Nucleotide Sequence and Structural Determinants of Specific Binding of Coat Protein or Coat Protein Peptides to the 3' Untranslated Region of Alfalfa Mosaic Virus RNA 4

FELICIA HOUSER-SCOTT, MARGARET L. BAER,[†] KAREL F. LIEM, JR.,[‡] JUN-MING CAI,
AND LEE GEHRKE*

*Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139,
and Program in Cell and Developmental Biology, Harvard Medical School, Boston, Massachusetts 02115*

Received 8 October 1993/Accepted 21 December 1993

The specific binding of alfalfa mosaic virus coat protein to viral RNA requires determinants in the 3' untranslated region (UTR). Coat protein and peptide binding sites in the 3' UTR of alfalfa mosaic virus RNA 4 have been analyzed by hydroxyl radical footprinting, deletion mapping, and site-directed mutagenesis experiments. The 3' UTR has several stable hairpins that are flanked by single-stranded (A/U)UGC sequences. Hydroxyl radical footprinting data show that five sites in the 3' UTR of alfalfa mosaic virus RNA 4 are protected by coat protein, and four of the five protected regions contain AUGC or UUGC. Electrophoretic mobility band shift results suggest four coat protein binding sites in the 3' UTR. A 3'-terminal 39-nucleotide RNA fragment containing four AUGC repeats bound coat protein and coat protein peptides with high affinity; however, coat protein bound poorly to antisense 3' UTR transcripts and poly(AUGC)₁₀. Site-directed mutagenesis of AUGC₈₆₅₋₈₆₈ resulted in a loss of coat protein binding and peptide binding by the RNA fragment. Alignment of alfalfa mosaic RNA sequences with those from several closely related ilarviruses demonstrates that AUGC₈₆₅₋₈₆₈ is perfectly conserved; moreover, the RNAs are predicted to form similar 3'-terminal secondary structures. The data strongly suggest that alfalfa mosaic virus coat protein and ilarvirus coat proteins recognize invariant AUGC sequences in the context of conserved structural elements.

Although it is clear that RNA-protein interactions are essential for protein synthesis (25), RNA processing (44), RNA localization (60), and other biological processes, our understanding of ribonucleoprotein particle assembly, structure, and functional significance is in an early stage of development. The best examples of high-resolution structural data for RNA-protein complexes are limited to two X-ray crystallographic analyses of tRNA-aminoacyl synthetase complexes (54, 55). These results and other studies demonstrate that RNA conformation can be altered significantly upon specific interaction with peptides or protein (4, 15, 33, 49, 63). Several classes of RNA-binding proteins have been identified by amino acid homology, and in some cases, the conserved domains are known to be important for binding RNA (41). One challenge is, however, to move beyond characterizing binding determinants to understand how the RNA-protein complexes accomplish their biological functions.

Alfalfa mosaic virus (AIMV) coat protein is required not only for virus assembly but also for virus replication (reviewed in reference 32). Coat protein binds to homologous sequences located in the 3' untranslated regions (UTRs) of the three genomic RNAs (numbered 1 to 3) (4, 15, 33, 49, 63). Several classes of RNA-binding proteins have been identified by amino acid homology, and in some cases, the conserved domains are known to be important for binding RNA (41). One challenge is, however, to move beyond characterizing binding determinants to understand how the RNA-protein complexes accomplish their biological functions.

Although it is clear that RNA-protein interactions are essential for protein synthesis (25), RNA processing (44), RNA localization (60), and other biological processes, our understanding of ribonucleoprotein particle assembly, structure, and functional significance is in an early stage of development. The best examples of high-resolution structural data for RNA-protein complexes are limited to two X-ray crystallographic analyses of tRNA-aminoacyl synthetase complexes (54, 55). These results and other studies demonstrate that RNA conformation can be altered significantly upon specific interaction with peptides or protein (4, 15, 33, 49, 63). Several classes of RNA-binding proteins have been identified by amino acid homology, and in some cases, the conserved domains are known to be important for binding RNA (41). One challenge is, however, to move beyond characterizing binding determinants to understand how the RNA-protein complexes accomplish their biological functions.

Coat protein binding sites on AIMV RNAs have been studied previously by isolating RNase T₁-protected RNA fragments that rebind coat protein after nitrocellulose filter selection (76-78). Common features of the bound fragments were the presence of two tetranucleotide AUGC sequences and potential to form a stable secondary structure (77). Koper-Zwartoff and Bol (35) proposed that the RNA determinant for coat protein binding consists of a stable hairpin flanked at its 3' end by the sequence AUGC.

In this study, we have used hydroxyl radical footprinting, deletion mapping, and site-directed mutagenesis to identify nucleotides required for coat protein binding. Binding sites were identified by testing the interaction of coat protein or coat protein peptides with (i) positive- and negative-strand transcripts of the 170-nucleotide 3' UTR, (ii) 5' and 3' half-UTR molecules, (iii) both a 26- and a 39-nucleotide truncated binding site, and (iv) site-directed AUGC→AAAA mutant RNAs. The footprinting data provide evidence for five coat protein interaction sites in the 3' UTR, while four specific RNA-protein complexes are identified by band shift analysis. A 3'-terminal 39-nucleotide minimal binding site fragment containing four AUGC repeats binds both full-length coat protein and N-terminal coat protein peptides with high affinity. However, a 26-nucleotide RNA comprising a predicted stable hairpin flanked by 5' and 3' AUGC sequence binds coat

* Corresponding author. Mailing address: HST Division, M.I.T. Building E25-545, Cambridge, MA 02139. Phone: (617) 253-7608. Fax: (617) 253-6693. Electronic mail address: lgehrke@mit.edu.

[†] Present address: University of California, San Francisco, San Francisco, CA 94143.

[‡] Present address: College of Physicians and Surgeons of Columbia University, New York, NY 10032.

protein poorly, as does the antisense 3' UTR transcript. Alignment of 3'-terminal nucleotide sequences of AIMV RNAs and RNAs from several related ilarviruses revealed a perfectly conserved AUGC that separates two predicted hairpin structures. Mutations in the conserved sequence element prevent coat protein binding. Comparative sequence analysis reveals covarying nucleotides that maintain hairpins. The results indicate that high-affinity binding of coat protein or peptides requires a combination of sequence and structural determinants located at the 3' end of the viral RNAs.

MATERIALS AND METHODS

Footprint analysis. Protocols based on either hydrogen peroxide (8) or molecular oxygen (38) were used for the hydroxyl radical footprinting experiments. RNAs were renatured prior to all treatments by heating at 65°C in REN buffer (10 mM Tris-HCl [pH 7.4], 3 mM MgCl₂, 50 mM NaCl, 0.1 mM EDTA) for 2 min and then slow cooling to room temperature over a 15-min period. A 10× stock of Fe(II)-EDTA cleavage reagent was prepared by mixing 10 µl of fresh 100 mM ammonium iron(II) sulfate hexahydrate (Aldrich) and 20 µl of 100 mM EDTA in a final volume of 100 µl. One picomole of renatured end-labeled RNA (10,000 to 100,000 cpm) was first incubated for 10 min in 10 µl of morpholinopropanesulfonic acid (MOPS) buffer (20 mM MOPS-HCl [pH 7.5], 1 mM EDTA, 55 pmol of tRNA) in the presence or absence of 5 pmol of coat protein. Next, 0.1 volume of the 10× stock Fe(II)-EDTA cleavage solution was added, and the samples were processed through either the molecular oxygen or the hydrogen peroxide procedure (see below).

For the molecular oxygen protocol, 0.1 volume of 50 mM dithiothreitol (DTT) was added, and RNA was cleaved during a 2-h incubation at room temperature. The reaction was quenched by the addition of thiourea to a final concentration of 10 mM. The solution was then extracted twice with phenol-chloroform (1:1), and the RNA was precipitated by adjusting the solution to 2.5 M ammonium acetate and then by adding 3 volumes of absolute ethanol. Tris-based buffers were avoided in all solutions because Tris scavenges free radicals and reduces cleavage (11).

For the hydrogen peroxide method, 0.1 volume of 100 mM ascorbate and 0.1 volume of 0.6% H₂O₂ were added to the RNA or RNA-protein complexes in MOPS buffer. Reaction mixtures were incubated for 10 to 15 min at room temperature prior to quenching with thiourea added to a final concentration of 10 mM. The RNA was deproteinized by phenol extraction and precipitated as described above. Precipitated RNAs were washed twice with 70% ethanol, resuspended in gel loading buffer (3.5 M urea in 90 mM Tris–90 mM boric acid–2 mM EDTA [pH 8] [TBE]), and then analyzed by electrophoresis into a 6 or 8% polyacrylamide, 0.4-mm-thick gel containing 8.3 M urea in TBE buffer. Nucleotide numbering was assigned by comparison of the hydroxyl radical cleavage patterns with an RNase T₁ ladder that was coelectrophoresed in an adjacent lane.

