1985) J. Mol. Biol. **184**, 99-105.
29. Carr, J.P., Antoniw, J.F., White, R.F. and Wilson, T.M.A. (1982) Biochem. Soc. Trans. **10**, 353-354.
30. Matsuoka, M. and Ohashi, Y. (1986) Plant Physiol. **80**, 505-510.
31. Somssich, I.E., Schmelzer, E., Bollmann, J. and Hahlbrock, K. (1986) Proc. Natl. Acad. Sci. USA **83**, 2427-2430.
32. Timko, M.P., Kausch, A.P., Castresana, P., Fassler, J., Herrera-Estrella, L., van den Broek, G., van Montagu, M., Schell, J. and Cashmore, A.R. (1985) Nature **318**, 579-582.
33. Simpson, J., Schell, J., van Montagu, M. and Herrera-Estrella, L.(1986) Nature **323**, 551-554.