MOLECULAR PLANT-MICROBE INTERACTIONS
Vol. 1, No. 3, pp. 107-112, 1988

A Virus-Inducible Tobacco Gene Encoding a Glycine-Rich Protein Shares Putative Regulatory Elements with the Ribulose Bisphosphate Carboxylase Small Subunit Gene

Jan A. L. van Kan,[1] Ben J. C. Cornelissen,[2] and John F. Bol[1]

[1]Department of Biochemistry, Leiden University, Wassenaarseweg 64, 2333AL Leiden and [2]Mogen International N.V., Einsteinweg 97, 2333 CB Leiden, The Netherlands.
Received 21 January 1988. Accepted 15 February 1988.

cDNA to an mRNA that is strongly induced in Samsun NN tobacco after tobacco mosaic virus (TMV) infection or salicylic acid treatment was used to probe a genomic blot and to screen a genomic library. The mRNA corresponds to a family of approximately eight genes, four of which were cloned. The sequence of the genes and flanking DNA in two clones was determined. One gene was found to contain an intron of 555 bp; S_1-nuclease mapping studies indicated that this gene is expressed. The other gene is interrupted by an intron of 1,954 bp and is probably not expressed after TMV infection. The genes encode a protein of 109 amino acids with a putative N-terminal signal peptide of 26 amino acids. The protein contains a high proportion of glycine (25%) and charged amino acids (29%), suggesting that it may be a cell wall component. A comparison of the upstream sequences of the genes encoding the glycine-rich protein and the pathogenesis-related protein 1a showed only limited homology, although both genes are TMV- and salicylic acid-inducible. However, the upstream sequence of the glycine-rich protein gene contains a 64-bp inverted repeat that occurs in a similar position in the tobacco ribulose bisphosphate carboxylase small subunit gene.

Additional keywords: glycine-rich proteins, induced resistance, pathogenesis-related proteins, plant defense genes

Plants reacting hypersensitively to infection by viruses, fungi, or bacteria start to accumulate host proteins that exert their function in the cytoplasm, the cell wall, or the intercellular space of the leaf. Collectively, these proteins are believed to be involved in defense mechanisms resulting in a systemically acquired resistance of the plant to further infection (Collinge and Slusarenko 1987; Bol and Van Kan 1988; Van Loon 1988). The induced cytoplasmic proteins include some key enzymes in the pathways leading to the synthesis of various aromatic compounds such as phytoalexins and lignin. The phytoalexins have an antimicrobial function whereas lignin is linked to pathogen-induced hydroxyproline-rich cell wall proteins in the formation of a barrier around the site of infection. The induced proteins that are excreted into the intercellular space of the leaf are known as pathogenesis-related proteins, or PRs. The ten acidic PRs induced by tobacco mosaic virus (TMV) infection of Samsun NN tobacco—proteins 1a, 1b, 1c, 2, N, O, P, Q, R, and S—have been characterized in most detail. Some of these proteins have β-1,3-glucanase activity (2, N, and O) or chitinase activity (P and Q) and are believed to be involved in a defense against fungal infection (Kauffmann *et al.* 1987; Legrand *et al.* 1987; Boller 1986). These acidic proteins are paralleled by basic equivalents as is the group of PR-1 proteins (Cornelissen *et al.* 1987). PR protein S is homologous to a maize alpha-amylase/trypsin inhibitor with a putative role in defense against insects (Richardson *et al.* 1987).

Address reprint requests to J.F. Bol.

Nucleotide and/or amino acid sequence data is to be submitted to GenBank as accession number J03670.

Circumstantial evidence suggests that some of the pathogen-induced proteins of yet unknown function might be involved in a defense against viral infection. Treatment of plants with salicylic acid results in the accumulation of a subset of PR proteins and an inhibition of more than 90% of virus multiplication (Hooft van Huijsduijnen *et al.* 1986a). To study the function of these proteins in more detail, we have cloned DNA copies of six classes of TMV-inducible tobacco mRNAs, which were called clusters A to F (Hooft van Huijsduijnen *et al.* 1986b). Two of these mRNAs, corresponding to clusters B and C, were strongly inducible by salicylic acid. Cluster B mRNAs encode the acidic PR proteins 1a, 1b, and 1c; the protein encoded by cluster C mRNAs does not correspond to any of the known PR proteins. The Samsun NN genome contains a family of approximately eight genes for acidic PR-1 proteins, and we have cloned and sequenced the PR-1a gene and two putative silent PR-1 genes (Cornelissen *et al.* 1987). Here we report that cluster C mRNAs also correspond to a family of approximately eight genes. Four of these genes were cloned and two were sequenced. As the protein encoded by cluster C mRNAs was found to contain a high proportion of glycine residues, it is tentatively denoted as glycine-rich protein (GRP). The flanking sequences of the GRP gene were compared with published sequences of other pathogen-inducible genes.

