Functional Domains of a Geminivirus Replication Protein*

Beverly M. Orozco‡, Ann B. Miller, Sharon B. Settlage, and Linda Hanley-Bowdoin

From the Department of Biochemistry, North Carolina State University, Raleigh, North Carolina 27695-7622

Tomato golden mosaic virus, a member of the geminivirus family, has a single-stranded DNA genome that is replicated and transcribed in infected plant cells through the concerted action of viral and host factors. One viral protein, AL1, contributes to both processes by binding to a directly repeated, double-stranded DNA sequence located in the overlapping (+) strand origin of replication and AL1 promoter. The AL1 protein, which occurs as a multimeric complex in solution, also catalyzes DNA cleavage during initiation of rolling circle replication. To identify the tomato golden mosaic virus AL1 domains that mediate protein oligomerization, DNA binding, and DNA cleavage, a series of truncated AL1 proteins were produced in a baculovirus expression system and assayed for each activity. These experiments localized the AL1 oligomerization domain between amino acids 121 and 181, the DNA binding domain between amino acids 1 and 181, and the DNA cleavage domain between amino acids 1 and 120. Deletion of the first 29 amino acids of AL1 abolished DNA binding and DNA cleavage, demonstrating that an intact N terminus is required for both activities. The observation that the DNA binding domain includes the oligomerization domain suggested that AL1-AL1 protein interaction may be a prerequisite for DNA binding but not for DNA cleavage. The significance of these results for AL1 function during geminivirus replication and transcription is discussed.

Geminiviruses are plant DNA viruses characterized by their single-stranded genomes and their double icosohedral particle morphology (for review, see Ref. 1). They replicate their small genomes through double-stranded DNA intermediates in the nuclei of infected plant cells using a rolling circle mechanism (2–4). These properties make geminiviruses unusual among plant viruses, most of which have RNA genomes and/or replication intermediates. Geminiviruses encode only a few proteins for their replication and depend on the host DNA replication machinery. Thus, geminiviruses represent excellent model systems for studying DNA replication mechanisms in plant cells.

Geminivirus genomes consist of either one or two circular DNA components. Each component contains divergent transcription units separated by a 5′-intergenic region. The intergenic region of all geminiviruses includes a hairpin motif with a conserved AT-rich loop sequence that contains the initiation site for (+) strand DNA replication (3, 5–7). A directly repeated sequence upstream from the hairpin is bound by the viral replication protein AL1 and is required for virus-specific DNA

replication (8) of tomato golden mosaic virus (TGMV)[1] and bean golden mosaic virus (BGMV). Similar motifs are found in a number of related geminivirus genomes (9), but their functional significance is not known. AL1 is the only viral protein required for replication of all geminiviruses (10, 11). (For some geminiviruses, the AL1 homologue is designated C1.) Many geminiviruses also encode a second protein, AL3, which greatly enhances replication (10, 12).

The AL1 protein plays key roles in viral DNA replication and transcription. AL1 confers virus-specific recognition of its cognate origin of replication (8) and initiates (+) strand DNA replication (6, 7). It represses its own expression at the level of transcription (13) and can enhance transcription of late genes of some geminiviruses (14). In addition, AL1 induces expression of a host DNA synthesis protein, proliferating cell nuclear antigen, in nondividing plant cells (15). Multiple biochemical activities have been described for AL1/C1 proteins. The AL1 proteins of TGMV and BGMV bind double-stranded viral DNA in a site- and virus-specific manner (16). Single-stranded DNA binding activity has also been reported for TGMV AL1 (17). The AL1/C1 proteins from tomato yellow leaf curl virus (TYLCV), wheat dwarf virus, and TGMV cleave the (+) DNA strand in the conserved loop sequence of the hairpin (6, 7). Covalent cross-linking of TYLCV C1 to the 5′-end of the cleaved DNA has also been detected (6). In addition, ATP and GTPase activity has been demonstrated for TYLCV C1 (18). Last, multiple protein interactions have been detected for AL1/C1 proteins. TGMV AL1 interacts with itself, AL3, and RRB1, a maize retinoblastoma homologue (19).[2] Interactions between wheat dwarf virus C1 and retinoblastoma proteins from human and maize have also been reported (21–24).

Recent experiments have begun to identify the functional domains of AL1. The first 211 amino acids of TYLCV C1 are sufficient to confer site-specific DNA cleavage *in vitro* (25). This region contains three amino acid motifs that are conserved among all geminivirus AL1/C1 proteins and many rolling circle initiator proteins from other systems (26, 27). The third motif includes a conserved tyrosine residue that functions in the DNA cleavage and joining reaction and mediates covalent linkage to the 5′-end of nicked DNA (28). The C terminus of AL1/C1 contains a fourth conserved motif, a P-loop, that is found in many NTP binding proteins (29). Mutation of a lysine residue in the P-loop of TYLCV C1 reduced or abolished ATPase activity of the protein (18, 30). Mutations in the third DNA cleavage motif and the NTP binding sequence also interfered with geminivirus replication *in vivo* (18, 30).

Genetic experiments using chimeric AL1 proteins mapped virus-specific origin recognition to the N-terminal third of AL1/C1 in TYLCV (31), beet curly top virus (32), TGMV, and

[1] The abbreviations used are: TGMV, tomato golden mosaic virus; BGMV, bean golden mosaic virus; TYLCV, tomato yellow leaf curl virus; GST, glutathione *S*-transferase; ACMV, African cassava mosaic virus.

[2] R. A., Ach, T. Durfee, A. B. Miller, P. Taranto, L. Hanley-Bowdoin, P. C. Zambriski, and W. Gruissem, submitted for publication.

BGMV.[3] However, the chimeric studies with TGMV and BGMV showed that interaction between the N terminus of AL1 and the cognate DNA binding motif is only part of the requirements for virus-specific origin recognition *in vivo*. No biochemical studies have addressed the protein domain involved in AL1-DNA binding. There is also no information regarding the various AL1 protein-protein interaction domains. In this article, we identified the domains of TGMV AL1 that mediate protein oligomerization, DNA binding, and DNA cleavage.

MATERIALS AND METHODS

Plasmid Constructs—Coding sequences corresponding to authentic AL1 and AL1 fused to the glutathione *S*-transferase (GST) domain were cloned into the baculovirus transfer vector pMON27025 (7) for expression in *Spodoptera frugiperda* Sf9 cells. The recombinant proteins were named according to their N- and C-terminal amino acids (Fig. 1). For example, $AL1_{1-120}$ includes AL1 amino acids 1–120. The baculovirus expression vectors encoding GST (pNSB313), GST-$AL1_{1-352}$ (pNSB314), and authentic $AL1_{1-352}$ (pNSB244) have been described previously (7, 33).

