The N-terminal protein of the polyprotein encoded by the potyvirus tobacco vein mottling virus is an RNA-binding protein

John D. Brantley† and Arthur G. Hunt*

Plant Physiology/Biochemistry/Molecular Biology Program, Department of Agronomy, University of Kentucky, Lexington, Kentucky 40546-0091, U.S.A.

The first predicted polypeptide encoded by the potyvirus tobacco vein mottling virus (TVMV) is a highly positively charged protein of predicted M, 29K that functions as a protease to perform the first predicted cleavage in the potyvirus polyprotein. We expressed this protein (P1 pro) fused with glutathione S-transferase (GST) and purified the fusion protein from engineered Escherichia coli. We found that the intact fusion protein, as well as samples in which the P1 pro portion was liberated from GST by pretreatment with thrombin, was able to bind RNA. Binding activity was optimal at relatively high KCl concentrations, suggesting an interaction dependent on a specific protein structure and not just on the binding of the negatively charged phosphate backbone by the positively charged P1 pro polypeptide. The TVMV P1 pro preferred ssRNA over DNA or dsRNA, and showed a possible preference for sequences containing oligo(G) tracts. Like other potyvirus-encoded proteins, the TVMV P1 pro therefore possesses more than one demonstrable biochemical activity and probably plays multiple roles in the TVMV life cycle.

Introduction

The potyviruses, the largest known group of plant viral pathogens, are flexuous rod-shaped, positive-sense ssRNA viruses. These viruses have genomes of approximately 9-5 kb that encode a single polyprotein that is co- or post-translationally cleaved into seven to nine mature gene products (Dougherty & Carrington, 1988; Riechmann et al., 1992). Functions have been assigned for many of these proteins. The coat protein (CP) encapsulates the viral RNA and plays a role in aphid transmission of this virus (Atreya et al., 1990) in conjunction with the helper component (HC-pro) (Thornbury et al., 1985). HC-pro is also the protease responsible for cleavage of the polyprotein at the C terminus of HC (Carrington et al., 1989). The C-terminal half of NIa is another protease (Dougherty et al., 1988; Hellmann et al., 1988), and the N-terminal half, or perhaps the whole protein, is the genome-linked protein (VPg) that binds to the 5' terminus of the viral RNA (Siaw et al., 1985; Murphy et al., 1990). Sequence analysis of Nb indicates some homology to RNA-dependent RNA polymerases from other positive-strand RNA viruses (Domier et al., 1987). Finally, CI, the principal protein component of the cylindrical inclusions that form in the cytoplasm of infected cells, is similar in sequence to the membrane-bound, replication-associated 2C protein of picornaviruses (Domier et al., 1987) and has dsRNA helicase activity (Lain et al., 1991).

The functions of some of these proteins, on the other hand, remain unclear. In particular, the role(s) of the first predicted protein [termed P1 pro here, which for tobacco vein mottling virus (TVMV) is a 29K protein] in the viral life cycle remain poorly understood. This protein participates in the proteolytic cleavage that liberates P1 pro from the rest of the polyprotein (Verchot et al., 1991). Computer analysis of the sequence of P1 pro of TVMV indicated that the net charge of this protein is +30, which is far higher than the charge of any of the other potyvirus-encoded proteins (A. G. Hunt, unpublished). A net charge of this magnitude could be indicative of a nucleic acid-binding protein. We therefore decided to characterize the possible RNA-binding properties of this protein. Here we show that the TVMV P1 pro does bind ssRNA, that the size of its binding site is consistent with its net charge and that some sequences are bound with greater affinity than others.