MATERIALS AND METHODS

Construction and screening of a tobacco genomic library. Nuclear DNA isolated from young tobacco leaves (*Nicotiana tabacum* cv. Samsun NN) according to Fischer and Goldberg (1982) was partially digested with *Sau*IIIa and cloned in the Charon 35 vector (Loenen and Blattner 1983) by the procedures of Zimmerman *et al.* (1980).

Approximately 2×10^6 independent isolates were plated and amplified to give a permanent library. The average tobacco DNA insert size was 15 kilobases (kb). The library was screened for GRP genes using the plaque hybridization technique of Benton and Davis (1977) and the cDNA clone cGRP-32 (Hooft van Huijsduijnen *et al.* 1986b) as probe. Labeling of the probe was done by nick-translation (Rigby *et al.* 1977).

Analysis of genomic clones. Recombinant phage DNA positively responding to the probe was isolated as described previously (Cornelissen *et al.* 1987). Four different clones were obtained. Fragments of the inserts, obtained by *Eco*RI digestion, were subcloned into pUC9 and subsequently into M13 derivatives tg130 and tg131 (Kieny *et al.* 1983). DNA was sequenced by the dideoxy chain termination method of Sanger *et al.* (1977) using [α-^{35}S]-dATP.

Southern blot analysis. Analysis of nuclear tobacco DNA and genomic clones by Southern blot hybridization was done as described previously (Cornelissen *et al.* 1987).

S_1-nuclease mapping. End-labeling of DNA and S_1-nuclease mapping was done as described by Maniatis *et al.* (1982).

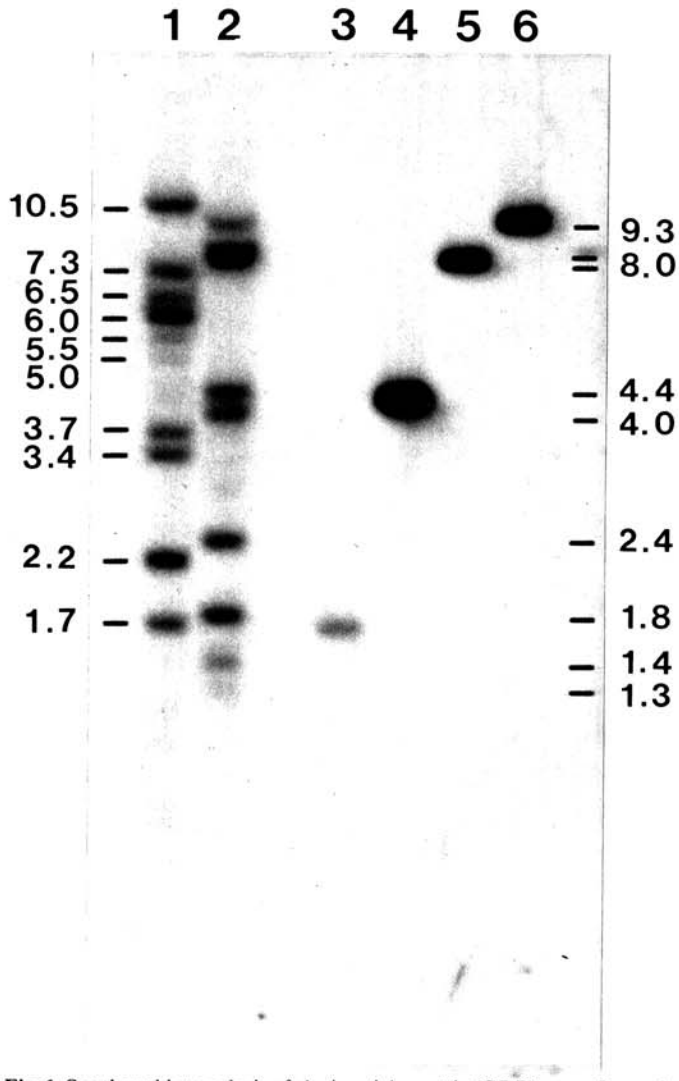

Fig. 1. Southern blot analysis of glycine-rich protein (GRP) genes. Lanes 1 and 2 were loaded with nuclear DNA of Samsun NN tobacco digested with *Hind*III and *Eco*RI, respectively; lanes 3, 4, 5, and 6 were loaded with *Eco*RI digests of the genomic clones gGRP-4, gGRP-8, gGRP-10, and gGRP-14, respectively. The blot was hybridized to the ^{32}P-labeled cDNA clone cGRP-32. The estimated size of the fragments is indicated.