RNase T₁ cleavage was carried out under three conditions: RNA alone (native conditions), RNA alone (denaturing conditions), and RNA plus coat protein. A T₁ ladder (denaturing conditions) was generated by digesting approximately 10,000 to 50,000 cpm of end-labeled RNA by using 0.1 to 0.2 U of enzyme in 5 µl of denaturing buffer (3.5 M urea, 16 mM sodium citrate [pH 5.0], 0.8 mM EDTA, 0.5 µg of tRNA per µl) for 12 min at 55°C. RNase T₁ hydrolysis under native conditions was performed by incubating labeled RNA in 10 µl of MOPS buffer (20 mM MOPS-HCl [pH 7.5], 1 mM EDTA)

FIG. 1. Secondary structure folding of the 3' UTR of AIMV RNA 4. The structure was proposed by Koper-Zwartoff and Bol (35) and Houwing and Jaspars (29) and tested by experimental enzymatic structure mapping (52). The termination codon for protein synthesis is located at position 700 of the AIMV RNA 4 nucleotide sequence. The lowercase letters indicate vector-derived nucleotides. Repeated (A/U)UGC sequences are underlined.

containing 55 pmol of tRNA and 0.1 to 0.2 U of RNase. RNase T₁ footprinting was conducted by adding the enzyme after preincubation of the RNA with coat protein as described for the hydroxyl radical footprinting.

Solutions containing RNAs digested with RNase T₁ under native conditions were adjusted to 3.5 M urea and loaded directly onto polyacrylamide gels. Phenol-extracted and precipitated/washed RNAs were resuspended in gel loading buffer (3.5 M urea, 1× TBE) and then analyzed by electrophoresis into a 6 or 8% polyacrylamide, 0.4-mm-thick gel containing 8.3 M urea in TBE buffer. Nucleotide numbering was assigned by comparing cleavage patterns with an RNase T₁ ladder that was coelectrophoresed in an adjacent lane.

RNA constructs and in vitro transcription. Plasmid pSP65A4, representing the complete 881-bp AIMV RNA 4 cDNA, was described previously (39). A 170-bp 3'-terminal fragment of the AIMV RNA 4 cDNA that contains high-affinity coat protein binding sites (4) was subcloned as a *Bst*XI-*Sma*I fragment into a transcription vector to permit synthesis of AIMV_{718–881} (Fig. 1). The 5' half-UTR molecule AIMV_{718–797} was generated by cleaving the AIMV_{718–881} construct DNA with restriction enzyme *Sau*3A (which cleaves at position 797) prior to in vitro transcription. The 3' half-molecule clone was generated by subcloning nucleotides 798 to 881 as a *Sau*3A-*Sma*I fragment into a transcription vector. RNAs were transcribed in vitro by using bacteriophage SP6 or T7 RNA polymerase (Gibco-BRL Life Sciences or Promega) according to the manufacturer's suggested protocols. Radiolabeled RNAs for use in electrophoretic mobility shift assays (EMSA) were generated in transcription reactions supplemented with [³⁵S]GTP (Dupont/New England Nuclear). The short 3'-terminal fragments of AIMV RNA 4 were synthesized by in vitro transcription using synthetic DNA templates (42). The transcription initiation site for the bacteriophage RNA polymerases in most constructs was GG in order to promote efficient transcription. All RNAs were purified by electrophoresis into a polyacrylamide-urea gel, followed by localization of the RNA band by UV light shadowing and buffer elution (9) of the RNA by diffusion overnight at 4°C.

Virion coat protein and peptides. Purified virion coat protein (a gift from Edward Halk) was prepared from virions essentially according to the protocol described by Kruseman et al. (37). The isolated form of the protein is a stable dimer (37). Peptides representing the amino-terminal 25 or 38 amino acids of AIMV coat protein were synthesized by using *t*-butyloxycarbonyl (T-Boc) chemistry on an Applied Biosystems model 430A solid-phase peptide synthesizer. High-pressure liquid

chromatography was used to purify the peptides. The N terminus of virion coat protein is acetyl-Ser (67); therefore, peptides were synthesized with the corresponding *N*-acetyl group. The sequence of peptide CP38 is *N*-acetyl-SSSKKAG GKAGKPTKRSQNYAALRKAQLPKPPALKVP, while that of CP25 is *N*-acetyl-SSSKKAGGKAGKPTKRSQNYAALR.

EMSA. EMSA conditions were adapted from Weinberger et al. (73) as described elsewhere (4). Prior to use, all AIMV RNAs were heated to 65°C for 2 min in REN buffer (50 mM NaCl, 10 mM MgCl₂, 1 mM EDTA, 10 mM Tris-HCl [pH 7.5]) to dissociate aggregates (61) and then cooled slowly to room temperature over approximately 15 min to permit renaturation. One picomole of radiolabeled, gel-purified RNA and 0 to 100 pmol of purified AIMV coat protein were incubated in a total volume of 10 µl of binding buffer (10 mM Tris-HCl [pH 7.5], 50 mM NaCl, 1 mM DTT, 1 mM EDTA, 5% glycerol, 1 U of RNasin per µl, 55 pmol of tRNA) at room temperature for 10 min. The molar concentration of coat protein was calculated on the basis of its isolated form as a dimer (37). When unlabeled competitor RNA was used, labeled and unlabeled RNAs were mixed before the addition of coat protein. The tracking dyes xylene cyanol FF and bromophenol blue were added, and the mixtures were analyzed by electrophoresis into a 7.5% polyacrylamide gel (acrylamide/bisacrylamide ratio of 46:1) in 0.5× TBE (44.5 mM Tris, 44.5 mM boric acid, 1.25 mM EDTA [pH 8.3]) at a constant power of 1 W (7 cm by 10 cm by 0.75 mm in a Bio-Rad Mini-PROTEAN II electrophoresis cell). The gels were prerun at the same power for approximately 1 h prior to loading of the samples. Dried gels were exposed to X-ray film (Kodak) or phosphor-imaging screens.

RESULTS

The experimental analysis of secondary structure in the 3' UTR of AIMV RNA 4 was reported previously (52), and a summary of the structure predicted by free-energy calculations coupled with experimental data obtained by using the computer program RNAMATRIX is shown in Fig. 1. The 3' UTR is characterized by a number of stable stem-loop structures separated by short stretches of single-stranded nucleotides. The tetranucleotide (A/U)UGC is found seven times within the 170-nucleotide UTR (36, 46, 76).

Hydroxyl radical and RNase T₁ footprinting analyses were used to localize coat protein interaction sites on the 3' terminus of AIMV RNA 4 (Fig. 2A). Advantages of Fe(II)-EDTA include the fact that the hydroxyl radical ($\cdot$ OH) group is very small and therefore offers high-resolution structural information; moreover, the reagent has little if any base specificity and cleaves double-stranded regions as well as single-stranded regions (12). Base-paired regions in the RNA become inaccessible to the T₁ enzyme following renaturation (Fig. 2A; compare lanes 2 and 3), and many T₁ sites are protected by coat protein (compare lane 3 with lanes 4 and 5). Hydroxyl radical footprinting analysis is presented in lanes 6 to 12; lanes 6, 9, and 12 represent RNA cleaved in the absence of coat protein, while lanes 7, 8, 10, and 11 show the cleavage pattern of RNA plus coat protein. The RNase and hydroxyl radical footprinting patterns are complementary; however, the data illustrate the superior resolution of the Fe(II)-EDTA reagent. For example, the absence of labeled RNA fragments in the 793–875 region of the T₁ footprinting data (lanes 4 and 5) suggests that the entire 3' end of the molecule is protected. However, the hydroxyl radical footprinting data presented in lanes 6 to 12 suggest three discrete protected RNA domains in the 793–875 region. In all, coat protein provides partial pro-

FIG. 2. Hydroxyl radical and RNase T₁ footprinting of AIMV coat protein bound to the 3' terminus of AIMV RNA 4. Procedural details are found in Materials and Methods. (A) Lane 1, RNA only, without treatment; lane 2, RNase T₁ digestion using denaturing (urea) conditions; lane 3, RNase T₁ digestion, native conditions; lanes 4 and 5, RNase T₁ footprints. RNA was incubated in the presence of a fivefold molar excess of coat protein and then digested with 0.5 (lane 4) or 1.0 (lane 5) U of RNase T₁. Lane 6, RNA only, hydrolyzed by the peroxide method using 0.2 volume of (Fe)II-EDTA plus 0.2 volume of peroxide-ascorbate solutions. The cleavage reaction was incubated for 15 min. Lanes 7 and 8, RNA was incubated with a fivefold molar excess of coat protein prior to addition of 0.2 volume of peroxide-ascorbate solutions. The reaction mixtures were incubated for 10 (lane 7) or 15 (lane 8) min before quenching and loading of the gels. Lanes 9 and 10, RNA only (lane 9) or RNA plus a fivefold molar excess of coat protein (lane 10) was incubated with 0.1 volume of Fe(II)-EDTA and 0.1 volume of DTT solution for 2 h prior to quenching and gel electrophoresis. Lanes 11 and 12, RNA alone (lane 12) or RNA plus a fivefold molar excess of coat protein (lane 11) was incubated with 0.1 volume of Fe(II)-EDTA and 0.1 volume of DTT solution for 2 h prior to quenching of the reactions and processing for analysis of the samples by gel electrophoresis. Numbers indicate nucleotides. (B) Schematic representation of the protection pattern. Nucleotides 738 to 749, 761 to 775, 810 to 819, 841 to 852, and 864 to 873 are protected from cleavage in the presence of coat protein.

tection from hydroxyl radical cleavage at five sites on the 3' RNA fragment. These sites appear as regions of diminished band intensity, and the protected regions (nucleotides 738 to 749, 761 to 775, 810 to 819, 841 to 852, and 864 to 873) are diagrammed schematically in Fig. 2B. Five of the seven (A/U)UGC repeats, therefore, are protected from Fe(II)-EDTA cleavage in this analysis, and all but one (761 to 775 as the

exception) of the protected regions contain at least one (A/U)UGC repeat. Where a protected (A/U)UGC was present, the 5' border of each site is adjacent to a single-stranded (A/U)UGC that separates two adjacent hairpins. Protection extends approximately five to eight nucleotides up the 5' side of the adjacent helix. Protected site 841–852 corresponds closely with an RNase T₁-resistant site described previously (77). Domain 761–775 can be distinguished from the other protected regions both by the absence of an (A/U)UGC sequence and by the fact that the protection includes the 3' side of a hairpin, including part of a bulge loop. We interpret the footprinting data as evidence that coat protein is in close proximity to the RNA at five sites and that some, although not necessarily all, of these protected sites are also coat protein binding domains.