Methods

Recombinant DNA manipulations. The structure of the gene encoding the glutathione S-transferase (GST)-P1 pro fusion protein used in these studies is shown in Fig. 1. To assemble this gene, nucleotides (nt) 206 to 1105 of the TVMV genome were amplified by PCR from a plasmid (pBS1220) containing a complete cDNA copy of the TVMV genome. For these reactions, two oligonucleotides were used; the 5' oligo-
The region that encodes the fusion protein is shown with the different coding region; HC', amino acids 257 to 300 of the TVMV polyprotein, protease are shown. For the nucleic acid and amino acid sequences corresponding to residues 1 to 43 of the HC-pro protein) and the corresponding amino acid compositions at the junctions of each section, the restriction enzyme sites used to assemble the fusion protein are given below each section. The nucleotide sequences and gene (underlined) and sites of cleavage by thrombin and the PW° protease are shown. For the nucleic acid and amino acid sequences shown, TVMV-derived sequences are highlighted with bold type.

nucleotide, 5' GGGGATCCATGTCACACTCACTGCACGT-CACA 3', contained 27 nt complementary to the 5' end of the P1° coding region and a BamHI cleavage site which was used for cloning into the expression vector; similarly, the 3' oligonucleotide, 5' AAGAATTCCTGACATTTAAAGCTGACATGGC 3', possessed 26 nt complementary to nt 1079 to 1105 of the TVMV genome and a flanking EcoRI site. Because there is an EcoRI site internal to the P1° sequence, the amplified product was subjected to a partial EcoRI digestion, followed by a BamHI digestion. The high Mf band was isolated by agarose gel electrophoresis and cloned into BamHI and EcoRI-digested pGEX-2T (Pharmacia). After confirmation of the presence of the correct plasmid in a minipreparation, the plasmid pGEX-P1° was used to transform Escherichia coli JM101 cells for production of the GST-P1° fusion protein.

Production of GST-P1° fusion protein. A 3 ml culture of JM101 containing the expression vector was grown overnight in LB with 100 µg/ml ampicillin (LB-amp). The next morning, 200 ml of LB-amp was inoculated with 200 µl of the overnight culture and grown at 37 °C until mid-log phase, at an approximate optical density at 600 nm of 0.7. At this point the expression of the fusion protein was induced with 1 mM-IPTG (Sigma). Growth was continued at 10 °C overnight to maximize the production of soluble fusion protein (Smith & Corcoran, 1990; Shirano et al., 1990).

Cells were collected by centrifugation at 4000 g, resuspended in 20 ml of PBS (0.15 M-NaCl, 0.016 M-Na2HPO4, 0.004 M-NaH2PO4) containing 0.1% Triton X-100 and 1 mM-PMSF, and ruptured by sonication. Cellular debris was centrifuged at 10,000 g at 37 °C until the supernatant was clear. Soluble fusion protein was purified from the supernatant by chromatography on glutathione-Sepharose 4B (Pharmacia) as recommended by the manufacturer; the eluate was concentrated using Centricon cartridges (Amicon) and stored in 50 mM-Tris-HCl pH 8.0. This preparation contained a single RNA-binding protein but was not pure owing to the presence of apparent breakdown products. Therefore the concentration of the fusion protein was estimated by Coomassie blue staining as compared with a series of BSA standards.

In some cases, the purified fusion protein was treated with thrombin prior to further analysis. For this, 100 µl of purified fusion protein (about 800 ng) was incubated with 2 µl (260 ng) of thrombin (Boehringer-Mannheim) for 1 h at 37 °C; this was used directly for subsequent experiments.