RESULTS

Cloning of GRP genes. Of the cluster C cDNA clones isolated by Hooft van Huijsduijnen *et al.* (1986b), the clone cGRP-32 had the longest insert. Sequence studies showed that it contained a reading frame for the C-terminal 86 amino acids of the GRP followed by a 3′-terminal noncoding sequence of 202 nucleotides. Screening of the tobacco genomic library with this cDNA clone yielded four unique genomic clones with inserts of approximately 15 (gGRP-4), 13 (gGRP-8), 20 (gGRP-10), and 16 kb (gGRP-14). Figure 1 shows a Southern blot analysis of *Eco*RI digests of these clones. Each clone yields one specific fragment hybridizing to the cDNA clone (Fig. 1, lanes 3–6). Analysis of tobacco nuclear DNA digested with *Hind*III or *Eco*RI revealed about eight bands of similar intensity (Fig. 1, lanes 1 and 2), indicating that the amphidiploid Samsun NN genome contains approximately eight GRP genes.

Structure of two GRP genes. The genomic clones gGRP-4 and gGRP-8 were analyzed in more detail. Sequence studies revealed that each clone contained one GRP gene. The mRNA sequence represented in the DNA sequence was interrupted by an intron of 555 nucleotides in clone gGRP-8 and an intron of 1,954 nucleotides in clone gGRP-4. Figure 2A shows a schematic representation of the two genes. Three

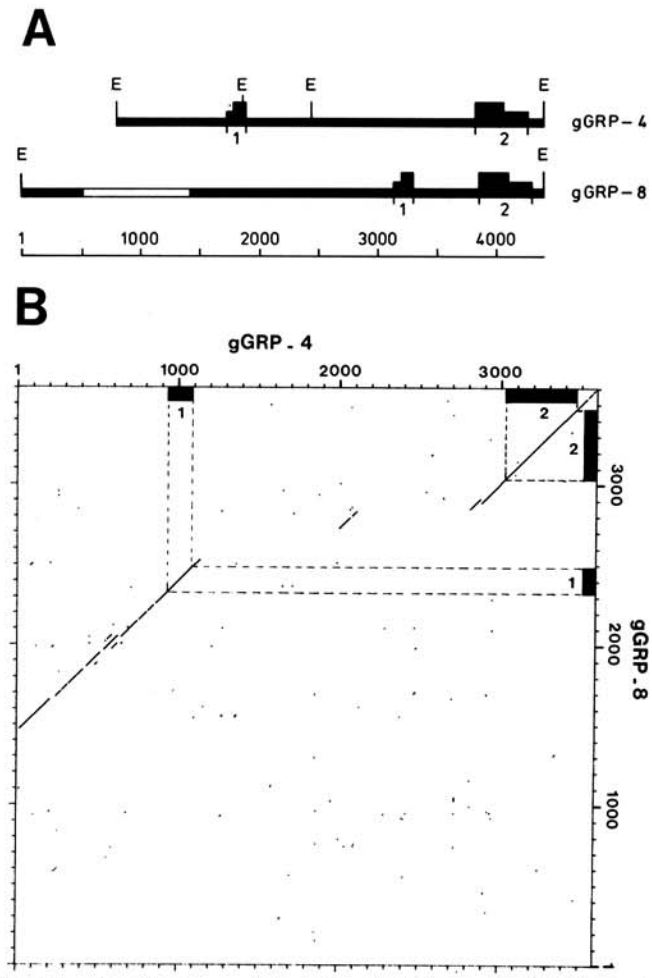

Fig. 2. A, Schematic representation of the *Eco*RI fragments of gGRP-4 and gGRP-8 that were sequenced (solid parts of the bars). The locations of exons 1 and 2 are indicated by boxes. The scale is in basepairs. **B,** Dot plot comparison of the sequence of gGRP-4 and gGRP-8. Nucleotide numbers are given along the axes; the locations of exons 1 and 2 are indicated. The sequences were compared with a window of 21 nucleotides and a stringency of 15 nucleotides.