EMSA was used to further investigate the number of coat protein binding sites. We reported elsewhere (4) that several band shifts are resolved in this analysis. For the experiments shown in Fig. 3A, the concentration of added coat protein was increased so that essentially all of the unbound labeled RNA was shifted into complexes. Under these conditions, four distinct complexes were observed (Fig. 3A). (A fifth complex with slowest relative mobility was visible in some experiments, but the specificity of this complex is uncertain.) Van Boxsel previously estimated the number of high-affinity binding sites to be between four and six (68).

We reasoned that if the four EMSA bands (Fig. 3A) represented coat protein dimers bound to separate sites on the UTR RNA (as the footprinting data suggest), then the protein should also bind to smaller fragments containing binding sites. Coat protein was therefore incubated with subfragments of the 170-nucleotide 3' UTR. A *Sau*3A restriction enzyme site at nucleotide 797 of the 3' UTR cDNA allowed us to generate 5' and 3' half-molecules, i.e., nucleotides 718 to 797 and 798 to 881, *in vitro* (Fig. 3B and C). Based on the data presented in Fig. 2, regions of hydroxyl radical protection are superimposed onto the schematic secondary structure maps for the fragments shown in Fig. 3B and C. Specificity was demonstrated by competitive binding with cognate and nonspecific RNAs (data not shown). When coat protein is added to radiolabeled 5' half-molecule AIMV_{718–797}, a single shifted band is observed on a nondenaturing polyacrylamide gel (Fig. 3B). The free RNA shifts abruptly into a complex with coat protein at approximately 3 μ M coat protein. Three shifted bands were observed, however, upon incubation of coat protein with the labeled 3' half-UTR fragment AIMV_{798–881} (Fig. 3C). Coat protein binds AIMV_{798–881} with an apparent K_d of about 300 nM, which is significantly higher affinity than the binding to AIMV_{718–797}. The relatively close spacing of the two lower bands in Fig. 3C, lanes 2 and 3 (bottom arrow), is interpreted as two different conformers of the RNA complexes with equal numbers of coat protein molecules. The results presented in Fig. 3 suggest that the 5' half-molecule AIMV_{718–797} (Fig. 3B) has a single coat protein binding site and the 3' half-molecule AIMV_{798–881} (Fig. 3C) has three coat protein binding sites.

The sequence and/or structural features of AIMV RNA 4 recognized by coat protein were further investigated by testing the binding of coat protein to minus-sense RNA. The genomic RNAs of AIMV are positive sense and must be copied into antisense RNAs by the viral replicase prior to new synthesis of plus-strand RNAs. It has been reported that AIMV coat protein is important in regulating the relative amounts of plus- and minus-sense genomic RNAs synthesized during virus replication (69); i.e., it seems to have a role in up-regulating plus-strand RNA synthesis and down-regulating minus-strand synthesis. Analysis of coat protein binding to antisense 3' UTR

RNA transcripts, which lack the repeated tetranucleotide 5'AUGC3' sequences (converted to 5'GCAU3') but retain the dyad symmetry elements found at 817 to 822/835 to 840 and 848 to 852/860 to 863, is presented in Fig. 4. A predicted secondary structure for the antisense 3'-terminal 170 nucleotides of AIMV RNA 4 (based on comparison with the plus-sense molecule [Fig. 1]) is shown in Fig. 4A. A single antisense AUGC sequence is found at positions 862 to 859. Although it is reasonable to predict that the secondary structure shown in Fig. 4A may form, tertiary interactions present in the positive-sense RNA (Fig. 1) may not be reflected in the minus-sense RNA. The results of binding coat protein to antisense AIMV_{718–881} are shown in Fig. 4B. The data demonstrate that positive-sense RNA is completely shifted into RNA-protein complexes at 2 μ M coat protein (Fig. 4B, lane 8). Conversely, only a small fraction of free antisense RNA shifts into a single complex (lane 3). The binding data show that the relative affinity of coat protein for antisense RNA is much lower than for positive-sense RNA. These results indicate that coat protein binding determinants present in the positive-sense RNA are, for the most part, absent in minus-sense RNA. On the basis of RNase protection of minus-sense genomic AIMV RNAs, Zuidema drew a similar conclusion (75).

Structural analysis of the coat protein-RNA 4 complexes will be facilitated by identifying small subfragments of the RNA that bind coat protein with high affinity; therefore, we have initiated attempts to define a minimal RNA binding site. The half-molecule band shift experiments (Fig. 3) suggested that the 84-nucleotide 3' molecule AIMV_{798–881} retains high-affinity coat protein binding. In addition, Zuidema et al. (77) described a 36-nucleotide 3'-terminal fragment of AIMV RNA 1 that was protected from RNase T₁ digestion and capable of rebinding coat protein, while Koper-Zwartoff and Bol (35) proposed that the coat protein would bind to a hairpin flanked by a 3' AUGC sequence. We transcribed two RNAs, i.e., 26-nucleotide AIMV_{843–868} (Fig. 5A) and 3'-terminal 39-nucleotide AIMV_{843–881} (Fig. 5B), and tested them in the band shift assay. Both of these RNAs contain the CUCAUGCAA (nucleotides 846 to 854) sequence found to be common to the RNase T₁-protected and rebound fragments described by Zuidema et al. (77). Coat protein affinity for the RNAs was evaluated by EMSA. As shown in Fig. 5A, band shifts were not observed with the 26-nucleotide RNA until the coat protein concentration reached 3 μ M. A complete shift of RNA into RNA-protein complexes was not observed with fragment AIMV_{843–868}, even at a coat protein concentration of 10 μ M (Fig. 5A, lane 8). Coat protein bound with much higher affinity to the 39-nucleotide AIMV_{843–881} RNA (Fig. 5B). Complexes were evident at 1.25 μ M coat protein (Fig. 5B, lane 3), and essentially all of the labeled 39-nucleotide RNA fragment was shifted into RNA-protein complexes at 2.5 μ M molar coat protein (lane 4). On the basis of the observation that the 26-nucleotide RNA (Fig. 5A) has a significantly diminished affinity for coat protein, we conclude that a single hairpin flanked by AUGC sequences is insufficient for high-affinity coat protein binding. The apparent K_d of the coat protein-AIMV_{843–881} interaction is approximately 500 nM, while that of the coat protein-AIMV_{843–868} complex is greater than 10 μ M.

To test the possibility of nucleotide sequence-specific coat protein binding to the 39-nucleotide fragment, the AUGC_{865–868} repeat was mutated to AAAA, and the resulting RNA was tested by EMSA (Fig. 6A). The structure of the 3' end of AIMV RNA 4 proposed by Houwing and Jaspars placed C-868 in a base pair with G-877 (29); therefore, we included a G-877→U-877 mutation in order to maintain the putative base