Preparation of labelled and unlabelled RNA. The RNAs used here were synthesized using T7 or T3 RNA polymerase (BRL Life Technologies) using standard conditions (Sambrook et al., 1989). For the production of labelled RNAs, [α-32P]ATP was included in the transcription reactions at a concentration of 1 to 2 µCi/ml. The template for preparing labelled RNA was a pBluescript (Stratagene) clone carrying nt +128 to +497 (with respect to the 35S transcription initiation site) of the cauliflower mosaic virus genome (pCAMV-57APA). In addition, this clone had a triple point mutation in which the AUAAA polyadenylation signal (located between nt +163 and +168) was replaced with the sequence UAGAAU (Mogen et al., 1990). This plasmid was digested with BglII before transcription with T7 RNA polymerase; after transcription, this would yield a 151 nt RNA with no known RNA processing signals. In some experiments, unlabelled competitor RNAs were prepared from a pBluescript (Stratagene) clone carrying nt -235 to +410 of the pea rbs-E9 polyadenylation signal (Coruzzi et al., 1984; Hunt & MacDonald, 1989). RNAs were recovered from transcription reactions by phenol/chloroform extraction and ethanol precipitation; in labelled RNA preparations, no further steps were taken to remove unincorporated nucleotide precursors. Double-stranded competitors were prepared by mixing equal amounts of the rbsE (+) and rbsE (-) RNAs (in water), heating to 85 °C for 5 min and cooling to 26 °C at the rate of 2 °C/min. Poly(A) and poly(G) both with average lengths of approximately 10000 nt (Boehringer Mannheim), were used as supplied by the manufacturer. TVMV RNA, isolated from purified virus, was provided by J. Shaw.

Northwestern blot experiments. Protein samples were separated by SDS-PAGE and transferred to nitrocellulose (Sambrook et al., 1989). Filters were then incubated at room temperature for the indicated times as follows: 10 min in 100 ml CGB (30 mM-Tris-HCl pH 8.0, 0.1 M-KCl, 1 mM-CaCl2, 1 mM-MgCl2 and 5 mM-EDTA); 15 min in 50 ml CGB with 50 µg/ml salmon sperm DNA; 30 min in 25 ml CGB with 250 µg/ml salmon sperm DNA and approximately 106 d.p.m. of labelled pCAMV-57APA RNA (the final RNA concentration was approximately 30 ng/ml), two washes (1 min each) in 25 ml CGB, two washes of 10 min each with 50 ml CGB with 50 µg/ml salmon sperm DNA; and 20 min in 200 ml CGB, three rinses with 25 ml each CGB (1 min per rinse), and one rinse (1 min) with H2O. The filter was then dried under vacuum at 80 °C for 15 min, exposed to X-ray film and the film was then developed to reveal any RNA-binding polypeptides.

Filter binding experiments. Measurement of the binding of binding on salt concentration, protein saturation curve determination and competition experiments were performed using filter binding as described elsewhere (Yang & Hunt, 1992). Reactions contained labelled RNA and protein (including either 8 units RNasin (Promega) or 1 unit Inhibitase (5Prime-3Prime)) in a total volume of 10 µl of 10 mM-Tris-HCl pH 8.0, 1 mM-EDTA and 250 mM-NaCl, except where otherwise noted; these were incubated for 15 min at 30 °C. After incubation, the samples were diluted with 1 ml of reaction buffer and applied onto a nitrocellulose filter disc which had been pre-wetted with buffer. The filters were washed with 10 ml of the same buffer at a flow rate of about 20 ml/min. The filter discs were dried at room temperature, placed in scintillation vials, covered with Ecolome scintillation cocktail (ICN) and counted.

Results

Production of a GST-P1° fusion protein

The TVMV P1° has not been isolated from infected tissue, and the expression of the P1° gene by itself in E. coli results in the product being trapped in
RNA binding by a potyvirus-encoded protein

Fig. 2. SDS–PAGE analysis of the P1<sup>pro</sup> derivatives. (a) Stained gel. Lane 1, M<sub>r</sub> standards (shown on the right). Lane 2, 200 ng of fusion protein that had been purified by chromatography on glutathione-Sepharose 4B. Lane 3, 200 ng of purified fusion protein after treatment with thrombin. Lane 4, 500 ng of gel-purified P1<sup>pro</sup>. Lane 5, 100 ng of purified GST. (b) Northwestern blot of purified fusion protein and P1<sup>pro</sup>. Lane 1, 200 ng of purified GST-P1<sup>wt</sup>, the same sample as in (a, lane 2). Lane 2, 500 ng of gel-purified P1<sup>pro</sup>, the same sample as in (a, lane 4). Samples were separated by SDS–PAGE, proteins being transferred to nitrocellulose, renatured, and assessed for RNA-binding as described in Methods. The positions of the fusion protein and purified P1<sup>pro</sup> are shown on the left and the positions of M<sub>r</sub> standards are shown on the right.