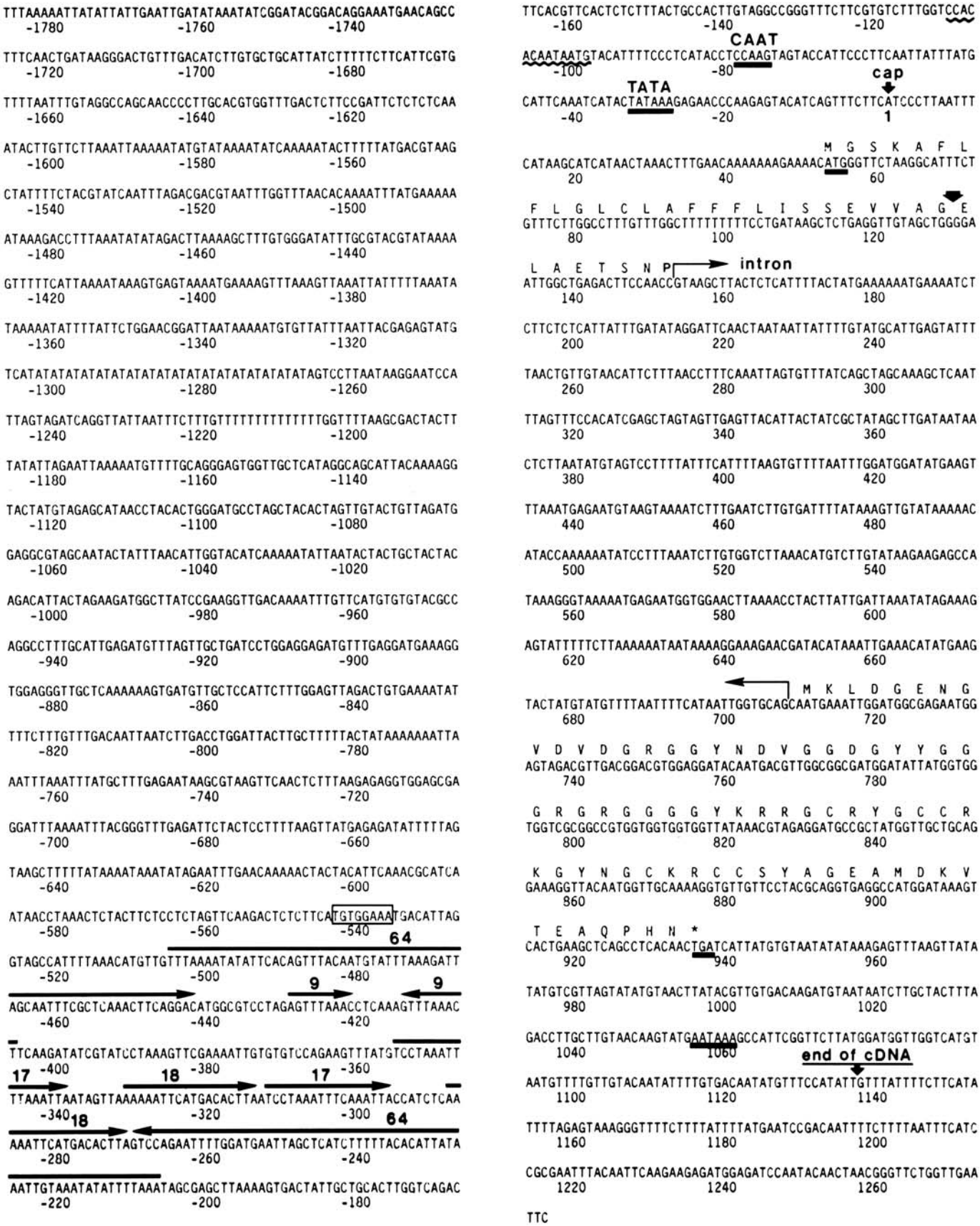

Fig. 3. Nucleotide sequence of the GRP gene and flanking sequences in clone gGRP-8. The GRP reading frame is aligned with the corresponding amino acid sequence. The bold arrow indicates the putative cleavage site involved in the removal of a signal peptide. Initiation and termination signals involved in transcription and translation are underlined. The locations of 17- and 18-bp direct repeats and 9- and 64-bp inverted repeats are indicated by arrows. Putative activator elements are indicated by a box and a wavy line.