FIG. 3. Electrophoretic mobility band shift analysis of RNA-protein complexes formed between AIMV coat protein and the 3' UTR of AIMV RNA 4. The RNA used in the EMSA reactions is shown at the left of each panel, and the hydroxyl radical footprinting data from Fig. 2 are superimposed on each RNA. The secondary structure and protected regions shown in panels B and C are drawn to facilitate comparison with Fig. 2, not to indicate thermodynamic stability. (A) Coat protein binding to the 170-nucleotide 3' UTR (see Fig. 1). The RNA fragment was radioactively labeled during in vitro transcription, and the EMSA was carried out as described in Materials and Methods, using an input concentration of 250 nM labeled RNA in each reaction. Lane 1, free (unbound) RNA. A small amount of slowly migrating RNA-RNA dimer is present. Reactions analyzed in lanes 2 to 5 contained RNA plus coat protein at concentrations of 250 nM, 625 nM, 1.25 μ M, and 2.5 μ M, respectively. The arrows mark the positions of the four RNA-protein complexes. (B) EMSA of coat protein binding to the 5' half-molecule. The plasmid DNA encoding the 3' UTR of AIMV RNA 4 was linearized with restriction enzyme *Sau*3A before transcription, thereby permitting synthesis of runoff transcripts of nucleotides 718 to 797. Each EMSA reaction included 100 nM labeled AIMV₇₁₈₋₇₉₇ RNA. Lane 1, RNA only. Reactions analyzed in lanes 2 to 5 contained RNA plus coat protein at concentrations of 1, 2, 3, and 4 μ M, respectively. The arrow indicates the position of a single shifted band. (C) Coat protein binding to the 3' half-molecule. A *Sau*3A-*Sma*I fragment including nucleotides 798 to 881 was subcloned into a transcription vector containing the bacteriophage SP6 promoter, and radiolabeled transcripts were synthesized and assayed for coat protein binding. Each EMSA reaction included 250 nM labeled AIMV₇₉₈₋₈₈₁ RNA. Lane 1, RNA only. Reactions analyzed in lanes 2 to 5 contained RNA plus coat protein at concentrations of 250 nM, 500 nM, 1.25 μ M, and 2.5 μ M, respectively. The bottom arrow at approximately the center of the gel points to two bands that, because of their close spacing, are likely to be alternate RNA conformers bound to a single coat protein dimer.

pairing through an A-U pair (Fig. 6A, lowercase letters). In contrast to the high-affinity binding by the wild-type 39-nucleotide RNA (Fig. 5B), coat protein binding by the AAAA₈₆₅₋₈₆₈ mutant RNA was not detectable by EMSA under identical conditions (Fig. 6A). Other data (4) demonstrate that the AAAA₈₆₅₋₈₆₈ mutant RNA (shown schematically in Fig. 6A) also fails to compete effectively for the binding of coat protein to the 39-nucleotide 3'-terminal RNA fragment (see also Fig. 9). The nucleotide sequence specificity of coat protein binding was further characterized by incubating coat protein with a labeled poly(AUGC)₁₀ transcript in the EMSA (Fig. 6B). No binding was detected in the band shift assay, strongly indicating that the sequence AUGC alone is insufficient for coat protein binding. The data presented in Fig. 6 suggest that the loss of coat protein binding to the AAAA₈₆₅₋₈₆₈ mutant RNA could be due to the altered nucleotide sequence of the RNA or to destabilization of the 3' hairpin by changing the putative base pair at 868 and 877 from G-C to A-U.

Comparative nucleotide sequence analysis is a powerful approach for analyzing secondary and higher-order structure in RNA molecules (31, 45); therefore, this method was used to further examine sequence and/or structural elements in AIMV RNA 4 that bind coat protein. The nucleotide sequence analysis of several ilarvirus RNAs has recently been completed (3, 58), and the 3'-terminal 50 nucleotides of 14 RNAs representing eight different viruses are presented in Fig. 7A. In addition to a common 3'-terminal AUGC, a second AUGC repeat is found 17 to 22 nucleotides from the 3' end of the RNA (comparable in position to AUGC₈₆₅₋₈₆₈ in AIMV RNA 4). There is particularly strong conservation of the terminal 18 to 23 nucleotides of the RNAs, which have the consensus sequence RAUGCCYCYX₀₋₄AARGRGAUGC (R = purine, Y = pyrimidine, and X = any nucleotide). The alignment shown in Fig. 7A is centered on the invariant upstream tetranucleotide AUGC (i.e., AUGC₈₆₅₋₈₆₈ in AIMV RNA 4), where flanking conserved sequences are also apparent.

Schematic representations of theoretical secondary structures of the 3'-terminal 50 nucleotides of the RNAs are presented in Fig. 7B and C. These structures are based on sequence alignment (Fig. 7A) and the experimental enzymatic structure mapping of AIMV RNA 4 (52). The RNAs fall into two general classes that differ in the structure and length of the upstream hairpin. The RNAs of parietaria mottle virus, spinach latent virus, elm mottle virus, and citrus variegation virus have a predicted 5' bulge loop in the upstream hairpin (Fig. 7B), while the same hairpin is continuous (Fig. 7C) in the RNAs of AIMV, prune dwarf virus, tobacco streak virus (TSV), and citrus leaf rugose virus. Other characteristics of the RNAs include a conserved purine (R) immediately upstream of the invariant AUGC₈₆₅₋₈₆₈, the maintenance of the downstream stem structure through the sequence 5'CYCY3' paired to 5'RGRG3', and the presence of two adenosine residues in the loop of the 3' hairpin. Nucleotide sequence conservation with covarying nucleotides strongly indicates that the upstream invariant AUGC (865 to 868 in AIMV RNA 4) is single stranded. All of the RNAs shown in Fig. 7A have the sequence RAUGCCYCY, and in all cases, the downstream sequence covaries to maintain base pairing in the CYCY stem. If C-868 of the invariant AUGC₈₆₅₋₈₆₈ were paired with G-877, then the CYCY stem of all RNAs excepting those of AIMV and prune dwarf virus would be destabilized both by introducing C-A base pairs and by decreasing the total number of potential pairs. We conclude, therefore, that it is likely that the invariant AUGC₈₆₅₋₈₆₈ is single stranded, and we favor the secondary structure model of AIMV₈₄₃₋₈₈₁ that is presented in Fig. 5B.

We recently reported an analysis of peptide binding to a

FIG. 4. Electrophoretic mobility band shift analysis of coat protein binding to antisense transcripts of the 3' terminus of AIMV RNA 4. (A) Schematic representation of the predicted secondary structure of the antisense 3'-end transcript. (B) Electrophoretic mobility band shift analysis of coat protein binding. RNA transcripts were radiolabeled during in vitro transcription, and each EMSA reaction mixture contained 100 nM RNA. Lanes 1 to 5, antisense RNA transcripts; lanes 6 to 10, sense transcripts. Lanes 1 and 6, minus-sense or plus-sense RNA only. Reactions analyzed in lanes 2 to 4 and 7 to 9 contained labeled RNA plus 1, 2, and 4 μ M coat protein, respectively. Lanes 5 and 10, EMSA reactions contained labeled RNA, 4 μ M coat protein, plus a 20-fold molar excess of unlabeled cognate RNA added as a competitor of the binding reaction. Higher apparent concentrations of coat protein were required to shift the RNA into complexes in this experiment (lanes 7 to 9) compared with data shown in Fig. 3A because the coat protein preparation was frozen and thawed numerous times and contained a fraction of molecules that were incapable of binding the RNA.

170-nucleotide 3' UTR fragment of AIMV RNA 4 (4), and we have extended the characterization of the peptide-RNA interaction by testing the binding to the 3'-terminal 39-nucleotide binding site (AIMV₈₄₃₋₈₈₁). Electrophoretic mobility band shift data (Fig. 8A) demonstrate that the N-terminal peptides CP25 (lanes 1 to 6) and CP38 (lanes 7 to 12) bind the 3'-terminal 39-nucleotide RNA AIMV₈₄₃₋₈₈₁, as indicated by a band shift to a complex that migrates more slowly than free RNA. The migration of AIMV₈₄₃₋₈₈₁ RNA is diminished only slightly by binding CP25 (lanes 3 to 6), while a greater mobility shift is observed when the RNA binds CP38 (lanes 7 to 12). Peptide binding to the AAAA₈₆₅₋₈₆₈ mutant RNA was also tested (Fig. 8B). The data show that there was no detectable binding of peptides CP25 and CP38 to the mutant AIMV_{843-881aaa} RNA. Therefore, mutating the AUGC₈₆₅₋₈₆₈ to AAAA₈₆₅₋₈₆₈ severely limits the ability of the RNA to bind coat protein (Fig. 5 and 6) or peptides (Fig. 8). These data strongly suggest that binding of both coat protein and N-terminal peptides to the 3'

FIG. 5. Electrophoretic mobility band shift analysis of coat protein binding to truncated fragments of the AIMV RNA 4' 3' terminus. Predicted secondary structures of the RNAs are shown at the left. (A) Radiolabeled 26-nucleotide AIMV₈₄₃₋₈₆₈ was incubated at an input RNA concentration of 100 nM in the presence of increasing amounts of coat protein prior to analysis of reaction products by native gel electrophoresis. Lane 1, RNA only. Reactions analyzed in lanes 2 to 8 contained RNA plus coat protein at concentrations of 200 nM, 500 nM, 1 μ M, 1.5 μ M, 3 μ M, 5 μ M, and 10 μ M, respectively. (B) Coat protein binding to 39-nucleotide AIMV₈₄₃₋₈₈₁. Each EMSA reaction mixture contained labeled RNA added at an input concentration of 250 nM. Lane 1, RNA only. Reactions analyzed in lanes 2 to 6 contained labeled RNA plus coat protein added at concentrations of 250 nM, 1.25 μ M, 2.5 μ M, 5 μ M, and 12.5 μ M, respectively.

AIMV RNA fragments requires the AUGC sequence located at positions 865 to 868. However, these results do not allow us to rule out the possibility that the G-877 $\rightarrow$ U-877 change in AIMV_{843-881aaa} RNA (Fig. 6A) also affects coat protein binding independently by changing the sequence and/or structure of the CYCY stem proposed in Fig. 7B and C.