Open reading frames for C-terminal truncated AL1 proteins (Fig. 1*A*) were generated by inserting an *Xba*I linker into repaired restriction sites at TGMV A positions 2242 (*Sal*I), 2059 (*Nco*I), and 1963 (*Eag*I) to create in-frame stop codons. Truncated AL1 open reading frames were subcloned as *Bgl*II-*Hin*dIII fragments into *Bam*HI- and *Hin*dIII-digested pMON27025 to give pNSB388 ($AL1_{1-120}$), pNSB517 ($AL1_{1-181}$), and pNSB392 ($AL1_{1-213}$). Plasmid pNSB310 was digested with *Sac*I-*Sma*I and trimmed with T4 DNA polymerase to release a 699-base pair fragment containing the GST coding sequence. Baculovirus expression cassettes corresponding to GST-AL1 fusion proteins (Fig. 1*B*) were generated by cloning this fragment into filled *Nde*I sites of pNSB388, pNSB517, and pNSB392 to give pNSB534 (GST-$AL1_{1-120}$), pNSB547 (GST-$AL1_{1-181}$), and pNSB535 (GST-$AL1_{1-213}$), respectively.

Coding sequences for N-terminal truncated AL1 proteins (Fig. 1*A*) were constructed by inserting an *Sph*I linker into repaired restriction sites at TGMV A positions 2442 (*Sal*I) and 2059 (*Nco*I) to create in-frame start codons. An *Sph*I linker was also inserted into a repaired *Hin*dIII site of pMON27025 to make pNSB448. The truncated AL1 open reading frames were subcloned as *Sph*I-*Eco*RI and *Sph*I-*Bam*HI fragments into the same sites of pNSB448 to give pNSB516 ($AL1_{121-352}$) and pNSB469 ($AL1_{182-352}$), respectively.

Engineered restriction sites and an endogenous *Nco*I site at TGMV A position 2059 were used to create open reading frames for GST-AL1 fusion proteins lacking N-terminal AL1 sequences. The AL1 coding sequence was modified at TGMV A positions 2516–2517 using the primer 5′-GTAATTGAGAAAGTACTTCTTCTTTGGAC to introduce an *Sca*I site and create pNSB162 (34). A *Bst*BI site was also introduced in the AL1 coding sequence by mutating TGMV A position 2404 using the primer 5′-GGCAGCAGTATTTTCCTTCGAACTGAATAAGC to make pNSB428. The *Sca*I-*Bam*HI and *Bst*BI-*Bam*HI fragments from pNSB162 and pNSB428, respectively, were repaired with *Escherichia coli* DNA polymerase (Klenow fragment) and fused in frame with the GST coding region of pNSB314 at an *Sma*I site. The resulting plasmids encoded GST-$AL1_{121-352}$ (pNSB564) and GST-$AL1_{66-352}$ (pNSB563). A trimmed *Sac*I-*Sma*I fragment from pNSB310 containing the GST coding region was inserted into a repaired *Nde*I site of pNSB516 to give pNSB547, encoding GST-$AL1_{182-352}$.

Expression, Purification, and Oligomerization of AL1 Proteins—Recombinant proteins were produced in Sf9 cells using a baculovirus expression system according to published protocols (7, 19). The GST-AL1 fusion proteins were purified by glutathione affinity chromatography (7) and analyzed *in vitro* for various AL1 activities. Aliquots (3 μg) of the purified proteins were fractionated by SDS-polyacrylamide gel electrophoresis and visualized by staining with Coomassie Brilliant Blue dye.

Protein extracts from Sf9 insect cells co-expressing authentic and GST-AL1 fusion proteins were also assayed for AL1 oligomerization by co-purification on glutathione-Sepharose (19). Co-purification was monitored by SDS-polyacrylamide gel electrophoresis followed by transfer to a nitrocellulose membrane (Schleicher & Schuell) and immunoblotting using the ECL detection system (Amersham Corp.). Primary antibodies were rabbit polyclonal anti-GST (Upstate Biotechnology Inc.) and anti-AL1 antisera (19).

The relative molecular masses of full-length AL1 ($AL1_{1-352}$) and the C-terminal truncated proteins $AL1_{1-120}$ and $AL1_{1-181}$ were determined by size exclusion chromatography of extracts from insect cells infected with the corresponding recombinant baculoviruses. Extracts were prepared by mixing cells for 30 min at 4 °C in column buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.15 M NaCl, and 1 mM dithiothreitol) supplemented with protease inhibitors (19). The extracts were clarified by centrifugation for 1 h at 100,000 × *g*. Approximately 0.2 mg of protein (1 mg/mL) was applied to a 50 × 1-cm column of Sepharose CL-6B in column buffer, chromatographed at 0.2 ml/min, and eluted as 0.5-ml fractions. To determine the elution positions of the various AL1 proteins, 75 μl of each fraction was analyzed by SDS-polyacrylamide gel electrophoresis and immunoblotting with anti-AL1 serum as described above. The column was calibrated with protein molecular weight markers (Sigma) individually diluted with column buffer. V_e values of protein standards were determined by monitoring the column effluent at A_{280}. The V_0 was determined from the elution volume of blue dextran. Relative molecular masses of AL1 proteins were estimated from linear regression analysis of V_e/V_0 *versus* the logarithm of the molecular masses of the protein standards.

In Vitro Assays for AL1 Function—DNA electrophoretic mobility shift assays and DNA cleavage assays were performed as described previously (7). For the binding assays, an 83-base pair *Eco*RI fragment containing the AL1-DNA binding motif (TGMV A positions 28–84) was isolated from pNSB378 and 3′-end-labeled using Klenow and [α-^{32}P]dATP. The radiolabeled DNA was incubated with purified GST-AL1 fusion proteins for 1 h at room temperature. DNA and protein concentrations are provided in the figure legends. The bound and free probes were resolved on 1% agarose gels, dried on Whatman DE-81 paper, and analyzed by autoradiography.

For DNA cleavage assays, a single-stranded oligonucleotide (5′-GTT-TAATATTACCGGATGGCCGC) corresponding to the loop and right side of the hairpin structure in the TGMV (+) strand origin was 5′-end-labeled using polynucleotide kinase and [γ-^{32}P]ATP. Approximately 5000 cpm of labeled DNA was incubated with 100 ng of purified GST-AL1 fusion protein in 10 μl of cleavage buffer (25 mM Tris-HCl, pH 7.5, 75 mM NaCl, 5 mM MgCl$_2$, 2.5 mM EDTA, and 2.5 mM dithiothreitol) for 30 min at 37 °C. The reactions were terminated by adding 6 μl of gel loading buffer (95% formamide, 20 mM EDTA, and 0.05% bromphenol blue) and heating to 90 °C for 2 min. The reaction products were resolved on 15% polyacrylamide denaturing gels.

ATPase assays were performed essentially as described by Desbiez *et al.* (18). Approximately 300 ng of GST-AL1 fusion proteins were incubated for 30 min at 37 °C in a buffer containing 25 mM Tris-HCl, pH 7.5, 20 mM NaCl, 2 mM MgCl, 0.01% Triton X-100, 40 μM ATP, and 110 fmol [γ-^{32}P]ATP. Free phosphate was extracted and measured according to the protocol described by Seto-Young and Perlin (35) with the following modifications. The reaction was stopped with 3 volume of 5% ammonium molybdate in 2 N SH_2O_4, and free phosphate was extracted with an equal volume of *N*-butanol. Radioactivity in a 50-μl aliquot was measured by liquid scintillation.