adjacent *Eco*RI fragments of clone gGRP-4 with a total length of 3,592 bp were completely sequenced. Only the fragment containing exon 2 was detected (Fig. 1, lane 3). Apparently, the 40 bp of exon 1 that were present in the cDNA clone were insufficient to detect the other two *Eco*RI fragments by hybridization. The 4.4-kb *Eco*RI fragment of gGRP-8 that was shown in lane 4 of Figure 1 was sequenced for the major part (solid bar in Fig. 2A). Figure 3 shows the nucleotide sequence of the 3′ terminal 3,062 bp of this fragment with the amino acid sequences encoded by the two exons. The transcription initiation site, designated nucleotide +1, was determined by S_1-nuclease mapping. An *Hae*III fragment containing nucleotides −135–+82 was labeled with kinase and γ-^{32}P-ATP. Subsequently, it was cleaved with *Ava*II at position −113, and the resulting 195-bp fragment was hybridized to poly(A)-RNA from healthy and TMV-infected tobacco. Figure 4A shows a band of material that is protected from S_1-nuclease degradation by both RNA preparations (lanes H and I), although protection with RNA from infected leaves is higher (lane I). No protection is obtained by hybridization to tRNA (lane T). Although the protected fragments are slightly heterogeneous, we assume that transcription initiates at the cap-site indicated in Figure 3.

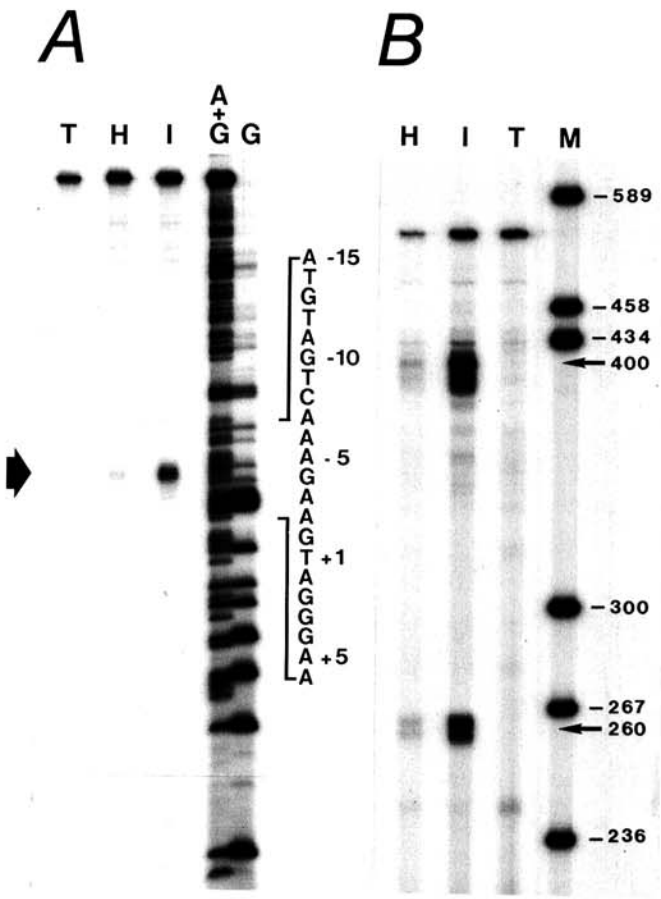

Fig. 4. S_1-nuclease mapping of the 5′ and 3′ ends of GRP mRNA. **A**, A 5′-labeled fragment of gGRP-8 containing the sequence from nucleotides −113–+82 was hybridized to tRNA (lane T) and to poly(A)-RNA from healthy (lane H) or tobacco-mosaic virus (TMV)-infected (lane I) tobacco. Fragments of the probe resistant to S_1-nuclease digestion (arrow) were electrophoresed in parallel to the products of chemical cleavages of the probe showing the position of A and G residues (lane A + G) or G residues (lane G). The sequence around the transcription initiation site in the minus-strand of gGRP-8 is given in the margin. **B**, A 3′-labeled fragment of gGRP-8 containing the sequence from nucleotides 742–1,271 was hybridized to tRNA (lane T) and to poly(A)-RNA from healthy (lane H) or TMV-infected (lane I) tobacco. Fragments of the probe resistant to S_1-nuclease digestion (arrows) were electrophoresed in parallel to marker DNAs (lane M). The size of the markers is given in base pairs.

To map the transcription termination site, we made use of an *Acc*I fragment of gGRP-8 containing the sequence from nucleotide 742 to a sequence in the vector downstream of the subcloned 4.4-kb *Eco*RI fragment. After 3′-end labeling by filling in the *Acc*I sites with Klenow polymerase, the fragment was cut with *Eco*RI at position 1,271 and the resulting 529-bp fragment was hybridized to poly(A)-RNA from healthy and TMV-infected plants. Two fragments with approximate sizes of 400 and 260 nucleotides were found to be protected from S_1-nuclease degradation (Fig. 4B, lanes H and I). The sequence 400 nucleotides downstream of the *Acc*I site at position 742 coincides with the 3′ end of the cDNA clone cGRP-32 (nucleotide 1,136). This position is preceded by the consensus polyadenylation signal AATAAA. The protected fragment of 260 nucleotides may have resulted from hybridization of the probe to a partially homologous GRP mRNA with mismatches around position 1,000.