Peptide binding to the 5' and 3' half-molecule RNAs described in Fig. 3B and C was also tested by EMSA. Although both peptides bound the downstream AIMV₇₉₈₋₈₈₁, peptide binding to the upstream half-molecule AIMV₇₁₈₋₇₉₇ was never observed (data not shown), suggesting that the peptides and full-length coat protein discriminate differentially for specific RNA binding sites. Discriminatory binding activity was therefore assessed by incubating coat protein or peptides with the 170-nucleotide wild-type AIMV₇₁₈₋₈₈₁ molecule in the presence of competitor RNAs containing or lacking the AAAA mutation at nucleotides 865 to 868 (Fig. 9). The AAAA₈₆₅₋₈₆₈ mutation described in Fig. 6A was introduced into AIMV₇₁₈₋₈₈₁ as shown in Fig. 9A. The results of the coat protein binding experiments are shown in Fig. 9B, in which the labeled RNA in all lanes is wild-type AIMV₇₁₈₋₈₈₁. In the presence of a 10-fold molar excess of coat protein, the majority of labeled AIMV₇₁₈₋₈₈₁ RNA is shifted into RNA-protein complexes (Fig. 9B; compare lanes 1 and 8 with lanes 2 and 5). Unlabeled cognate AIMV₇₁₈₋₈₈₁ RNA competed effectively for coat protein binding, resulting in minimal shift of labeled RNA (Fig. 9B, lane 3). The 170-nucleotide AIMV_{718-881aaa} mutant RNA (shown in Fig. 9A) is a weaker competitor than wild-type AIMV₇₁₈₋₈₈₁ RNA (Fig. 9B, lane 4). The fact that the 170-nucleotide RNA containing the AUGC $\rightarrow$ AAAA mutation

FIG. 6. (A) Electrophoretic mobility band shift analysis of coat protein binding to 39-nucleotide mutant AIMV_{843-881aaa} RNA. The AUGC₈₆₅₋₈₆₈ sequence was changed to AAAA by transcription directed by a chemically synthesized DNA template containing the mutated nucleotides. The concentration of radiolabeled RNA in each reaction was 100 nM. Lane 1, RNA only. Reactions analyzed in lanes 2 to 6 contained RNA plus AIMV coat protein added at concentrations of 100 nM, 500 nM, 1 μ M, 2 μ M, and 5 μ M, respectively. (B) Electrophoretic mobility band shift analysis of coat protein binding to poly(AUGC)₁₀. Radiolabeled poly(AUGC)₁₀ was synthesized by in vitro transcription using a chemically synthesized DNA template and added to the EMSA reactions at a concentration of 100 nM. Lane 1, RNA only. Reactions analyzed in lanes 2 to 10 contained RNA plus coat protein added at concentrations of 50 nM, 100 nM, 200 nM, 500 nM, 1 μ M, 1.5 μ M, 3 μ M, 5 μ M, and 10 μ M, respectively.

retained weak competitive activity (Fig. 9B, lane 4) is consistent with the conclusion that the 3' UTR has multiple coat protein binding sites (Fig. 3). The 39-nucleotide RNAs were also used as competitors. Although the 39-nucleotide AIMV₈₄₃₋₈₈₁ competitor RNA blocked coat protein binding (Fig. 9B, lane 6), the mutant RNA carrying the 865-868aaa mutation failed to compete (Fig. 9B, lane 7). The latter result is consistent with the data presented in Fig. 6A, in which coat protein binding to AIMV_{843-881aaa} was undetectable.

The negative band shift that results from the binding of CP25 and CP38 to the AIMV₇₁₈₋₈₈₁ fragment (4) is shown in Fig. 9C and D (compare lanes 1 and 8 with lanes 2 and 5). The wild-type AIMV₇₁₈₋₈₈₁ RNA 3' UTR fragment is a potent competitor of peptide binding (Fig. 9C and D, lanes 3). However, the mutant 170-nucleotide AIMV_{718-881aaa} RNA fails to block formation of the fast-migrating peptide-RNA complexes (Fig. 9C and D, lanes 4). These results demonstrate that although the AIMV_{718-881aaa} RNA (Fig. 9A) partially blocks coat protein binding to the AIMV₇₁₈₋₈₈₁ RNA (Fig. 9B, lane 4), it does not block peptide binding (Fig. 9C and D, lanes 4). Data presented in Fig. 8B also show that this mutant RNA fails to bind peptide. The 39-nucleotide RNA AIMV₈₄₃₋₈₈₁ competed effectively for peptide binding to the labeled AIMV₇₁₈₋₈₈₁ RNA (Fig. 9C and D, lanes 6); however, the same RNA carrying the AAAA mutation failed to compete (Fig. 9C and D, lanes 7). The data presented in Fig. 8B and 9B to D strongly indicate that peptide binding requires the AUGC₈₆₅₋₈₆₈ site.

Because the AUGC $\rightarrow$ AAAA mutation at 865 to 868 also included a compensatory change at G-877 (Fig. 9A), we were not able to conclude that mutations in the invariant

FIG. 7. Nucleotide sequence alignment of the 3' termini of AIMV RNAs and ilarvirus RNAs. Shown are the 3'-terminal 50 nucleotides of the following viral RNAs: prune dwarf virus (PDV) (3), TSV (14, 35), spinach latent virus (SLV) (58), parietaria mottle virus (ParMV) (58), elm mottle virus (EMV) (58), citrus leaf rugose virus (CiLRV) (58), citrus variegation virus (CVV) (58), and AIMV (7, 36). The conserved AUGC sequences relative to nucleotide positions 865 to 868 in AIMV RNAs are underlined and presented in boldface. Other AUGC sequences in the RNAs are underlined. (B) Predicted secondary structure for the 3'-terminal 50 nucleotides of parietaria mottle virus, spinach latent virus, elm mottle virus, and citrus variegation virus. (C) Predicted secondary structure for the 3'-terminal 50 nucleotides of prune dwarf virus, TSV, citrus leaf rugose virus, and AIMV.

AUGC₈₆₅₋₈₆₈ alone disrupted coat protein or peptide binding. Therefore, the role of the invariant AUGC₈₆₅₋₈₆₈ was further explored by introducing additional mutations that did not affect the base pairing in the 3' CYCY stem. The peptide binding assay was used to analyze the effects of mutations in the invariant AUGC region. The RNA used in the binding assay shown in Fig. 10A is the 170-nucleotide 3' UTR of AIMV RNA 4 containing two nucleotide changes in the invariant AUGC₈₆₅₋₈₆₈ sequence (AUGC→AACC). The migration of free RNA is shown in lane 1, while the negative band shift resulting from peptide binding to the wild-type RNA (shown in Fig. 1) is seen in lane 2. Lane 3 represents the migration pattern of the AUGC→AACC mutant RNA alone, and lanes 4 to 6 show the migration of RNA in the presence of increasing amounts of peptide CP25. Unlike the mobility change observed with the wild-type RNA (lane 2), the mobility of the AUGC→AACC mutant RNA is unchanged in the presence of peptide, indicating that peptide fails to bind this RNA. A different result was obtained by mutating only a single nucleotide in the invariant AUGC₈₆₅₋₈₆₈ as shown in Fig. 10B. Here, the sequence was changed from AUGC to AAGC, and the effect of added peptide is shown in Fig. 10B, lanes 4 to 6. In the presence of peptide, the RNA migration of the mutant RNA is increased in the same way (lanes 4 to 6) as that of the wild-type RNA (lane 2). Therefore, the single U-866→A mutation has a minimal effect on binding, as demonstrated by the marked increase in electrophoretic mobility in the presence of peptide (Fig. 10B), while the double mutant AUGC→AACC (Fig. 10A) has a marked effect on peptide binding. We have also mutated C-868 to A-868 without effect on peptide binding (data not shown). The data presented in Fig. 10 are evidence that modification of nucleotides within the invariant AUGC₈₆₅₋₈₆₈ alone is sufficient to disrupt peptide binding. Altogether, the results shown in Fig. 5 to 10 indicate that specific coat protein or peptide binding requires a combination of sequence determinants present in the invariant AUGC in addition to structural determinants present in the conserved 3' hairpins.

DISCUSSION

This report focuses on AIMV as a model system for the characterization of features of sequence and/or structure (determinants) in RNA and protein that underlie formation of a specific ribonucleoprotein complex. Although a number of classes of RNA-binding proteins have been described (reviewed in reference 41), the binding determinants of many RNA-binding proteins (for example, ribosomal proteins, some phage proteins, viral coat proteins, some heterogeneous nuclear ribonucleoproteins) are not yet known. AIMV coat protein has a lysine-rich N-terminal arm that is necessary and sufficient for specific binding to AIMV RNAs (4). AIMV coat protein binds to its own mRNA (AIMV RNA 4) and to the 3' termini of the three AIMV viral genomic RNAs.