RESULTS

The AL1 Oligomerization Domain—Recent experiments showed that TGMV AL1 interacts with itself to form a multimeric complex (19). To map the AL1 oligomerization domain, we generated a series of truncated AL1 proteins (Fig. 1*A*) and assayed their ability to interact with a full-length AL1 protein fused to glutathione *S*-transferase (GST-$AL1_{1-352}$; Fig. 1*B*). The GST-$AL1_{1-352}$ and truncated AL1 proteins were co-expressed in baculovirus-infected insect cells, and GST-$AL1_{1-352}$ was purified using glutathione-Sepharose resin. Total protein extracts from insect cells (Fig. 2, *lanes 1–7*) and proteins bound to glutathione resin (*lanes 8–14*) were resolved by SDS-polyacrylamide gel electrophoresis and visualized by immunoblotting using antibodies against AL1 and GST. Full-length AL1 (Fig. 2, *lanes 1* and *8*) and the C-terminal truncations $AL1_{1-213}$ (Fig. 2, *lanes 2* and *9*) and $AL1_{1-181}$ (Fig. 2, *lanes 3* and *10*) co-purified with GST-$AL1_{1-352}$. However, further deletion to amino acid 120 in $AL1_{1-120}$ (Fig. 2, *lanes 4* and *11*) abolished complex formation. Similarly, the N-terminal truncation $AL1_{121-352}$ (Fig. 2, *lanes 5* and *12*), co-purified with GST-$AL1_{1-352}$, whereas $AL1_{182-352}$ (Fig. 2, *lanes 6* and *13*) did not interact with GST-$AL1_{1-352}$. $AL1_{1-352}$ did not bind GST alone (Fig. 2,
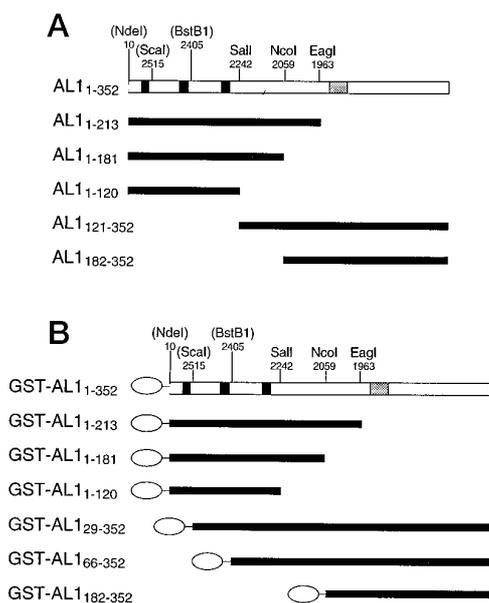
FIG. 1. **Diagram of the AL1 protein constructs.** *A,* schematic of authentic TGMV AL1 protein and truncated AL1 proteins. AL1$_{1-352}$ is shown as an *open box,* with the *solid* and *shaded boxes* marking motifs conserved among rolling circle replication initiator proteins and the ATP binding consensus, respectively. Restriction sites used for cloning are indicated by name and position in the TGMV A genome (20) with engineered sites shown in *parentheses. Solid lines* correspond to truncated AL1 proteins expressed in insect cells. The proteins are designated on the *left* by their terminal amino acids. *B,* schematic of the GST-AL1 fusion proteins. AL1$_{1-352}$ is shown as described for *A* except with GST (*ovals*) fused to the N terminus. The truncated GST-AL1 proteins are represented by *solid lines* and designated on the *left* by their AL1 terminal amino acids.

lanes 7 and *14*), establishing the specificity of AL1-AL1 protein interactions in this experiment. Together, these results located the AL1 oligomerization domain between amino acids 121 and 181.

AL1 oligomerization was further examined by size exclusion chromatography through Sepharose CL-6B. The chromatographic properties of AL1$_{1-352}$ and AL1$_{1-181}$, both of which are predicted to oligomerize, were compared with those of AL1$_{1-120}$, which lacks the oligomerization domain. The column was calibrated with seven globular proteins of known molecular masses from 12.4 to 660 kDa. Linear regression analysis demonstrated a linear relationship between the elution volumes of the protein standards and the logarithms of their reported molecular masses. Immunoblot analysis indicated that AL1$_{1-352}$, AL1$_{1-181}$, and AL1$_{1-120}$ eluted with apparent molecular masses of 318, 156, and 13.7 kDa (Fig. 2*B*), respectively, whereas their predicted monomeric molecular masses are 40.5, 20.8, and 14.0 kDa. Thus, the elution profiles of AL1$_{1-352}$ and AL1$_{1-181}$ demonstrated that they form large protein complexes but that AL1$_{1-120}$ occurs as a monomer in solution, corroborating the glutathione affinity chromatography data (Fig. 2*A*).

The AL1-DNA Binding Domain—AL1 binds specifically to a directly repeated DNA sequence in the 5′-intergenic region of the TGMV genome (8, 33). We used purified GST fusion proteins truncated at AL1 amino acids 213, 181, and 120 (Fig. 1*B*) to map the C-terminal boundary of the AL1-DNA binding domain *in vitro.* GST-AL1$_{1-352}$, the GST-AL1 truncations, or GST alone were expressed in insect cells and purified by binding to glutathione resin. The affinity-purified proteins were pure or highly enriched, as determined by Coomassie Brilliant Blue staining of SDS-polyacrylamide gels (Fig. 3, *lanes 1–5*).

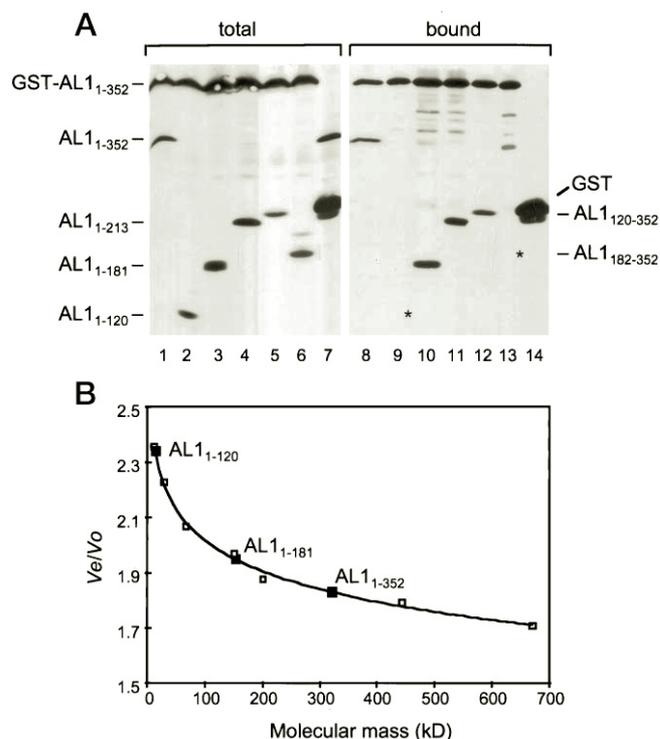The GST-AL1 proteins were tested for DNA binding in elec-