The incomplete mRNA sequence represented in clone cGRP-32 showed 35 nucleotide substitutions and four small deletions or insertions of 3–9 nucleotides, compared to the genomic sequence in gGRP-8. Also, the cDNA sequence did not completely match with the GRP gene in gGRP-4. The S_1-nuclease mapping data demonstrate, however, that gGRP-8 corresponds to an expressed gene. When the S_1 experiments were performed with probes from gGRP-4, no protected fragments were obtained, indicating that this clone corresponds to a gene that is not expressed after TMV infection (data not shown).

Figure 3 shows that the gene in gGRP-8 consists of a leader sequence of 52 nucleotides, a coding region of 109 triplets, and a 3′-noncoding region of approximately 200 nucleotides. The coding region is interrupted by an intervening sequence of 555 nucleotides in the proline codon at amino acid-position 34. In clone gGRP-4 the intron is at the same position, but has a length of 1,954 nucleotides. The splice donor sequence C/GTAAG in gGRP-8 has been changed to G/ATAAG in gGRP-4. This may result in a defective splicing, being responsible for our inability to detect a transcript of this gene by S_1-nuclease mapping. As in gGRP-8, the coding region in gGRP-4 contained 109 triplets.

Figure 2B shows a dot plot comparison of the sequences of gGRP-4 and gGRP-8 made by using the UWGCG computer program. The exons and the flanking regions of the two genes are highly homologous, but there is only limited homology between the introns. The intron in gGRP-8 is extended by 1,400 bp in gGRP-4 due to a replacement of 220 nucleotides (positions 197–417 in Fig. 3) by a nonhomologous sequence of 930 nucleotides and the insertion of a sequence of 690 nucleotides at position 508/509.

As indicated in Figure 3, the upstream sequence of the GRP gene shows a number of inverted and direct repeats and other elements that may be involved in the regulation of its expression. These are discussed below.

Properties of the GRP. The N-terminus of the GRP is strongly hydrophobic and the amino acid sequence around position 26/27 (bold arrow in Fig. 3) shows resemblance to the cleavage site that is used in the removal of the signal peptide of several PR proteins, in particular the bean chitinase (Broglie *et al.* 1986). Therefore, we assume that the N-terminal 26 amino acids encoded by the GRP gene represent a signal peptide. Table 1 shows that the putative

mature protein of 83 amino acids contains 11 acidic and 13 basic residues. A hydrophilicity plot revealed that this protein is strongly hydrophilic over its entire length (data not shown). Moreover, Table 1 illustrates that the putative mature GRP contains a high proportion of glycine, that is, 21 residues or 25%. These glycine residues are particularly clustered near the central part of the polypeptide (encoded by nucleotides 750–820).

DISCUSSION

The mRNA encoding the GRP rapidly accumulates to high levels in tobacco sprayed with 5 mM salicylic acid or infected with TMV (Hooft van Huijsduijnen *et al.* 1986b). However, the GRP itself has not yet been identified in these plants. The hybrid-selected GRP mRNA is efficiently translated in the reticulocyte cell-free system, indicating that it is a functional messenger. The *in vitro*-made translation product is not precipitated by an antiserum to a mixture of tobacco PR proteins (Hooft van Huijsduijnen *et al.* 1986b), but it is precipitated by an antiserum raised against a synthetic C-terminal peptide specified by the genomic sequence (Van Kan *et al.*, unpublished results). With this latter antiserum we were unable to detect the GRP in extracts from TMV-infected or salicylic-acid treated tobacco (Van Kan *et al.*, unpublished results). This may be due to a post-translational modification of the C-terminus of the protein or to a difficult accessibility of the protein in the plant. Recently, the structure has been reported of a petunia gene encoding a GRP that is likely to function as a structural cell wall component (Condit and Meagher 1986). Moreover, there is evidence that GRPs fulfill a structural role in many plant species (Varner and Cassab 1986). If the tobacco GRP turns out to be a cell wall protein, it may be functionally equivalent to another class of pathogen-inducible plant defense proteins, the hydroxyproline-rich glycoproteins that are found in cell walls (Ecker and Davis 1987).

