



Published in final edited form as:

Science. 2004 December 17; 306(5704): 2108–2111.

Cofolding Organizes Alfalfa Mosaic Virus RNA and Coat Protein for Replication

Laura M. Guogas^{1,3}, David J. Filman², James M. Hogle^{2,*}, and Lee Gehrke^{1,3,*}

¹Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115, USA.

²Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.

³Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Abstract

Alfalfa mosaic virus genomic RNAs are infectious only when the viral coat protein binds to the RNA 3' termini. The crystal structure of an alfalfa mosaic virus RNA-peptide complex reveals that conserved AUGC repeats and Pro-Thr-x-Arg-Ser-x-x-Tyr coat protein amino acids cofold upon interacting. Alternating AUGC residues have opposite orientation, and they base pair in different adjacent duplexes. Localized RNA backbone reversals stabilized by arginine-guanine interactions place the adenosines and guanines in reverse order in the duplex. The results suggest that a uniform, organized 3' conformation, similar to that found on viral RNAs with transfer RNA-like ends, may be essential for replication.

A general problem in positive-strand RNA virology is understanding how viral RNA replication is initiated by the RNA-dependent RNA polymerase (replicase) on the correct template and nucleotide in an infected cell. Alfalfa mosaic virus (AMV) and ilarviruses are unusual positive-sense viruses, the genomic RNAs of which are replicated only in the presence of the viral coat protein (CP) (1,2). These viruses are distinguished from many other members of the virus family Bromoviridae because they lack canonical features of the tRNA-like structure (TLS) common at the 3' termini of the viral RNA genomes. The TLS is a necessary and sufficient feature for recruitment of the bromovirus replicase (3,4). CP-induced structural organization of the AMV RNA 3' terminus may create a functional homolog of the tRNA tail and thereby permit recognition by the RNA-dependent RNA polymerase.

CP binds specifically to the 3' untranslated regions (3'UTRs) found on all four RNAs of the segmented AMV genome (5). The 180-nucleotide 3'UTR secondary structure likely consists of six hairpins, most of which are separated by single-stranded tetranucleotide AUGC repeats (5-8). These repeats are characteristic of AMV and ilarvirus RNA sequences and are important for CP binding (8-11). We previously identified a 39-nucleotide minimal high affinity AMV CP-binding site, consisting of the two terminal hairpins and their flanking AUGC nucleotides (nucleotides 843 to 881 in RNA4; i.e., AMV₈₄₃₋₈₈₁) (8,12,13) (fig. S1A). This fragment is competent to bind either full-length CP or a 26-amino acid peptide (CP26, fig. S1B) (13) representing the N-terminal RNA binding domain (14). The CP N terminus contains a Pro-Thr-x-Arg-Ser-x-x-Tyr (PTxRSxxY) RNA binding domain conserved among AMV and ilarvirus CPs (14). The arginine at position 17 is critical for both RNA binding and virus

*To whom correspondence should be addressed. E-mail: Lee_Gehrke@hms.harvard.edu (L.G.); James_Hogle@hms.harvard.edu (J.M.H.)

replication (14-16). Circular dichroism experiments suggest that the CP N terminus is unstructured in solution (17). Previous virus crystallization attempts required proteolytic cleavage of the AMV CP N terminus (18,19).

Crystals of the AMV N-terminal CP peptide CP26 in complex with 5-bromouridine-labeled AMV₈₄₃₋₈₈₁ RNA were grown in hanging drops by vapor diffusion. The structure was solved to 3.0 resolution with phases obtained by multiple-wavelength anomalous dispersion (table S1) (13). The cocrystal structure reveals that the RNA and peptides undergo cofolding events that substantially alter their structure from their unbound forms. The RNA forms two hairpins oriented at approximately right angles (Fig. 1A). Each of the RNA hairpins is extended by two base pairs formed between nucleotides from adjacent AUGC sequences (Figs. 1 and 2). Loop nucleotides could not be placed into the final model (fig. S2) (13). The two CP26 peptides each form an α helix with an extended N-terminal tail (Fig. 1A). Residues 12 to 26 are ordered in peptide 1, whereas residues 9 to 26 are ordered in peptide 2. The extreme N termini of both peptides remained flexible and were not visible.