FIG. 2. **Mapping the AL1 oligomerization domain.** *A,* protein extracts from baculovirus-infected cells (*total, lanes 1–7*) were incubated with glutathione resin and eluted in SDS sample buffer (*bound, lanes 8–14*). Total and bound fractions were resolved by SDS-polyacrylamide gel electrophoresis and analyzed by immunoblotting using antibodies against AL1 and GST. GST-AL1$_{1-352}$ was co-expressed with AL1$_{1-352}$ (*lanes 1* and *8*), AL1$_{1-213}$ (*lanes 2* and *9*), AL1$_{1-181}$ (*lanes 3* and *10*), AL1$_{1-120}$ (*lanes 4* and *11*), AL1$_{121-352}$ (*lanes 5* and *12*), and AL1$_{182-352}$ (*lanes 6* and *13*). GST alone was also co-expressed with AL1$_{1-352}$ (*lanes 7* and *14*). The positions of the proteins are shown to the *left* and *right* of the blot and are designated by their N- and C-terminal amino acids. Missing bands in the bound fractions (*lanes 9* and *13*) are indicated by *stars. B,* the elution positions of globular protein standards on a Sepharose CL-6B column were determined by monitoring column effluent at A_{280}. The protein standards and their approximate molecular masses in *parentheses* were thyroglobulin (669 kDa), apoferritin (443 kDa), β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), albumin (66 kDa), carbonic anhydrase (29 kDa), and cytochrome *c* (12.4 kDa). The ratios of their V_e values and the column V_0 were plotted against their reported molecular masses (□). The logarithmic dependence of V_e/V_0 was calculated and is shown as the fitted *line.* Protein extracts from baculovirus-infected insect cells expressing AL1$_{1-352}$, AL1$_{1-181}$, or AL1$_{1-120}$ were chromatographed through the calibrated column, and their V_e/V_0 values were plotted on the standard curve (■).

trophoretic mobility shift assays by titrating in a radiolabeled DNA fragment (TGMV A positions 28–84) that included the AL1 binding site, 5′GGTAGTAAGGTAG. Shifted complexes were observed with full-length GST-AL1$_{1-352}$ (Fig. 4*A*) and with the C-terminal truncated proteins GST-AL1$_{1-213}$ (Fig. 4*B*) and GST-AL1$_{1-181}$ (Fig. 4*C*). In contrast, no DNA binding was detected for GST-AL1$_{1-120}$ (Fig. 4*D*). Together, these results located the C-terminal boundary of the AL1-DNA binding domain between amino acids 121 and 181.

These experiments uncovered two interesting aspects of AL1/DNA binding. First, full-length GST-AL1$_{1-352}$ (Fig. 4*A*) required nearly 4-fold more protein than GST-AL1$_{1-213}$ (Fig. 4*B*) and GST-AL1$_{1-181}$ (Fig. 4*C*) to show equivalent levels of DNA binding. Hence, the presence of AL1 amino acids 214–352 interfered with the efficiency of DNA binding. Second, the protein-DNA complexes formed by the truncated GST-AL1 proteins displayed only slightly increased mobilities in agarose gels (*cf.* Fig. 4, *A–C*) relative to complexes containing full-length GST-AL1$_{1-352}$. The small differences in mobility ob-
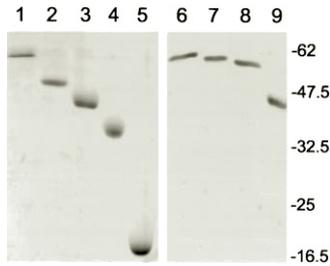
FIG. 3. **Purified GST-AL1 proteins.** Full-length and truncated GST-AL1 proteins were expressed in Sf9 cells and purified using glutathione-Sepharose resin. Eluted proteins (~3 μg) were fractionated on polyacrylamide gels and visualized by Coomassie Brilliant Blue staining. The proteins are GST-AL1$_{1-352}$ (*lanes 1* and *6*), C-terminal truncations GST-AL1$_{1-213}$ (*lane 2*), GST-AL1$_{1-181}$ (*lane 3*), and GST-AL1$_{1-120}$ (*lane 4*), GST alone (*lane 5*), and N-terminal truncations GST-AL1$_{29-352}$ (*lane 7*), GST-AL1$_{65-352}$ (*lane 8*), and GST-AL1$_{182-352}$ (*lane 9*).
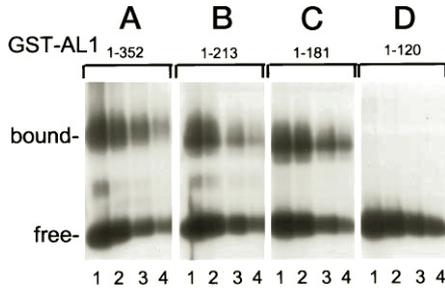

FIG. 4. **DNA binding of full-length and C-terminal truncated GST-AL1 proteins.** GST-AL1 proteins were analyzed by electrophoretic mobility shift assays for binding to a radiolabeled DNA fragment containing the TGMV AL1 binding site. Protein concentrations were constant for each panel. In *lanes 1–4,* 560 ng of GST-AL1$_{1-352}$ (*A*) and GST-AL1$_{1-120}$ (*D*) and 150 ng of GST-AL1$_{1-213}$ (*B*) and GST-AL1$_{1-181}$ (*C*) were used. The amounts of radiolabeled DNA in all of the panels were 0.5 ng (*lane 1*), 0.25 ng (*lane 2*), 0.12 ng (*lane 3*), and 0.06 ng (*lane 4*).

served with agarose gels may reflect the large size of the AL1/DNA complexes demonstrated in Fig. 2*B*.

The N-terminal boundary of the AL1-DNA binding domain was mapped *in vitro* using GST-AL1$_{29-352}$ and GST-AL1$_{66-352}$ (Fig. 1*B*), which lacked the first 28 or 65 amino acids of AL1, respectively. GST-AL1$_{29-352}$ and GST-AL1$_{66-352}$ were each expressed in baculovirus-infected insect cells and purified using glutathione affinity chromatography. The glutathione-eluted proteins were pure or highly enriched, as shown by Coomassie Brilliant Blue staining of SDS-polyacrylamide gels (Fig. 3, *lanes 7* and *8*). GST-AL1$_{1-352}$ (Fig. 3, *lane 6*) was co-electrophoresed for size comparison. The DNA binding activities of purified GST-AL1$_{1-352}$, GST-AL1$_{29-352}$, and GST-AL1$_{66-352}$ were compared using three concentrations of a radiolabeled TGMV A DNA fragment (positions 28–84) containing the AL1 binding site. Although shifted bands were readily detected for GST-AL1$_{1-352}$ at all three probe concentrations (Fig. 5*A*, *lanes 1–3*), no bound DNA was observed for either GST-AL1$_{29-352}$ (Fig. 5*A*, *lanes 4–6*) or GST-AL1$_{66-352}$ (Fig. 5*A*, *lanes 7–9*) under any reaction conditions. These results demonstrated that sequences within the first 28 amino acids of AL1 are essential for protein-DNA interaction.