A defining feature of AIMV and the ilarviruses is that coat protein is required for activating the early stages of virus replication (17). The coat proteins of AIMV and the ilarviruses are known to cross-activate replication; that is, AIMV coat protein activates replication not only of AIMV RNAs but also of the RNAs of TSV, a related ilarvirus (70). Other examples of cross-activation of replication have also been described (19-22). The cross-activation phenomenon is striking because the coat proteins are serologically unrelated or highly variable at the primary sequence level yet capable of activating replication of RNAs that have 3' sequence homology (Fig. 7). One challenge is to understand how these RNA-binding proteins can be functionally equivalent (in genome activation) yet structurally dissimilar at the level of primary sequence or serological cross-reactivity.

Little is known about AIMV replicase, which is encoded on genomic RNAs 1 and 2 (32). Quadri et al. (51) have reported that AIMV coat protein copurifies in replicase preparations, suggesting that it may form a part of the replicase complex. The coat protein of TSV, which is a member of the ilarvirus group, apparently contains zinc, and Sehnke et al. (59) have proposed that the TSV coat protein and possibly AIMV coat protein (50) contain zinc finger domains common to other nucleic acid-binding proteins (34, 64, 66). The putative zinc

FIG. 8. (A) EMSA of peptide binding to the 39-nucleotide AIMV₈₄₃₋₈₈₁. Peptides CP25 and CP38 represent the N-terminal 25 or 38 amino acids of AIMV coat protein (see amino acid sequences in Materials and Methods). The concentration of radiolabeled RNA in each reaction was 100 nM. Lanes 1 and 7, RNA only. Reactions analyzed in lanes 2 to 6 contained labeled RNA plus peptide CP25 concentrations of 100 nM, 500 nM, 1 μ M, 2 μ M, and 5 μ M, respectively. Reactions analyzed in lanes 8 to 12 contained labeled RNA plus peptide CP38 at concentrations of 100 nM, 500 nM, 1 μ M, 2 μ M, and 5 μ M, respectively. (B) Electrophoretic mobility band shift analysis of peptide binding to the 39-nucleotide AIMV_{843-881aaa} mutant RNA containing the AUGC $\rightarrow$ AAAA mutation at nucleotides 865 to 868. The concentration of labeled RNA in each EMSA reaction mixture was 100 nM. Lanes 1 and 7, RNA only. Reactions analyzed in lanes 2 to 6 contained RNA plus peptide CP25 added at concentrations of 100 nM, 500 nM, 1 μ M, 2 μ M, and 5 μ M, respectively. Reactions analyzed in lanes 8 to 12 contained labeled RNA plus peptide CP38 added at concentrations of 100 nM, 500 nM, 1 μ M, 2 μ M, and 5 μ M, respectively.

finger domain of AIMV coat protein is far removed from the N-terminal RNA binding domain, and the relevance of this feature to AIMV coat protein function is not clear.

The presence of seven (A/U)UGC repeats within the 180-nucleotide 3' UTR is an obvious feature of the AIMV RNAs. The homology alignment (Fig. 7A) demonstrates, however, that although AUGC₈₆₅₋₈₆₈ and the 3'-terminal AUGC are common, multiple AUGC repeats are not found uniformly in related viral RNAs. The functional significance of the multiple repeats in AIMV is not known but may be relevant to the viral assembly process. The nuclease or hydroxyl radical cleavage data (Fig. 2) suggest a similar footprint pattern among the (A/U)UGC repeats in AIMV RNA 4, where a single-stranded (A/U)UGC forms the base of a protected region that continues up the 5' strand of the hairpin. The single protected area that diverges from this pattern (761 to 775) is one that lacks an (A/U)UGC and, unlike the others, shows partial protection of a 3' bulge loop and the 3' strand of hairpin. The isolated 5' half-molecule (AIMV₇₁₈₋₇₉₇) seems to have only a single coat

FIG. 9. Discriminatory binding of AIMV coat protein and peptides to AIMV RNAs containing the AUGC $\rightarrow$ AAAA mutation. (A) Secondary structure of RNA AIMV_{718-881aaa}. The underbar shows the position of the AUGC $\rightarrow$ AAAA mutation. (B to D) EMSA reaction mixtures containing 100 nM labeled AIMV₇₁₈₋₈₈₁ and 1 μ M coat protein (B, lanes 2 to 7) or 2.5 μ M peptide (C and D, lanes 2 to 7) were incubated with competitor RNAs. Competitor RNA concentrations are expressed as moles of nucleotide. Lanes 1 and 8, RNA only; lanes 2 and 5, RNA plus coat protein or peptide; lanes 3, 25-fold excess of competitor AIMV₇₁₈₋₈₈₁ RNA; lanes 4, 25-fold excess of competitor mutant AIMV_{718-881aaa} RNA (shown in panel A); lanes 6, with added 24-fold excess of competitor AIMV₈₄₃₋₈₈₁ RNA; lanes 7, with added 24-fold excess of AIMV_{843-881aaa} competitor RNA. Arrows indicate the free RNA (F) and the negative shift of the peptide-bound (B) RNA.

protein binding site in EMSA (Fig. 3B), perhaps suggesting that only one of the two sites (738 to 749 or 761 to 775) in the intact 3' UTR is a coat protein binding site, while other site is protected by coat protein molecules bound nearby. An alternate explanation is that site 761-775 in the 5' half-molecule (Fig. 5B) fails to bind because the truncated RNA lacks the proper structural features, including several paired nucleotides. Nevertheless, the combination of footprinting data (Fig. 2) and band shift analysis of coat protein binding to the 5' and 3' half-molecules (Fig. 3) provides evidence for multiple coat protein binding sites on AIMV RNA 4.

Although coat protein binds specifically to both half-UTR molecules, we have not detected CP25 or CP38 peptide binding to the 5' half-molecule AIMV₇₁₈₋₇₉₇. These findings, coupled with the mutagenesis results (Fig. 6A and 9B), lead to our conclusion that peptides CP25 and CP38 require AUGC₈₆₅₋₈₆₈ for binding. Results presented in Fig. 9 strongly indicate that peptides and coat protein discriminate differen-

FIG. 10. EMSA of peptide binding to RNAs containing mutations of the invariant AUGC₈₆₅₋₈₆₈ sequence within AIMV₇₁₈₋₈₈₁. Mutations were introduced by PCRs primed with oligonucleotides containing nucleotide substitutions. All EMSA reaction mixtures contained 100 nM radiolabeled RNA. Peptides were added as indicated prior to analysis by electrophoresis into a nondenaturing polyacrylamide gel. The peptide used is CP25, representing the N-terminal 25 amino acids of AIMV coat protein. (A) Lane 1, wild-type AIMV₇₁₈₋₈₈₁ RNA; lane 2, wild-type AIMV₇₁₈₋₈₈₁ RNA plus a 10-fold molar excess of CP25. Lanes 3 to 6 contain AIMV₇₁₈₋₈₈₁ RNA that is mutated from AUGC₈₆₅₋₈₆₈ to AACC₈₆₅₋₈₆₈ and incubated without peptide (lane 3) or with 10-, 25-, and 50-fold molar excesses of peptide CP25 added (lanes 4 to 6, respectively). (B) Lane 1, wild-type AIMV₇₁₈₋₈₈₁ RNA; lane 2, wild-type AIMV₇₁₈₋₈₈₁ RNA plus a 10-fold molar excess of CP25. Lanes 3 to 6 contain AIMV₇₁₈₋₈₈₁ RNA that is mutated from AUGC₈₆₅₋₈₆₈ to AAGC₈₆₅₋₈₆₈ and incubated without peptide (lane 3) or with 10-, 25-, and 50-fold molar excesses of peptide CP25 added (lanes 4 to 6, respectively).

tially for binding to the 170-nucleotide UTR. Weeks and Crothers reported related observations (72). When the human immunodeficiency virus Tat-TAR interaction was analyzed, 14- and 38-amino-acid peptides of the basic region of human immunodeficiency virus Tat protein had similar binding affinities for the TAR RNA sequence, but the shorter peptide dissociated more rapidly and discriminated less well between specific and nonspecific sites. Although AIMV coat protein will bind to the 3' UTR fragment containing the AUGC→AAAA mutation (shown by competitive binding in Fig. 9B), the affinity is diminished by the mutation; moreover, the peptides fail to bind this RNA (Fig. 9).

Koper-Zwartoff and Bol (35) and Zuidema et al. (77) predicted that coat protein recognition of AIMV RNA 4 requires a combination of RNA sequence and structural features. The nucleotide sequence homology data (Fig. 7A), coupled with comparable secondary structure folding based on covarying nucleotides, provide evidence to support this prediction. Although Koper-Zwartoff and Bol (35) proposed that the coat protein would bind to a hairpin flanked by a 3' AUGC sequence, weak binding to the 26-nucleotide AIMV₈₄₃₋₈₆₈ (Fig. 5A) in light of high-affinity binding to the 39-nucleotide AIMV₈₄₃₋₈₈₁ (Fig. 5B) shows that the single hairpin/AUGC is

insufficient. Recent deletion data suggest that AIMV₈₄₃₋₈₈₀ may be the minimal high-affinity binding site (1).