A notable feature of the structure is that cofolding events promote base pairing between the AUGC repeats in the presence of CP. The U(844) and C(846) nucleotides pair with A(865) and G(867), respectively, at the base of duplex 1 (Figs. 1C and 2A). The C₀G pair is a Watson-Crick pair, whereas the A-U pair is a noncanonical UA reverse Watson-Crick pair. An identical interaction is seen at the base of duplex 2, which is extended by two base pairs when U(866) and C(868) hydrogen bond with A(878) and G(880), respectively (Figs. 1C and 2B). The formation of these base pairs introduces two types of distortion into the AMV RNA. First, the interhelical AUGC₈₆₅₋₈₆₈ (purple in Figs. 1 and 2) participates in base pairing interactions in both duplex 1 and duplex 2. This requires alternating bases of AUGC₈₆₅₋₈₆₈ to be oriented in opposite directions. Second, to form C₀G base pairs, G(867) and G(880) must be positioned upstream of A(865) and A(878) in their respective stacks, out of numerical order (Fig. 1B). The insertion of G(867) and G(880) between A(864) and A(865) or G(877) and A(878), respectively, introduces an unusual kink into the ribose-phosphate backbone (Fig. 1). In duplex 1, A(864) and A(865) are separated by the insertion of G(867) into the base stack. U(866) is flipped out of the base stack of duplex 1 and instead participates in reverse Watson-Crick base pairing at the base of duplex 2 (Fig. 2A). The same interaction is seen at the base of duplex 2 (Fig. 2B), where the distance between G(877) and A(878) is also increased to accommodate the insertion of G(880) between them to form a base pair with C(868).

Amino acids in the conserved RNA binding motif PTxRSxxY play key roles in stabilizing the AUGC base pairing interactions. Biochemical evidence suggesting the importance of arginine 17 (14) is consistent with its central position within the RNA protein-binding site. Both the G(867)-C(846) and G(880)-C(868) base pairs are displaced from the helical axis toward the minor groove, allowing the Arg¹⁷ sidechain guanidinium to insert into the base stack at the major groove. Arg¹⁷ is positioned to make polydentate hydrogen bonds with the G(867) (peptide 1) and G(880) (peptide 2) bases, thus extending the stacking interaction of the guanine bases with A(865) and A(878), respectively (Fig. 3A). The importance of arginine residues in RNA-peptide interactions was demonstrated previously in the human immunodeficiency virus Tat/Tar complex (20). The bottoms of both hairpins are capped by conserved tyrosine (Tyr²¹) sidechains from peptides 1 and 2. Between Arg¹⁷ and Pro¹⁴, both copies of the CP26 peptide transition from a helix to extended chain. This transition is mediated by a tight turn between Thr¹⁵ and Pro¹⁴, forming an L shape that cradles the extended RNA duplex (Fig. 1A). The L turn allows the N-terminal tail of peptide 2 to straddle a groove in the RNA surface formed by duplex 1 and duplex 2, and to make additional contacts between the RNA backbone of duplex 1 and Lys¹³ and Gln¹⁹ of peptide 2 (Fig. 1B). This conformation is facilitated by the conserved Pro¹⁴ and is stabilized by an exchange of hydrogen bonds between the sidechain hydroxyls and the mainchain amides of Thr¹⁵ and Ser¹⁸ within the PTxRSxxY motifs (Fig.