GST-AL1$_{29-352}$ and GST-AL1$_{66-352}$ were also analyzed for their ability to interact with authentic AL1$_{1-352}$ (Fig. 5*B*) to verify that the truncated proteins were properly folded and functional. The truncated GST-AL1 proteins were co-expressed with authentic AL1 in insect cells, extracted (Fig. 5*B, lanes 1–4*), and bound to glutathione resin (Fig. 5*B, lanes 5–8*). Immunoblot analysis showed that authentic AL1$_{1-352}$ co-purified with both GST-AL1$_{29-352}$ (Fig. 5*B, lanes 2* and *6*) and GST-AL1$_{66-352}$ (Fig. 5*B, lanes 3* and *7*), demonstrating AL1
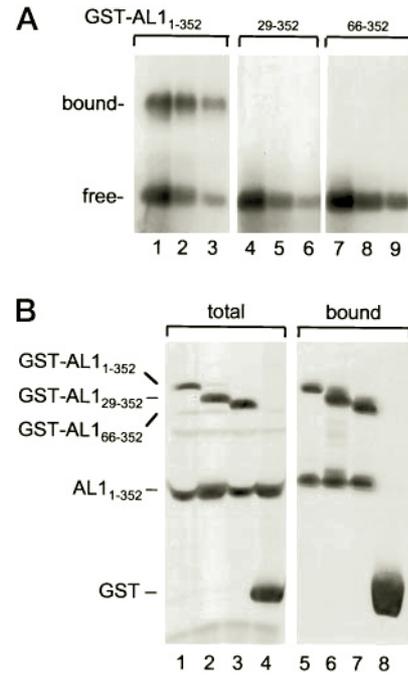

FIG. 5. **Mapping the N-terminal boundary of the AL1-DNA binding domain.** *A,* GST-AL1 proteins were analyzed by electrophoretic mobility shift assays for binding to a radiolabeled DNA fragment containing the TGMV AL1 binding site. Approximately 560 ng of GST-AL1$_{1-352}$ (*lanes 1–3*), GST-AL1$_{29-352}$ (*lanes 4–6*), and GST-AL1$_{65-352}$ (*lanes 7–9*) were incubated with 0.2 ng (*lanes 1, 4,* and *7*), 0.1 ng (*lanes 2, 5,* and *8*), or 0.05 ng (*lanes 3, 6,* and *9*) of radiolabeled DNA. *B*, protein extracts from baculovirus-infected cells (*total, lanes 1–4*) were incubated with glutathione resin and eluted in SDS sample buffer (*bound, lanes 5–8*). Total and bound fractions were resolved by SDS-polyacrylamide gel electrophoresis and analyzed by immunoblotting. GST- AL1$_{1-352}$ (*lanes 1* and *5*), the N-terminal truncations GST-AL1$_{29-352}$ (*lanes 2* and *6*) and GST-AL1$_{65-352}$ (*lanes 3* and *7*), and GST alone (*lanes 4* and *8*) were co-expressed with authentic AL1$_{1-352}$.

oligomerization with the truncated proteins. Co-purification of AL1$_{1-352}$ with GST-AL1$_{1-352}$ (Fig. 5*B, lanes 1* and *5*) and GST (Fig. 5*B, lanes 4* and *8*) were analyzed in parallel as positive and negative controls, respectively, for specificity of interaction. Together, these results established that GST-AL1$_{29-352}$ (Fig. 5*B, lanes 2* and *6*) and GST-AL1$_{66-352}$ are competent for AL1 interaction but deficient for DNA binding. Thus, the functional domain for DNA binding is located between amino acids 1 and 181 and overlaps the AL1 oligomerization domain between amino acids 121 and 181.

The AL1-DNA Cleavage Domain—TGMV AL1 catalyzes site-specific DNA cleavage of the (+) strand origin of replication. The truncated GST-AL1 fusion proteins were used to map the boundaries of the DNA cleavage domain. A radiolabeled, single-stranded oligonucleotide containing the origin nick site (TGMV A positions 129–151) was incubated with full-length GST-AL1$_{1-352}$ and GST fusions lacking AL1 C-terminal amino acids (Fig. 6*A*). The mobility of a marker corresponding to the cleavage product is shown (Fig. 6*A, lane 7*). Cleavage products were detected for GST-AL1$_{1-352}$ (Fig. 6*A, lane 1*), GST-AL1$_{1-213}$ (Fig. 6*A, lane 2*), GST-AL1$_{1-181}$ (Fig. 6*A, lane 3*), and GST-AL1$_{1-120}$ (Fig. 6*A, lane 4*). No cleavage product was seen with GST alone (Fig. 6*A, lane 5*) or in the absence of protein (Fig. 6*A, lane 6*), indicating that the nicking activity was specific to AL1. These results showed that the DNA cleavage domain of TGMV AL1 is in the first 120 amino acids of the protein, and that GST-AL1$_{1-120}$ was folded correctly and functional. Although GST-AL1$_{1-213}$ and GST-AL1$_{1-181}$ formed AL1 oligomers and bound DNA, neither activity was detected for GST-AL1$_{1-120}$ (Figs. 2 and 4). However, because GST-AL1$_{1-120}$ cleaved DNA,
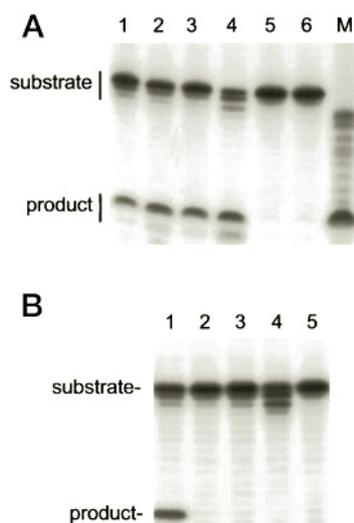
TGMV AL1 Protein Domains

FIG. 6. **Mapping the DNA cleavage domain of TGMV AL1.** GST-AL1 proteins were analyzed in DNA cleavage assays using a radiolabeled oligonucleotide corresponding to the loop and right side of the hairpin in the (+) strand origin. *A,* cleavage results for GST-AL1$_{1-352}$ (*lane 1*), the C-terminal truncations GST-AL1$_{1-213}$ (*lane 2*), GST-AL1$_{1-181}$ (*lane 3*), and GST-AL1$_{1-120}$ (*lane 4*), and GST alone (*lane 5*). *Lane 6,* uncleaved substrate; *lane 7,* radiolabeled marker corresponding to the predicted product (*M*). *B,* cleavage results for GST-AL1$_{1-352}$ (*lane 1*) and the N-terminal truncated proteins GST-AL1$_{29-352}$ (*lane 2*), GST-AL1$_{65-352}$ (*lanes 3*), and GST-AL1$_{121-352}$ (*lane 4*). The bands directly below the substrate in *lane 4* of both panels were due to a contaminating nuclease activity detected in some protein preparations.

the loss of DNA and AL1 interactions can be attributed to deletion of essential amino acids for this activity and not to protein misfolding. The endonucleolytic activity of GST-AL1$_{1-120}$ also demonstrated that AL1 oligomerization is not required for DNA cleavage.

Three amino acid motifs proposed to be part of the DNA cleavage domain are conserved in the N termini of geminivirus AL1/C1 proteins (Fig. 1*A*) and in the initiator proteins from other rolling circle replication systems (26). GST-AL1$_{29-352}$, GST-AL1$_{66-352}$, and GST-AL1$_{182-352}$, which sequentially delete each motif, were assayed for DNA cleavage activity (Fig. 6*B*). GST-AL1$_{1-352}$ specifically cleaved DNA containing the origin nick site (Fig. 6*B, lane 1*), whereas all three truncated proteins were deficient for DNA cleavage activity (Fig. 6*B, lanes 2–4*). Purified GST alone did not cleave DNA (Fig. 6*B, lane 5*), indicating that the product resulted specifically from AL1 activity. These results demonstrated that sequences in the first 28 amino acids of AL1, which contain motif I, are essential for DNA cleavage.