Data reported here establish AIMV RNA fragment AIMV₈₄₃₋₈₈₁ as a high-affinity coat protein-binding element containing two invariant AUGC repeats (AUGC₈₆₅₋₈₆₈ and 3' terminal) with surrounding structural determinants. The RNA-protein contact points are not yet known, however. A full understanding will require additional biochemical characterization, including modification interference analysis, coupled with structural analysis using nuclear magnetic resonance spectroscopy or X-ray crystallography. Coat protein may interact directly with the bases of the AUGC element or, alternatively, the AUGC sequence may facilitate formation of a structural motif that is favorable for coat protein binding. Steitz (62) has emphasized that the potential of an RNA structure to be deformed in order to conform with protein conformation is likely to be an important parameter for determining binding affinities. AIMV RNAs lack a pseudoknot structure or a tRNA-like 3' structure that might serve as a recognition signal for the viral replicase (18, 47, 57); therefore, Houwing and Jaspars proposed that coat protein binding is accompanied by a conformational change that generates the replicase recognition signal (27, 28). We find that peptide binding to the 3' UTR of AIMV RNA 4 is accompanied by increased electrophoretic mobility of the complexes in native polyacrylamide gels in addition to an altered circular dichroism spectrum (4), lending support to the possibility of a conformational change.

The functional significance of mRNA 3' untranslated sequences is being reevaluated (30, 74) as roles in regulation of mRNA stability (2), translation (13, 16, 23, 65), mRNA cytolocalization (40, 60), and cellular differentiation (47) are being described. Specific protein binding to the 3' UTR is certain to play an important role. The 3' UTR of AIMV RNA 4 is multifunctional through its roles in mRNA stabilization (39), virus replication (reviewed in reference 32), and probably virus assembly, although an origin of assembly site has not been identified for AIMV. AIMV coat protein is not a translational repressor (43, 56, 71) in the manner that R17 coat protein represses viral replicase expression (10, 53); however, we have recently found that deletion of the 3' UTR significantly decreases translation of AIMV RNA 4 in a HeLa cell extract (24). One additional role for the coat protein-AIMV RNA 4 interaction may be in coat protein-mediated virus cross-protection, whereby coat protein expression in transgenic plants minimizes symptom development following virus superinfection (5, 6, 48). A detailed analysis of structure and structure-function relationships will be required to appreciate the complexity and significance of specific RNA-protein interactions in this system.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant GM42504 from the National Institutes of Health.

We thank L. Sue Loesch-Fries and Nancy Jarvis for the AIMV RNA 4 cDNA clone and Edward Halk for providing purified AIMV coat protein. We are very grateful to Simon Scott and Xin Ge (Clemson University) for sharing ilarvirus nucleotide sequence data prior to publication. Jamie Williamson, Sue Loesch-Fries, and members of the Gehrke laboratory provided useful criticism and discussions. Peptides were synthesized by the Massachusetts Institute of Technology Biopolymers laboratory.

REFERENCES

1. Ansel, P. L., A. Anderson-Chang, C. Frederick, and L. Gehrke. Unpublished data.
2. Atwater, J. A., R. Wisdom, and I. M. Verma. 1990. Regulated messenger RNA stability. *Annu. Rev. Genet.* 24:519-541.

3. Bachman, E. J., S. W. Scott, G. Xin, and V. B. Vance. The complete nucleotide sequence of prune dwarf virus RNA 3. *Virology*, in press.
4. Baer, M., F. Houser, L. S. Loesch-Fries, and L. Gehrke. Specific RNA binding by amino-terminal peptides of alfalfa mosaic virus coat protein. *EMBO J.*, in press.
5. Beachy, R. N., S. Loesch-Fries, and N. E. Tumer. 1990. Coat protein-mediated resistance against virus infection. *Annu. Rev. Phytopathol.* **28**:451-474.
6. Beachy, R. N., R. W. Nelson, J. Register III, R. T. Fraley, and N. Tumer. 1987. Transformation to produce virus-resistant plants, p. 47-53. *In* R. T. Fraley, N. M. Frey, and J. Schell (ed.), *Genetic improvements of agriculturally important crops*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
7. Brederode, R. T., E. C. Koper-Zwartoff, and J. F. Bol. 1980. Complete nucleotide sequence of alfalfa mosaic virus RNA 4. *Nucleic Acids Res.* **8**:2213-2223.
8. Burkhoff, A. M., and T. D. Tullius. 1987. The unusual conformation adopted by the adenine tracts in kinetoplast DNA. *Cell* **48**:935-943.
9. Calnan, B. J., S. Biancalana, D. Hudson, and A. D. Frankel. 1991. Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA protein recognition. *Genes Dev.* **5**:201-210.
10. Carey, J., V. Cameron, P. L. de Haseth, and O. C. Uhlenbeck. 1983. Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. *Biochemistry* **22**:2601-2610.
11. Celander, D. (University of Colorado). Personal communication.
12. Celander, D. W., and T. R. Cech. 1990. Iron(II)-ethylenediaminetetraacetic acid catalyzed cleavage of RNA and DNA oligonucleotides: similar reactivity toward single- and double-stranded forms. *Biochemistry* **29**:1355-1360.
13. Ch'ng, J. L. C., D. L. Shoemaker, P. Schimmel, and E. W. Holmes. 1990. Reversal of creatine kinase translational repression by 3' untranslated sequences. *Nature (London)* **248**:1003-1006.
14. Cornelissen, B. J. C., H. Janssen, D. Zuidema, and J. F. Bol. 1984. Complete nucleotide sequence of tobacco streak virus RNA 3. *Nucleic Acids Res.* **12**:2427-2437.
15. Daly, T. J., J. R. Rusche, T. E. Maione, and A. D. Frankel. 1990. Circular dichroism studies of the HIV-1 Rev protein and its specific RNA binding site. *Biochemistry* **29**:9791-9795.
16. Danthinne, X., J. Seurinck, F. Meulewaeter, M. Vanmontagu, and M. Cornelissen. 1993. The 3' untranslated region of satellite tobacco necrosis virus RNA stimulates translation in vitro. *Mol. Cell. Biol.* **13**:3340-3349.
17. Francki, R. I. B. 1985. Polyhedral virions with tripartite genomes. *Plant Viruses* **1**:309.
18. Garcia-Arenal, F. 1988. Sequence and structure at the genome 3' end of the U2-strain of tobacco mosaic virus, a histidine-accepting tobamovirus. *Virology* **167**:201-206.
19. Gonsalves, D., and R. W. Fulton. 1977. Activation of Prunus necrotic ringspot virus and rose mosaic virus by RNA 4 components of some ilarviruses. *Virology* **81**:398-407.
20. Gonsalves, D., and S. M. Garnsey. 1974. Infectivity of the multiple nucleoprotein and RNA components of citrus leaf rugose virus. *Virology* **64**:343-353.
21. Gonsalves, D., and S. M. Garnsey. 1975. Functional equivalence of an RNA component and coat protein for infectivity of citrus leaf rugose virus. *Virology* **64**:25-31.
22. Gonsalves, D., and S. M. Garnsey. 1975. Nucleic acid components of citrus variegation virus and their activation by coat protein. *Virology* **67**:311-318.
23. Han, J., G. Huez, and B. Beutler. 1991. Interactive effects of the tumor necrosis factor promoter and 3'-untranslated regions. *J. Immunol.* **146**:1843-1848.
24. Hann, L. E., and L. Gehrke. 1992. Identification of *cis*-acting sequence required for selective translation of alfalfa mosaic virus RNA 4 in poliovirus-infected HeLa cell extracts, abstr. 127. *Abstr. Cold Spring Harbor Symp. Translational Control*.
25. Hershey, J. W. B. 1991. Translational control in mammalian cells. *Annu. Rev. Biochem.* **60**:717-755.
26. Houwing, C., and E. M. J. Jaspars. 1980. Preferential binding of 3'-terminal fragments of alfalfa mosaic virus RNA to virions. *Biochemistry* **19**:5261-5263.
27. Houwing, C. J., and E. M. J. Jaspars. 1978. Coat protein binds to the 3'-terminal part of RNA 4 of alfalfa mosaic virus. *Biochemistry* **17**:2927-2933.
28. Houwing, C. J., and E. M. J. Jaspars. 1980. Complexes of alfalfa mosaic virus RNA 4 with one and three coat protein dimers. *Biochemistry* **19**:5244-5260.
29. Houwing, C. J., and E. M. J. Jaspars. 1982. Protein binding sites in nucleation complexes of alfalfa mosaic virus RNA 4. *Biochemistry* **21**:3408-3414.
30. Jackson, R. J. 1993. Cytoplasmic regulation of mRNA function: the role of the 3' untranslated region. *Cell* **74**:9-14.
31. Jaeger, J. A., J. Santalucia, and I. Tinoco. 1993. Determination of RNA structure and thermodynamics. *Annu. Rev. Biochem.* **62**:255-287.
32. Jaspars, E. M. J. 1985. Interaction of alfalfa mosaic virus nucleic acid and protein, p. 155-221. *In* J. W. Davies (ed.), *Molecular plant virology*. CRC Press, New York.
33. Kjems, J., M. Brown, D. D. Chang, and P. A. Sharp. 1991. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc. Natl. Acad. Sci. USA* **88**:683-687.
34. Klug, A., and R. Danciel. 1987. "Zinc fingers": a novel protein motif for nucleic acid recognition. *Trends Biochem. Sci.* **12**:464-469.
35. Koper-Zwartoff, E. C., and J. F. Bol. 1980. Nucleotide sequence of the putative recognition site for coat protein in the RNAs of alfalfa mosaic virus and tobacco streak virus. *Nucleic Acids Res.* **8**:3307-3318.
36. Koper-Zwartoff, E. C., F. T. Brederode, P. Walstra, and J. F. Bol. 1979. Nucleotide sequence of the 3'-noncoding region of alfalfa mosaic virus RNA 4 and its homology with the genomic RNAs. *Nucleic Acids Res.* **7**:1887-1900.
37. Kruseman, J., B. Kraal, E. Jaspars, J. Bol, F. Brederode, and H. Veldstra. 1971. Molecular weight of the coat protein of alfalfa mosaic virus. *Biochemistry* **10**:447-455.
38. Latham, J. A., and T. R. Cech. 1989. Defining the inside and outside of a catalytic RNA molecule. *Science* **245**:276-282.
39. Loesch-Fries, L. S., N. P. Jarvis, K. J. Krahn, S. E. Nelson, and T. C. Hall. 1985. Expression of alfalfa mosaic virus RNA 4 cDNA transcripts *in vitro* and *in vivo*. *Virology* **146**:177-187.
40. Macdonald, P. M., and G. Struhl. 1988. *Cis*-acting sequences responsible for anterior localization of bicoid mRNA in *Drosophila* embryos. *Nature (London)* **336**:595-598.
41. Mattaj, J. W. 1993. RNA recognition: a family matter? *Cell* **73**:837-840.
42. Milligan, J. F., D. R. Groebe, G. W. Witherell, and O. C. Uhlenbeck. 1987. Oligoribonucleotide synthesis using T7 polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**:8783-8798.
43. Mohier, E., L. Hirth, M. A. Le Meur, and P. Gerlinger. 1976. Analysis of alfalfa mosaic virus 17S RNA translational products. *Virology* **71**:615-621.
44. Moore, M. J., C. C. Query, and P. A. Sharp. 1993. Splicing of precursors to mRNA by the spliceosome, p. 303-357. *In* R. Gesteland and J. F. Atkins (ed.), *The RNA world*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
45. Noller, H. F., and C. R. Woese. 1981. Secondary structure of 16S ribosomal RNA. *Science* **212**:403-411.
46. Pinck, L., and M. Pinck. 1979. Sequence homology at the 3'-ends of alfalfa mosaic virus RNAs. *FEBS Lett.* **107**:61-65.
47. Pleij, C. W. A., K. Rietveld, and L. Bosch. 1985. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.* **13**:1717-1731.
48. Powell-Abel, P., R. S. Nelson, B. De, N. Hoffmann, S. G. Rogers, R. T. Fraley, and R. N. Beachy. 1986. Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene. *Science* **232**:738-743.
49. Puglisi, J. D., R. Tan, B. J. Calnan, A. D. Frankel, and J. R. Williamson. 1992. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **257**:76-80.
50. Quadt, R., and E. M. J. Jaspars. 1991. Effect of removal of zinc on alfalfa mosaic virus RNA-dependent RNA polymerase. *FEBS Lett.* **278**:61-62.