When the plasmid pGEX-2T was induced with IPTG and grown at 37 °C, high levels of soluble GST were produced. Unfortunately, under the same conditions, the product expressed from pGEX-P1<sup>pro</sup> formed insoluble inclusion bodies. One method for producing soluble protein in such cases is to induce at a lower temperature (Smith & Corcoran, 1990; Shirano & Shibata, 1990). Induction of the fusion protein at 10 °C produced enough soluble protein for purification, but the majority of the protein remained insoluble (data not shown).

When proteins produced in cells carrying pGEX-P1<sup>pro</sup> that had been induced at 10 °C were purified on a glutathione-Sepharose 4B column, analysis of the products eluted with glutathione by SDS–PAGE revealed several proteins, with the predominant polypeptide being of approximately 58K (Fig. 2a, lane 2). This size is consistent with that predicted for the GST–P1<sup>pro</sup> fusion protein (56K), and slightly smaller than that predicted (61K) for the fusion protein if cleavage by the P1<sup>pro</sup> protease did not occur. Since this polypeptide was recognized by antisera against P1<sup>pro</sup> (data not shown), we assumed that this protein was in fact the predicted GST–P1<sup>pro</sup> fusion protein.

As we indicated, our construction of pGEX-P1<sup>pro</sup> preserved the thrombin cleavage site engineered into pGEX-2T. This allowed the cleavage of the GST from P1<sup>pro</sup> (Fig. 2a, lane 3). Our experiments indicated that the intact fusion protein behaved identically to the cleaved protein (see below). For this reason, and because Northwestern blot analysis indicated that the intact fusion protein was the only RNA-binding species in the sample eluted from the glutathione-Sepharose 4B column with glutathione (Fig. 2b, lane 1), we used the
intact fusion protein to study the RNA-binding properties of TVMV P1^pr°.

The concentration of the fusion protein in our preparation was estimated by comparison of the density of Coomassie blue R-250 staining of the fusion protein band with the density of known BSA standards. Routinely, a 200 ml culture yielded 4 to 6 μg of fusion protein, sufficient for several experiments.

RNA binding experiments

Initial Northwestern blot assays indicated that the GST–P1^pr° fusion protein, as well as P1^pr° that had been overproduced in E. coli, were able to bind labelled RNA (Fig. 2b, lanes 1 and 2). In these experiments, an additional 14K species was seen to bind RNA in the P1^pr° sample; however, this was apparently an E. coli-encoded protein since it was also present in control extracts lacking any P1^pr°-related species (data not shown). This species could be separated from the GST–P1^pr° fusion protein by chromatography on glutathione–Sepharose 4B, since the GST–P1^pr° fusion protein was the only RNA-binding species in affinity-purified preparation (Fig. 2b, lane 1). The RNA-binding properties of this latter preparation were thus characterized in more detail.

Because of the highly charged nature of P1^pr°, it was possible that the observed binding reflected a non-physiological association of negatively charged RNA with positively charged P1^pr°. Accordingly, we examined the ionic strength dependence of the binding of RNA by GST–P1^pr° fusion proteins using a filter binding assay, a more quantitative measure of binding than Northwestern blot assays. The salt dependence profile of RNA binding showed two major peaks, one at 10 to 15 mM-NaCl, the other at 220 to 225 mM-NaCl (Fig. 3). The binding at low ionic strength is reminiscent of the salt dependence of the non-specific binding of a highly charged polymer to a nucleic acid, for example the binding of polylysine to homopolymer (Mascotti & Lohman, 1990). The peak of binding at 225 mM-NaCl cannot be explained on this basis. Thus the binding at 225 mM-NaCl suggests a possible physiological activity.