3B). The side chain of Ser¹⁸ occupies a pocket that provides an optimal fit. It forms hydrogen bonds with the phosphate oxygen of G(845) (peptide 1) or G(867) (peptide 2) and the amide nitrogen of Thr¹⁵ (Fig. 3B). Pro¹⁴ makes van der Waals contacts with the ribose ring and phosphate oxygen of G(845) (peptide 1) and G(867) (peptide 2) (Fig. 1B). The edge of the Pro¹⁴ sidechain makes additional favorable packing contacts with the ribose-phosphate backbone above and below the plane of the proline ring, contributing to the stability of the complex. Mutation of Pro¹⁴, Ser¹⁸, or Tyr²¹ diminishes CP26-AMV₈₄₃₋₈₈₁ RNA binding affinity and replication activity (21), underscoring the importance of these contacts for the stability of the RNA-protein complex.

The structure correlates well with biochemical and functional data. Disruption of either the CP-binding site on the RNA or the RNA binding site on the CP prevents formation of the complex and blocks RNA replication (12,15-17,22). The contacts related to the L turn, which were not predicted by amino acid sequence homology, provide a structural basis for the observed protection of nucleotides 847 to 850 during hydroxyl radical footprinting experiments (8,12). In vitro RNA selection experiments (10,11) and hydroxyl radical footprinting data (8, 14) suggested that peptides bind to the base of the hairpins without substantial loop interactions. The inter-AUGC base pairing contacts are consistent with previous chemical modification data and in vitro selection showing that modification of U(844) or C(846) diminished CP binding (12), whereas degenerate in vitro selection data show the strong conservation of nucleotide identity at the AUGC positions (11). The importance of the central AUGC and the dual hairpin contacts made in this region with peptide 2 (Fig. 1) was also anticipated on the basis of biochemical data revealing that peptide binding was abolished if tandem AUGC repeats separated the two hairpins (23).

The unusual inter-AUGC base pairing seen in the structure and the conservation of AUGC sequences among AMV and ilarvirus 3'UTRs (fig. S3A) prompted us to test whether CP can bind RNAs containing nucleotide substitutions that preserve base pairing, even though the substitutions are not strictly structurally equivalent. We changed all three AUGC sequences in AMV₈₄₃₋₈₈₁ to CGUA, GCAU, or UACG. None of these RNAs bound CP26 in mobility shift assays (fig. S3B) (13). The positions of G(867) and G(880) are likely to be critical for CP binding because of their close interaction with the critical Arg¹⁷ residue. An additional UACG mutant, designed to leave the position of the C0G base pair unchanged, also did not bind CP26. This observation is consistent with the proposal that the extended stacking of Arg¹⁷-G(867) and Arg¹⁷-G(880) with A(865) and A(878), respectively, contributes to the stability of the unusual base pair structure at the base of the stem loops.

Features of the AMV RNA-CP peptide complex may shed light on the mechanisms at work during the initial phases of virus replication. The structure suggests that the RNA assumes a highly ordered conformation in the presence of CP. The AUGC regions, single stranded and potentially flexible in the absence of CP (24), form base pairs when bound to CP26 and force the RNA to assume a more conformationally limited form. The recurring pattern of peptide interactions with duplex 1 and duplex 2 RNAs (Figs. 1B and 2) suggests that these types of contacts may also occur at other U-C-hairpin-A-G motifs throughout the 3'UTR (8) (Fig. 4A). We modeled CP binding along the entire length of the 3'UTR by aligning the AUGC regions of multiple copies of the AMV₈₄₃₋₈₈₁-CP26 structure, while ignoring the differing lengths of the hairpins. The resulting molecule forms a long rodlike structure with the hairpins projecting out from a central axis (Fig. 4, B and C). A transition from a more flexible to a more constrained 3' end upon CP binding agrees with a replication model suggesting that an RNA conformational change accompanies CP binding to the 3' ends of the genomic RNAs (25). Circular dichroism analyses and polyacrylamide gel electrophoresis experiments also suggest that the AMV 3'UTR RNA becomes more compact in the presence of CP (11,17). Sequential loading of CP along the length of the 3'UTR may specify switches between replication, translation, and

assembly processes. As a general theme in nonpolyadenylated positive-strand RNA viruses, the structural organization of the extreme 3' terminus (here, by binding of two CP subunits at the base of the last two hairpins) may present a single conformation that is recognized by replicase enzymes. Organization of the entire length of the AMV 3'UTR (binding of CP molecules at the base of all hairpins) may specify assembly functions through the formation of a rodlike structure to nucleate bacilliform particle formation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References and Notes