The N-terminal truncation, GST-AL1$_{182-352}$, was deficient for AL1-AL1 interaction and DNA cleavage and did not contain the domain responsible for DNA binding. To verify that GST-AL1$_{182-352}$ was properly folded, we assayed for ATP hydrolysis, because previous studies showed that ATP and GTPase activity is located in the C terminus of TYLCV C1 (18). Equivalent amounts of GST-AL1$_{1-352}$ and GST-AL1$_{182-352}$ hydrolyzed 80 and 50% of the radiolabeled ATP, respectively, whereas GST showed background levels of free phosphate (data not shown). Thus, all of the truncated GST-AL1 proteins possessed at least one of the activities described for AL1.

DISCUSSION

Small DNA viruses with their limited coding capacities frequently specify proteins that have multiple roles during infection. The range of activities and the complexity of multifunctional viral proteins is best exemplified by SV40 large T antigen, which is involved in replication, transcription, and

host induction (36). Recent studies established that the TGMV AL1 protein displays a similar range of functions during geminivirus infection (37, 38). To begin to understand the organization of the AL1 protein and how the different activities are coordinated, we mapped the functional domains for TGMV AL1 oligomerization, DNA binding, and DNA cleavage. Our experiments showed that all three functions are mediated by overlapping domains in the N terminus of the AL1 protein.

We mapped the AL1 oligomerization domain by examining the capacities of truncated proteins to co-purify with GST-AL1$_{1-352}$. In this assay, AL1$_{1-181}$, AL1$_{1-213}$, and AL1$_{120-352}$ interacted with GST-AL1$_{1-352}$. The only region common to all three proteins is from amino acids 121 to 181. Two proteins, AL1$_{1-120}$ and AL1$_{182-352}$, which lacked this sequence, failed to co-fractionate with GST-AL1$_{1-352}$. However, GST fusions of AL1$_{1-120}$ and AL1$_{182-352}$ were active for DNA cleavage and ATP hydrolysis, respectively, indicating that the truncated proteins were properly folded and that the loss of protein interaction was due to deletion of sequences required for AL1 oligomerization. This conclusion was further supported by gel filtration data showing that the apparent and predicted monomeric molecular mass of AL1$_{1-120}$ are equivalent, consistent with it occurring as a single subunit in solution. In contrast, the apparent molecular masses of AL1$_{1-352}$ and AL1$_{1-181}$ complexes are approximately eight times greater than their predicted monomeric masses. However, the precise stoichiometry of the AL1 subunits could not be determined, because the complexes may have included other AL1-interacting proteins in the crude extracts. Together, these data demonstrated that native AL1 is oligomeric and that amino acids 121–181 are required and sufficient for AL1 oligomerization.

Truncated GST-AL1 proteins were used in electrophoretic mobility shift assays to map the TGMV AL1-DNA binding domain. The failure of GST-AL1$_{29-352}$ to bind DNA demonstrated that the first 28 amino acids of AL1 are essential for protein-DNA interactions. The loss of DNA binding activity by GST-AL1$_{1-120}$ but not GST-AL1$_{1-181}$ located the C-terminal boundary of the DNA binding domain between amino acids 121 and 181. These results showed that the functional domain for DNA binding is between AL1 amino acids 1 and 181. Hong and Stanley (39) reported that the first 57 amino acids of the C1 protein of African cassava mosaic virus (ACMV) are sufficient to repress C1 expression in tobacco protoplasts and proposed that the DNA binding domain of ACMV C1 is located in this region (39). In similar studies, we found that deletion of as little as 39 amino acids from the TGMV AL1 C terminus abrogated transcriptional regulation,[4] indicating that the DNA binding domain of TGMV AL1 is not the only requirement for repression *in vivo.* One potential explanation for the observed differences between the TGMV and ACMV proteins may be that the putative ACMV C1 binding site does not contain directly repeated motifs such as those found in the TGMV AL1 binding site. Thus, TGMV AL1 and ACMV C1 may contact their respective promoters differently and may repress transcription through different mechanisms.

AL1 recognition of the (+) strand origin is essential for virus-specific DNA replication (8). Chimeric virus studies showed that the N-terminal third of C1 confers virus-specific replication to closely related strains of TYLCV (31) or beet curly top virus (32). Replication studies using chimeric origins and AL1 expression cassettes established that amino acids 1–116 of TGMV and BGMV AL1 specifically recognize the repeated DNA binding motifs in their respective (+) strand origins *in vivo.*[3] In contrast, we showed that AL1 amino acids

---

[4] P. A. Eagle and L. Hanley-Bowdoin, unpublished results.

1–181 are necessary for DNA binding *in vitro*. The additional sequences between amino acids 121 and 181 required for *in vitro* DNA binding may contribute essential DNA contacts that are conserved between TGMV and BGMV. Alternatively, AL1 oligomerization, which has been mapped to amino acids 121–181, may be a prerequisite for AL1-DNA binding. Chimeric studies only reveal amino acid differences involved in AL1-DNA interactions and, therefore, cannot distinguish between these possibilities.

DNA binding proteins frequently interact with DNA as dimeric or multimeric complexes (40, 41). Two observations support the idea that TGMV AL1 binds DNA as a multimer. First, the TGMV AL1 binding site contains a repeated motif, such that two AL1 subunits could interact simultaneously with the site. Protein dimer interactions with directly repeated sequences have been described for α-2 protein (42) and HAP1 (43). Second, electrophoretic mobility shift assays suggested that AL1 binds DNA as a large multimeric complex, with AL1-DNA complexes failing to enter polyacrylamide gels and only being resolved on agarose gels. Binding experiments using circularly permuted DNA fragments indicated that it is unlikely that the low electrophoretic mobility of the AL1-DNA complex is due to unusual DNA structure or bending.[5] Several assays failed to determine the stoichiometry of the AL1-DNA complexes. Electrophoretic mobility shift assays with full-length and truncated AL1 proteins bound to DNA did not distinguish heterodimer formation (data not shown). In addition, fusion to GST, which dimerizes with itself (44), or addition of AL1 antibodies did not restore DNA binding activity to $AL1_{1-120}$.[5] Based on these results, we think that it is unlikely that fusion to a heterologous protein interaction domain will restore DNA binding activity to $AL1_{1-120}$ and that a different strategy will be necessary to address the relationship between DNA binding and oligomerization.