Because the PR-1 genes and GRP genes are both inducible by TMV-infection and salicylic acid treatment, their flanking sequences were searched for possible common regulatory elements. A dot plot matrix did not reveal any significant homology between the two classes of genes. As shown in Figure 5A, however, visual inspection did reveal some sequence similarities just upstream of the cap-site of the two genes. The significance of this homology is not yet clear.

Between nucleotides −510 and −200 the upstream sequence of the GRP gene contains numerous repeats. Direct repeats of 17 and 18 bp and inverted repeats of 9 and 64 bp are indicated in Figure 3. A comparison by computer analysis with published sequences of other pathogen-inducible plant genes (Bol and Van Kan 1988) did not show any conservation of these repeats or other sequence similarities. However, comparison with a gene encoding the small subunit of ribulose bisphosphate carboxylase (rbcS) from tobacco (Mazur and Chui 1985) revealed that the 64-bp inverted repeat is present in the upstream sequence of both the GRP and rbcS gene. Figure 5B shows that the mismatches between the two sequences are of the same order of magnitude as those occurring between the repeats in the GRP gene (asterisks in Fig. 5B). In this respect it is interesting to note that in a study on developmental regulation of mRNAs that are induced upon cytokinin stress, the GRP mRNA was found to be regulated in an organ-specific and light-regulated manner, similar to the rbcS mRNA (J. Memelink, J. H. C. Hoge, and R. A. Schilperoort, unpublished data). This may suggest that the 64-bp inverted repeat has a regulatory function. Further analysis of the upstream sequence of the GRP gene revealed a homology around position −540 with the SV-40 enhancer (box in Fig. 3) and a homology around position −110 (wavy line in Fig. 3) with the activator element that has been identified in the cytokinin gene from the octopine T-DNA of *Agrobacterium tumefaciens* and a number of plant genes (De Pater 1987). Experiments in which the flanking regions of the GRP gene are fused to a reporter gene are being carried out to further characterize the elements that regulate the expression of the GRP gene in response to light, TMV-infection, or salicylic acid treatment.

ACKNOWLEDGMENTS

We thank Frans Th. Brederode and Miranda van de Rhee for technical assistance. This work was sponsored in part by The Netherlands

Table 1. Amino acid composition of the glycine-rich protein

	Total protein	Putative signal peptide	Putative mature protein
Ala	7	3	4
Val	6	2	4
Leu	7	5	2
Ile	1	1	0
Pro	2	0	2
Phe	5	5	0
Trp	0	0	0
Tyr	7	0	7
Met	3	1	2
Gly	24	3	21
Ser	5	3	2
Thr	2	0	2
Cys	7	1	6
Asn	5	0	5
Gln	1	0	1
Asp	6	0	6
Glu	6	1	5
Lys	6	1	5
Arg	8	0	8
His	1	0	1
	109	26	83

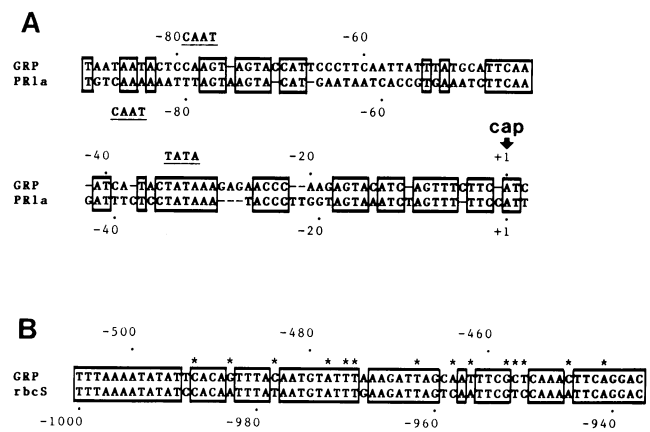

Fig. 5. Comparison of the upstream sequence of the GRP gene with those of the tobacco PR-1a gene (**A**) and the tobacco ribulose bisphosphate carboxylase small subunit (rbcS) gene (**B**). Identical nucleotides are boxed. The sequence shown in **B** is that of a 64-bp inverted repeat occurring both in the GRP and the rbcS gene. The asterisks show the mismatches in the 64-bp inverted repeat of the GRP gene. The PR-1a sequence is from Cornelissen *et al.* (1987); the rbcS sequence is from Mazur and Chui (1985). Numbering of the PR-1 and GRP sequence is from the cap-site; numbering of the rbcS sequence is from the ATG initiation codon.