1. Bol JF, van Vloten-Doting L, Jaspars EM. *Virology* 1971;46:73. [PubMed: 5124242]
2. Jaspars, EMJ. *Molecular Plant Virology*. Davies, JW., editor. CRC Press; New York: 1985. p. 155-221.
3. Dreher TW, Bujarski JJ, Hall TC. *Nature* 1984;311:171. [PubMed: 6472477]
4. Chapman MR, Kao CC. *J. Mol. Biol* 1999;286:709. [PubMed: 10024445]
5. Koper-Zwarthoff EC, Bol JF. *Nucleic Acids Res* 1980;8:3307. [PubMed: 6160470]
6. Houwing CJ, Jaspars EM. *Biochemistry* 1982;21:3408. [PubMed: 6810924]
7. Quigley GJ, Gehrke L, Roth DA, Auron PE. *Nucleic Acids Res* 1984;12:347. [PubMed: 6320093]
8. Houser-Scott F, Baer ML, Liem KF Jr, Cai JM, Gehrke L. *J. Virol* 1994;68:2194. [PubMed: 8139004]
9. Reusken CB, Neeleman L, Bol JF. *Nucleic Acids Res* 1994;22:1346. [PubMed: 8190624]
10. Houser-Scott F, Ansel-McKinney P, Cai JM, Gehrke L. *J. Virol* 1997;71:2310. [PubMed: 9032367]
11. Rocheleau GA, Petrillo JE, Guogas LM, Gehrke L. *J. Virol* 2004;78:8036. [PubMed: 15254175]
12. Ansel-McKinney P, Gehrke L. *J. Mol. Biol* 1998;278:767. [PubMed: 9614941]
13. Materials and methods are available as supporting material on Science Online
14. Ansel-McKinney P, Scott SW, Swanson M, Ge X, Gehrke L. *EMBO J* 1996;15:5077. [PubMed: 8890181]
15. Yusibov VM, Loesch-Fries LS. *Virology* 1995;208:405. [PubMed: 11831728]
16. Yusibov V, Loesch-Fries LS. *Virology* 1998;242:1. [PubMed: 9501033]
17. Baer ML, Houser F, Loesch-Fries LS, Gehrke L. *EMBO J* 1994;13:727. [PubMed: 8313916]
18. Fukuyama K, Abdel-Meguid SS, Johnson JE, Rossmann MG. *J. Mol. Biol* 1983;167:873. [PubMed: 6876169]
19. Kumar A, et al. *J. Virol* 1997;71:7911. [PubMed: 9311881]
20. Calnan BJ, Tidor B, Biancalana S, Hudson D, Frankel AD. *Science* 1991;252:1167. [PubMed: 1709522]
21. Petrillo, JE.; Guogas, LM.; Gehrke, L. unpublished data
22. Reusken CB, Bol JF. *Nucleic Acids Res* 1996;24:2660. [PubMed: 8758992]
23. Laforest SM, Gehrke L. *RNA* 2004;10:48. [PubMed: 14681584]
24. Ansel-McKinney, P. thesis. Harvard University; 1996.
25. Houwing CJ, Jaspars EM. *Biochemistry* 1978;17:2927. [PubMed: 99164]
26. We thank S. Ginell and R. Zhang for support at beamline ID19 of the Advanced Photon Source (APS) and B. Appleton and T. Auperin for help with data collection. B. Eichman, T. Hollis, and E. Toth provided useful discussions and advice. Use of the Argonne National Laboratory Structural Biology Center beamlines at the APS was supported by the U.S. Department of Energy, Office of Energy Research, under Contract no. W-31-109-ENG-38. This work was supported by NIH awards GM42504 (L.G.) and AI20566 (J.H.) and by a fellowship from the Albert J. Ryan Foundation (L.M.G.). Structure factors and atomic coordinates have been deposited in the Protein Data Bank (accession number 1XOK)