AL1 initiates rolling circle replication by introducing a nick into the (+) strand origin of the viral DNA. Heyraud-Nitschke *et al.* (25) showed that the first 211 amino acids of the TYLCV C1 protein possess DNA cleavage activity. Our results showed that the first 120 amino acids of TGMV AL1 specifically cleaves single-stranded DNA containing the (+) strand origin *in vitro* and that AL1 oligomerization and DNA binding were not prerequisites for cleavage of a single-stranded DNA *in vitro*. However, AL1-DNA binding may be required for cleavage of the double-stranded viral genome during rolling circle replication *in vivo*. Three motifs in the N termini of all geminivirus AL1/C1 proteins are also conserved among initiator proteins from other rolling circle systems (26, 27). Motif I (FLTY) is located between amino acids 16 and 19. Motif II (HLH) is a putative metal binding site consisting of two histidines within a region of bulky hydrophobic residues. Motif III includes a highly conserved tyrosine residue that is required for DNA cleavage and ligation by TYLCV C1 (28). Although the role of motif I is unknown, deletion of the N-terminal 28 amino acids of TGMV AL1 abolished DNA cleavage activity, suggesting that this conserved element may be essential for DNA cleavage. The loss of DNA cleavage activity by $GST-AL1_{29-352}$ precluded any conclusions about motifs II or III based on other N-terminal truncations of AL1.

The AL1/C1 protein sequences from 17 dicot-infecting geminiviruses were compared using the EMBL Predict program (45) to determine whether the N terminus of AL1 contains any conserved structural motifs that might contribute to DNA binding, DNA cleavage, or oligomerization. This analysis revealed two sets of α-helices that are predicted with greater than 80%
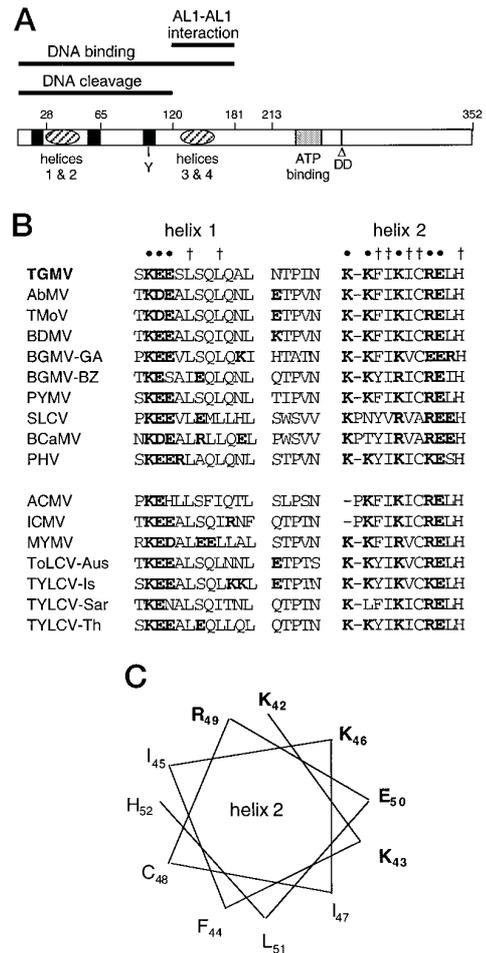

FIG. 7. **Schematic of the AL1 protein domains and predicted structural motifs.** *A,* schematic of the TGMV AL1 protein. *Solid boxes* mark the location of the three motifs conserved among replication initiator protein; the *shaded box* shows the location of the ATP binding domain; and the *hatched ovals* indicate the predicted sets of α-helices. *Solid lines* above the protein mark the location of the functional domains for AL1-AL1 interaction, DNA binding, and DNA cleavage. *B,* sequence alignment for 17 geminivirus AL1/C1 proteins shows two predicted helices (*helix 1* and *helix 2*) separated by 5 amino acids. Conserved charged residues (●) and aliphatic residues (†) are indicated above the amino acid sequence. All charged residues are shown in **bold**. The sequences are grouped according to new world (*top*) and old world (*bottom*) geminiviruses. The AL1 sequences were obtained from GenBank using the following accession numbers: TGMV, K02029; abutilon mosaic virus (*AbMV*), X15983; tomato mottle virus (*TMoV*), L14640; bean dwarf mosaic virus (*BDMV*), M88179; BGMV, Guatamalan isolate (*BGMV-GA*), M91604; BGMV, Brazilian isolate (*BGMV-BZ*), M88686; potato yellow mosaic virus (*PYMV*), D00940; squash leaf curl virus (*SLCV*), M38183; bean calico mosaic virus (*BCaMV*), L27264; pepper huasteco virus (*PHV*), X70418; ACMV, X17095; Indian cassava mosaic virus (*ICMV*), Z24758; mung bean yellow mosaic virus (*MYMV*), D14703; tomato leaf curl virus, Australian isolate (*ToLCV-Aus*), S53251; TYLCV, Isreali isolate (*TYLCV-Is*), X15656; TYLCV, Sardinian isolate (*TYLCV-Sar*), X61153; and TYLCV, Thailand isolate (*TYLCV-Th*), M59838. *C,* helical wheel analysis of the TGMV AL1 helix 2 amino acid sequence. The helix is amphipathic with conserved charged residues clustered on one side (shown in *bold*).

probability (Fig. 7*A*). Helices 1 and 2 are between TGMV AL1 amino acids 25–52 in the overlapping DNA binding and cleavage domains and might be involved in these activities. The sequences of both helices show a high degree of homology among different geminiviruses (Fig. 7*B*), especially helix 2, which is conserved at 9 of 11 positions and displays a strong amphipathic character (Fig. 7*C*). Most known DNA binding motifs include α-helical regions that recognize and contact DNA (46). However, the AL1 N terminus shows no obvious

homology to the α-helical motifs of basic/helix-loop-helix, ets, homeodomain, zinc finger, or basic/leucine zipper proteins (for review, see Refs. 46–48). AL1 helices 1 and 2, which are separated by a 5-amino acid loop, most resemble the helix-turn-helix motif, but our sequence comparison failed to uncover a nearby third helix characteristic of most helix-turn-helix DNA binding domains (49). The second set of predicted α-helices is located between TGMV AL1 amino acids 131 and 152 in the overlapping AL1-DNA binding and oligomerization domains (Fig. 7A). Several classes of DNA binding proteins, including members of the basic/helix-loop-helix, homeodomain, and basic/leucine zipper families, use α-helices for dimerization as well as DNA contacts (46, 50, 51). The significance of these predicted structures in DNA binding and/or protein interactions is being investigated.

The oligomerization, DNA binding, and DNA cleavage domains are located to the N-terminal half of AL1, whereas very little is known about the C terminus of the protein. To date, the only biochemical activity that has been attributed to the AL1 C terminus is ATP and GTP hydrolysis (18). We found that deletion of only 39 amino acids from the TGMV AL1 C terminus abolished DNA replication and repression *in vivo*,[5] further demonstrating the functional importance of this region. We also observed that deletion of the C-terminal 139 amino acids of AL1 enhanced DNA binding activity approximately 4-fold *in vitro*, suggesting that a negative effector of DNA binding may be located in this region. Many transcription factors contain regions that inhibit their DNA binding activity unless complexed with other proteins or co-factors (43, 52, 53). The AL1 C terminus may also mediate interaction with other proteins, single-stranded DNA binding, nuclear localization, and/or attachment to the nuclear matrix. We are continuing to map the functional domains of TGMV AL1 to gain a more complete understanding of AL1 structure and function. We are also constructing a series of site-directed mutations in the N terminus of TGMV AL1 to address the functional significance of the predicted α-helices and the relationship between DNA binding and protein oligomerization.