In order to evaluate the possibility that RNA binding might involve the GST portion of the fusion protein, filter binding experiments were undertaken with protein samples which had been cleaved with thrombin. No significant differences between thrombin-treated and untreated fusion protein samples were seen (Fig. 4). Also, binding experiments performed with purified GST alone showed no RNA-binding activity. This indicated that the P1^pr° portion of the fusion protein was responsible for RNA binding.

When the amount of fusion protein was varied in these reactions, maximal binding was seen at a protein:RNA ratio of 5:1 (w/w; Fig. 4). This value was not affected by prior thrombin treatment of the fusion protein, indicating that the GST portion of the fusion protein apparently had little effect on the binding of the fusion protein to RNA. Since the fusion protein is approximately 55K and the 140 nt RNA is about 47K, this translated to roughly four GST–P1^pr° molecules bound per RNA molecule. In turn, this implied a maximum binding site size of about 35 nt.

Competition with different unlabelled nucleic acids
was done to assess the binding specificity of P1<sup>pro</sup> (Table 1). The ssRNA molecules showed 50% inhibition ranging from a competitor:probe (w/w) ratio of 2 for poly(G) to a ratio of about 130 for the rbcS (−) RNA. An ssDNA, M13 ssDNA, reached 50% inhibition at a ratio of 120. Fifty percent inhibition for dsRNA was seen at a ratio of 330, and the ratio at which dsDNA (Bluescript linearized with EcoRI) inhibited binding by 50% was 500. From these results, it is apparent that the RNA-binding properties of the GST–P1<sup>pro</sup> fusion protein are somewhat dependent on the nucleic acid involved; RNA is preferred over DNA [with the curious exception of the rbcS (−) RNA] and single-stranded nucleic acids over double-stranded molecules.

### Discussion

We have shown that the TVMV P1<sup>pro</sup> is an RNA-binding protein. Our experiments were performed on a fusion protein owing to the tendency of the protein to produce insoluble inclusion bodies when overexpressed in *E. coli*. Even so, most of the fusion protein produced was also insoluble; highly charged proteins are likely to produce insoluble fusion proteins (Smith & Johnson, 1988). Fortunately, it was possible to obtain sufficient soluble material to study the binding characteristics of the protein.

The charge characteristics of the protein are also reflected in the salt dependence for binding. Our data indicate two peaks of activity, one at 10 mm-NaCl, the other at 225 mm-NaCl. We interpret the activity at low ionic strength as representing non-specific binding and the binding at the higher salt concentration as representing physiological activity.

The difference in affinity among the ssRNA-like molecules varies by a factor of 50. P1<sup>pro</sup> has the highest affinity for poly(A), about five times higher than TVMV RNA and poly(A). P1<sup>pro</sup> binds roughly seven times more tightly to the positive-sense strand rbcS RNA used here than it does to the minus-sense rbcS RNA. There is a run of three G residues in the rbcS (+) RNA whereas the rbcS (−) lacks any stretches of G (Coruzzi et al., 1984). With the exception of poly(A), the affinity of P1<sup>pro</sup> for an RNA sequence seems to correlate with the presence of poly(G) tracts. Whether or not this preference is reflected in the association of P1<sup>pro</sup> with TVMV (+)- or (−)-RNAs remains to be determined.

Our experiments suggest that P1<sup>pro</sup> binds to roughly 35 nt. This is much larger than the binding site of tobacco mosaic virus P30 protein, which binds 4 to 7 nt (Citovsky et al., 1990), but is comparable to the 28 to 30 nt binding site reported for the *Agrobacterium* VirE2 protein (Citovsky et al., 1989). This size of binding site is also consistent with the number of phosphate groups that would be neutralized by binding of a protein of charge +30 (Mascott & Lohman, 1990).