Foundation for Chemical Research (S.O.N.) with financial aid from The Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

LITERATURE CITED

Benton, W. D., and Davis, R. W. 1977. Screening lambda gt recombinant clones by hybridization to single plaques *in situ*. Science 196:180-182.

Bol, J. F., and van Kan, J. A. L. 1988. The synthesis and possible functions of virus-induced proteins in plants. Microbiol. Sci. 5:47-52.

Boller, T. 1986. Hydrolytic enzymes in plant disease resistance. In: Plant-Microbe Interactions: Molecular and Genetic Perspectives. E. W. Nester and T. Kosuge, eds., pp. 385-413. Macmillan, New York.

Broglie, K. E., Gaynor, J. J., and Broglie, R. M. 1986. Ethylene-regulated gene expression: Molecular cloning of the genes encoding an endochitinase from *Phaseolus vulgaris*. Proc. Natl. Acad. Sci. USA 83:6820-6824.

Collinge, D. B., and Slusarenko, A. J. 1987. Plant gene expression in response to pathogens. Plant Mol. Biol. 9:389-410.

Condit, C. M., and Meagher, R. B. 1986. A gene encoding a novel glycine-rich structural protein of petunia. Nature 323:178-181.

Cornelissen, B. J. C., Horowitz, J., van Kan, J. A. L., Goldberg, R. B., and Bol, J. F. 1987. Structure of tobacco genes encoding pathogenesis-related proteins from the PR-1 group. Nucleic Acids Res. 15:6799-6811.

De Pater, B. S. 1987. Plant expression signals of the *Agrobacterium* T-cyt gene. Ph.D. thesis, University of Leiden.

Ecker, J. R., and Davis, R. W. 1987. Plant defense genes are regulated by ethylene. Proc. Natl. Acad. Sci USA 84:5202-5206.

Fischer, R. L., and Goldberg, R. B. 1982. Structure and flanking regions of soybean seed protein genes. Cell 29:651-660.

Hooft van Huijsduijnen, R. A. M., Alblas, S. W., De Rijk, R. H., and Bol, J. F. 1986a. Induction by salicylic acid of pathogenesis-related proteins and resistance to alfalfa mosaic virus infection in various plant species. J. Gen. Virol. 67:2135-2143.

Hooft van Huijsduijnen, R. A. M., Van Loon, L. C., Bol, J. F. 1986b. cDNA cloning of six mRNAs induced by TMV-infection of tobacco and a characterization of their translation products. EMBO J. 5:2057-2061.

Kauffmann, S., Legrand, M., Geoffroy, P., and Fritig, B. 1987. Biological function of pathogenesis-related proteins: Four PR-proteins have β-1,3-glucanase activity. EMBO J. 6:3209-3212.

Kieny, M. P., Lathe, R., and Lecocq, J. P. 1983. New versatile cloning and sequencing vectors based on bacteriophage M13. Gene 26:91-99.

Legrand, M., Kauffmann, S., Geoffroy, P., and Fritig, B. 1987. Biological function of pathogenesis-related proteins: Four tobacco pathogenesis-related proteins are chitinases. Proc. Natl. Acad. Sci. USA 84:6750-6754.

Loenen, W. A. M., and Blattner, F. R. 1983. Lambda charon vectors (Ch 32, 33, 34 and 35) adapted for DNA cloning in recombination deficient hosts. Gene 26:171-179.

Maniatis, T., Fritsch, E. F., and Sambrook, J. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, New York.

Mazur, B. J., and Chui, C. F. 1985. Sequence of a genomic DNA clone for the small subunit of ribulose bisphosphate carboxylase-oxygenase from tobacco. Nucleic Acids Res. 13:2373-2386.

Richardson, M., Valdes-Rodriguez, S., and Blanco-Labra, A. 1987. A possible function for thaumatin and a TMV-induced protein suggested by homology to a maize inhibitor. Nature 327:432-434.

Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick-translation with DNA polymerase I. J. Mol. Biol. 113:237-251.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Van Loon, L. C. 1988. Stress proteins in infected plants. In: Plant-Microbe Interactions: Molecular and Genetic Perspectives. E .W. Nester and T. Kosuge, eds., Macmillan, New York. In press.

Varner, J. E., and Cassab, G. I. 1986. A new protein in petunia. Nature 323:110.

Zimmerman, C. R., Orr, W. C., Leclerc, R. F., Barnard, E. C., and Timberlake, W. E. 1980. Molecular cloning and selection of genes regulated in *Aspergillus* development. Cell 21:709-715.