REFERENCES

1. Lazarowitz, S. (1992) *Crit. Rev. Plant Sci.* **11,** 327–349
2. Stenger, D. C., Revington, G. N., Stevenson, M. C., and Bisaro, D. M. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 8029–8033
3. Stanley, J. (1995) *Virology* **206,** 707–712
4. Heyraud, F., Matzeit, V., Schaefer, S., Schell, J., and Gronenborn, B. (1993) *Biochimie* (*Paris*) **75,** 605–615
5. Heyraud, F., Matzeit, V., Kammann, M., Schaefer, S., Schell, J., and Gronenborn, B. (1993) *EMBO J.* **12,** 4445–4452
6. Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S. G., Schell, J., and Gronenborn, B. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 3879–3883
7. Orozco, B. M., and Hanley-Bowdoin, L. (1996) *J. Virol.* **270,** 148–158
8. Fontes, E. P. B., Gladfelter, H. J., Schaffer, R. L., Petty, I. T. D., and Hanley-Bowdoin, L. (1994) *Plant Cell* **6,** 405–416
9. Arguello-Astorga, G. R., Guevara-Gonzalez, R. G., Herrera-Estrella, L. R., and Rivera-Bustamante, R. F. (1994) *Virology* **203,** 90–100
10. Elmer, J. S., Sunter, G., Gardiner, W. E., Brand, L., Browning, C. K., Bisaro, D. M., and Rogers, S. G. (1988) *Plant Mol. Biol.* **10,** 225–234
11. Hanley-Bowdoin, L., Elmer, J. S., and Rogers, S. G. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 1446–1450
12. Sunter, G., Hartitz, M. D., Hormuzdi, S. G., Brough, C. L., and Bisaro, D. M. (1990) *Virology* **179,** 69–77
13. Eagle, P. A., Orozco, B. M., and Hanley-Bowdoin, L. (1994) *Plant Cell* **6,** 1157–1170
14. Hofer, J. M. I., Dekker, E. L., Reynolds, H. V., Woolston, C. J., Cox, B. S., and Mullineaux, P. M. (1992) *Plant Cell* **4,** 213–223
15. Nagar, S., Pedersen, T. J., Carrick, K., Hanley-Bowdoin, L., and Robertson, D. (1995) *Plant Cell* **7,** 705–719
16. Fontes, E. P. B., Luckow, V. A., and Hanley-Bowdoin, L. (1992) *Plant Cell* **4,** 597–608
17. Thommes, P., Osman, T. A. M., Hayes, R. J., and Buck, K. W. (1993) *FEBS Lett.* **319,** 95–99
18. Desbiez, C., David, C., Mettouchi, A., Laufs, J., and Gronenborn, B. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 5640–5644
19. Settlage, S. B., Miller, B., and Hanley-Bowdoin, L. (1996) *J Virol.* **70,** 6790–6795
20. Hamilton, W. D. O., Stein, V. E., Coutts, R. H. A., and Buck, K. W. (1984) *EMBO J.* **3,** 2197–2205
21. Xie, Q., Suarezlopez, P., and Gutierrez, C. (1995) *EMBO J.* **14,** 4073–4082
22. Collin, S., Fernandezlobato, M., Gooding, P. S., Mullineaux, P. M., and Fenoll, C. (1996) *Virology* **219,** 324–329
23. Grafi, G., Burnett, R. J., Helentjaris, T., Larkins, B. A., DeCaprio, J. A., Sellers, W. R., and Kaelin, W. G., Jr. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 8962–8967
24. Xie, Q., Sanzburgos, P., Hannon, G. J., and Gutierrez, C. (1996) *EMBO J.* **15,** 4900–4908
25. Heyraud-Nitschke, F., Schumacher, S., Laufs, J., Schaefer, S., Schell, J., and Gronenborn, B. (1995) *Nucleic Acids Res.* **23,** 910–916
26. Koonin, E. V., and Ilyina, T. V. (1992) *J. Gen. Virol.* **73,** 2763–2766
27. Ilyina, T. V., and Koonin, E. V. (1992) *Nucleic Acids Res.* **20,** 3279–3285
28. Laufs, J., Schumacher, S., Geisler, N., Jupin, I., and Gronenborn, B. (1995) *FEBS Lett.* **377,** 258–262
29. Gorbalenya, A. E., and Koonin, E. V. (1989) *Nucleic Acids Res.* **17,** 8413–8440
30. Hanson, S. F., Hoogstraten, R. A., Ahlquist, P., Gilbertson, R. L., Russell, D. R., and Maxwell, D. P. (1995) *Virology* **211,** 1–9
31. Jupin, I., Hericourt, F., Benz, B., and Gronenborn, B. (1995) *FEBS Lett.* **362,** 116–120
32. Choi, I. R., and Stenger, D. C. (1995) *Virology* **206,** 904–912
33. Fontes, E. P. B., Eagle, P. A., Sipe, P. A., Luckow, V. A., and Hanley-Bowdoin, L. (1994) *J. Biol. Chem.* **269,** 8459–8465
34. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 488–492
35. Seto-Young, D., and Perlin, D. S. (1991) *J. Biol. Chem.* **266,** 1383–1389
36. Fanning, E., and Knippers, R. (1992) *Annu. Rev. Biochem.* **61,** 55–85
37. Laufs, J., Jupin, I., David, C., Schumacher, S., Heyraud-Nitschke, F., and Gronenborn, B. (1995) *Biochimie* (*Paris*) **77,** 765–773
38. Hanley-Bowdoin, L., Eagle, P. A., Orozco, B. M., Robertson, D., and Settlage, S. B. (1996) in *Biology of Plant-Microbe Interactions* (Stacey, G., Mullin, B., and Gresshoff, P. M., eds) pp. 287–292, International Society of Molecular and Plant-Microbe Interactions, St. Paul, MN
39. Hong, Y. G., and Stanley, J. (1995) *J. Gen. Virol.* **76,** 2415–2422
40. Sun, X.-H., and Baltimore, D. (1991) *Cell* **64,** 459–470
41. Voronova, A., and Baltimore, D. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 4722–4726
42. Smith, D. L., and Johnson, A. D. (1992) *Cell* **68,** 133–142
43. Zhang, L., Bermingham-McDonogh, O., Turcotte, B., and Guarente, L. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 2851–2855
44. Mannervik, B., and Danielson, U. H. (1988) *CRC Crit. Rev. Biochem.* **23,** 283–337
45. Rost, B., and Sander, C. (1993) *J. Mol. Biol.* **232,** 585–599
46. Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem.* **61,** 1053–1095
47. Harrison, S. C. (1991) *Nature* **353,** 715–719
48. Nelson, H. C. M. (1995) *Cur. Opin. Genet. & Dev.* **5,** 180–189
49. Wintjens, R., and Rooman, M. (1996) *J. Mol. Biol.* **262,** 294–313
50. Murre, C., Schonleber, P., and Baltimore, D. (1989) *Cell* **56,** 777–783
51. Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubotokunaga, K., Schier, A. F., Resendezperez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994) *Cell* **78,** 211–223
52. Janknecht, R., and Nordheim, A. (1993) *Biochim. Biophys. Acta* **1155,** 346–356
53. Bardwell, V. J., and Treisman, R. (1994) *Genes Dev.* **8,** 1664–1677