Several potyvirus-encoded proteins apparently interact with RNA as a result of their respective activities. The CI protein has been shown to unwind RNA–RNA duplexes (Lain et al., 1991), a function apparently reflective of a role in separating progeny from template RNA molecules during replication. The NIa protein probably associates transiently with RNA at some point during replication since the VPG is located in this protein. The NIb protein, the putative RNA-dependent RNA polymerase, necessarily must be able to interact with RNA templates. The CP is the protein that forms stable complexes with viral RNA so that mature virions may form. The biological role of RNA binding by P1<sup>pro</sup>, if any, is not known. We can speculate that the difference in affinity of P1<sup>pro</sup> for ssRNA and dsRNA may reflect a role in unfolding the viral RNA by perturbing the equilibrium between single- and double-stranded regions, thus facilitating replication. Alternatively, P1<sup>pro</sup> might convert random coil regions into more regular strands without disturbing double-stranded regions, similar to the effect of the single-stranded binding protein. It is conceivable that this complex could play a role in cell-to-cell movement of the virus, similar to that proposed for the tobacco mosaic virus 30K protein (Citovsky et al., 1990), by converting the molecule into an easily transportable form.

There may be other roles that P1<sup>pro</sup> plays in other aspects of the viral life cycle. Several replication-associated functions needed for the production of TVMV progeny have yet to be linked to particular virus-encoded genes; these include factors that might anchor viral RNA to intracellular sites of replication and agents responsible for controlling the relative levels of (+)- and (−)-strand TVMV RNAs. The nature of P1<sup>pro</sup> seems particularly well-suited to the former of these possibilities; it has the

### Table 1. Competition for binding of different classes of nucleic acids

<table>
<thead>
<tr>
<th>Competitor</th>
<th>Type*</th>
<th>(I_{50})†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poly(G)</td>
<td>ssRNA</td>
<td>2</td>
</tr>
<tr>
<td>Poly(A)</td>
<td>ssRNA</td>
<td>5</td>
</tr>
<tr>
<td>TVMV RNA</td>
<td>ssRNA</td>
<td>10</td>
</tr>
<tr>
<td>rbcS (+)</td>
<td>ssRNA</td>
<td>20</td>
</tr>
<tr>
<td>rbcS (−)</td>
<td>ssRNA</td>
<td>130</td>
</tr>
<tr>
<td>M13</td>
<td>ssDNA</td>
<td>120</td>
</tr>
<tr>
<td>rbcS-DNA</td>
<td>dsDNA</td>
<td>330</td>
</tr>
<tr>
<td>Bluescript</td>
<td>dsDNA</td>
<td>500</td>
</tr>
</tbody>
</table>

*Type of nucleic acid used for competition.
† The ratio (w/w) of the competitor to probe that results in a 50% reduction of binding of probe to a standard amount of 34K fusion protein. RNA-binding reactions contained 2 ng of labelled RNA, 16 ng unlabelled competitor RNAs, all in 10 mM-Tris-HCl pH 8.0, 1 mM-EDTA, 225 mM-NaCl.
ability to bind RNA, and should be able to associate with negatively charged phospholipids in membranes, as do positively charged peptides (Kim et al., 1991; Mosior & McLaughlin, 1992). Further studies will explore these and other possibilities.

We are grateful to Drs J. Shaw and E. Rodriguez-Cerezo for gifts of TVMV RNA, anti-P1 \(^{\text{ victorious}}\) antisera and purified P1 \(^{\text{ victorious}}\) protein. We thank Dr J. Carrington for helpful suggestions concerning P1 \(^{\text{ victorious}}\) nomenclature. This work was supported by grants from the University of Kentucky FRHI (4E021) and the USDA (88-37263-3709).

References


(Received 3 November 1992; Accepted 4 January 1993)