51. **Quadt, R., H. J. M. Rosdorff, T. W. Hunt, and E. M. J. Jaspars.** 1991. Analysis of the protein composition of alfalfa mosaic virus RNA-dependent RNA polymerase. *Virology* **182**:309–315.
52. **Quigley, G. J., L. Gehrke, D. A. Roth, and P. E. Auron.** 1984. Computer-aided nucleic acid secondary structure modeling incorporating enzymatic digestion data. *Nucleic Acids Res.* **12**:347–366.
53. **Romaniuk, P. J., P. Lowary, H.-N. Wu, G. Stormo, and O. C. Uhlenbeck.** 1987. RNA binding site of R17 coat protein. *Biochemistry* **26**:1563–1568.
54. **Rould, M. A., J. J. Perona, D. Soll, and T. A. Steitz.** 1989. Structure of *E. coli* glutamyl-tRNA synthetase complexed with tRNA^{Gln} and ATP at 2.8 Å resolution. *Science* **246**:1135–1142.
55. **Ruff, M., S. Krishnaswamy, M. Boeglin, A. Poterszman, A. Mitschler, A. Podjarny, B. Rees, J. C. Thierry, and D. Moras.** 1991. Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA^{Asp}. *Science* **252**:1682–1689.
56. **Rutgers, A. S.** 1977. In vitro and in vivo translation of the RNAs of alfalfa mosaic virus. Ph.D. thesis. University of Leiden, Leiden, The Netherlands.
57. **Schimmel, P.** 1989. RNA pseudoknots that interact with components of the translation apparatus. *Cell* **58**:9–12.
58. **Scott, S., and X. Ge. (Clemson University).** Personal communication.
59. **Sehnke, P. C., A. M. Mason, S. J. Hood, R. M. Lister, and J. E. Johnson.** 1989. A “zinc-finger”-type binding domain in tobacco streak virus coat protein. *Virology* **168**:48–56.
60. **Singer, R. H.** 1993. Spatial organization of messenger RNA within cells. *J. Cell. Biochem.* **52**:125–126.
61. **Srinivasan, S., and E. M. J. Jaspars.** 1978. Influence of a few coat protein subunits on the base-paired structure of the RNA species of alfalfa mosaic virus. *Biochim. Biophys. Acta* **520**:237–241.
62. **Steitz, T. A.** 1990. Structural studies of protein nucleic acid interaction—the sources of sequence-specific binding. *Q. Rev. Biophys.* **23**:205.
63. **Tan, R. Y., and A. D. Frankel.** 1992. Circular dichroism studies suggest that TAR RNA changes conformation upon specific binding of arginine or guanidine. *Biochemistry* **31**:10288–10294.
64. **Theunissen, O., F. Rudt, U. Guddat, H. Mentzel, and T. Pieler.** 1992. RNA and DNA binding zinc fingers in *Xenopus* TFIIIA. *Cell* **71**:679–690.
65. **Timmer, R. T., L. A. Benkowski, D. Schodin, S. R. Lax, A. M. Metz, J. M. Ravel, and K. S. Browning.** 1993. The 5′ and 3′ untranslated regions of satellite tobacco necrosis virus RNA affect translational efficiency and dependence on a 5′ cap structure. *J. Biol. Chem.* **268**:9504–9510.
66. **Vallee, B. L., J. E. Coleman, and D. S. Auld.** 1991. Zinc fingers, zinc clusters, and zinc twists in DNA-binding protein domains. *Proc. Natl. Acad. Sci. USA* **88**:999–1003.
67. **van Beynum, G. M. A., J. M. de Graaf, A. Castel, B. Kraal, and L. Bosch.** 1977. Structural studies of the coat protein of alfalfa mosaic virus. *Eur. J. Biochem.* **72**:63–78.
68. **van Bossel, J. A. M.** 1976. High-affinity sites for coat protein on alfalfa mosaic virus. Ph.D. thesis. University of Leiden, Leiden, The Netherlands.
69. **van der Kuyl, A. C., L. Neeleman, and J. F. Bol.** 1991. Role of alfalfa mosaic virus coat protein in regulation of the balance between viral plus and minus strand RNA synthesis. *Virology* **185**:496–499.
70. **van Vloten-Doting, L.** 1975. Coat protein is required for infectivity of tobacco streak virus: biological equivalence of the coat proteins of tobacco streak and alfalfa mosaic virus. *Virology* **65**:215–225.
71. **Van Vloten-Doting, L., T. Rutgers, L. Neeleman, and L. Bosch.** 1975. In vitro translation of the RNAs of alfalfa mosaic virus. *Colloq. Inst. Natl. Sante Rech. Med.* **47**:233–237.
72. **Weeks, K. M., and D. M. Crothers.** 1992. RNA binding assays for Tat-derived peptides—implications for specificity. *Biochemistry* **31**:10281–10287.
73. **Weinberger, J., D. Baltimore, and P. Sharp.** 1986. Distinct factors bind to apparently homologous sequences in the immunoglobulin heavy-chain enhancer. *Nature (London)* **322**:846–848.
74. **Wickens, M.** 1993. Messenger RNA—springtime in the desert. *Nature (London)* **363**:305–306.
75. **Zuidema, D.** 1983. Specific binding sites on RNAs and coat protein of alfalfa mosaic virus involved in genome activation. Ph.D. thesis. University of Leiden, Leiden, The Netherlands.
76. **Zuidema, D., M. F. A. Bierhuizen, B. J. C. Cornelissen, J. F. Bol, and E. M. J. Jaspars.** 1983. Coat protein binding sites on RNA 1 of alfalfa mosaic virus. *Virology* **125**:361–369.
77. **Zuidema, D., R. H. Cool, and E. M. J. Jaspars.** 1984. Minimum requirements for specific binding of RNA and coat protein of alfalfa mosaic virus. *Virology* **136**:282–292.
78. **Zuidema, D., and E. M. J. Jaspars.** 1985. Specificity of RNA and coat protein interaction in alfalfa mosaic virus and related viruses. *Virology* **140**:342–350.