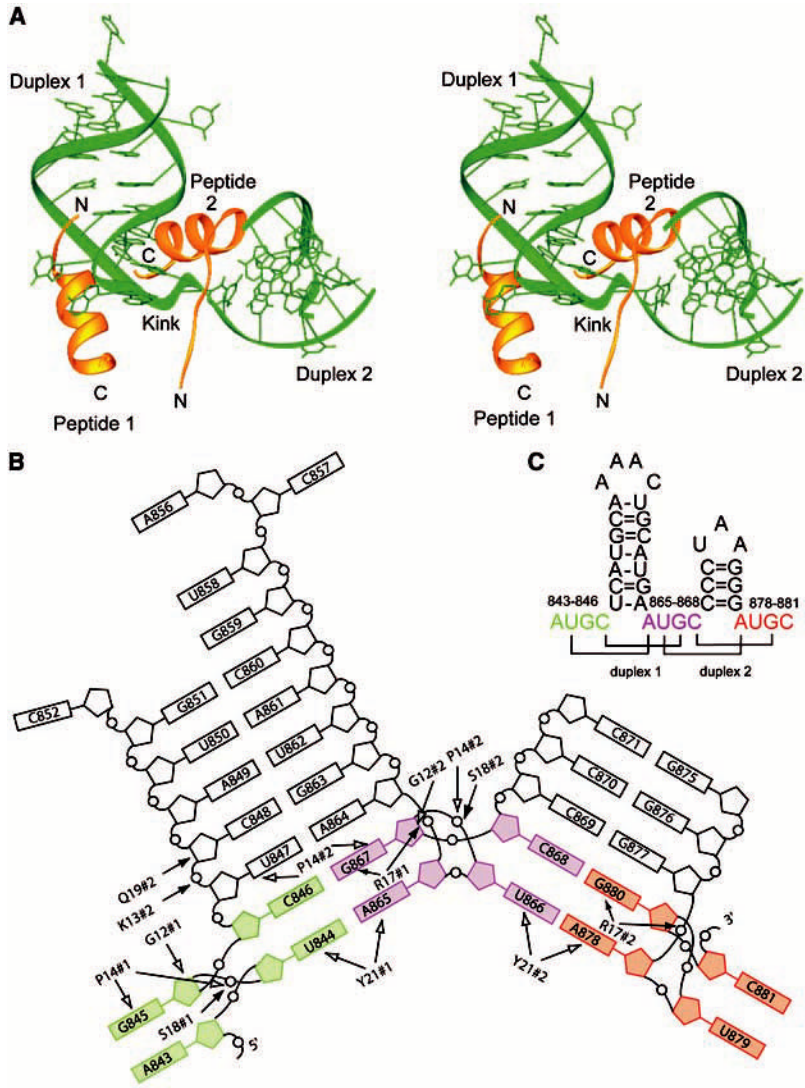


Fig 1. (A) Stereoview of the AMV RNA-CP complex. The RNA is shown in green and the CP26 peptides are shown in gold. (B) Schematic diagram of the AMV RNA and CP26 protein contacts. The AUGC sequences have been color coded. The 5' AUGC is colored green, the interhelical AUGC sequence is colored purple, and the 3' AUGC sequence is colored red. Hydrogen-bonding contacts between the RNA and peptide are indicated with filled arrows and van der Waals contacts are indicated by open arrows. #1, peptide 1; #2, peptide 2; G, Gly; K, Lys; P, Pro; Q, Gln; R, Arg; S, Ser; Y, Tyr. (C) Secondary structure of the 39-nucleotide minimal binding domain found at the 3' end of AMV RNAs, AMV₈₄₃₋₈₈₁. The AUGC sequences are color coded as in (B). Base pairs are indicated by brackets.

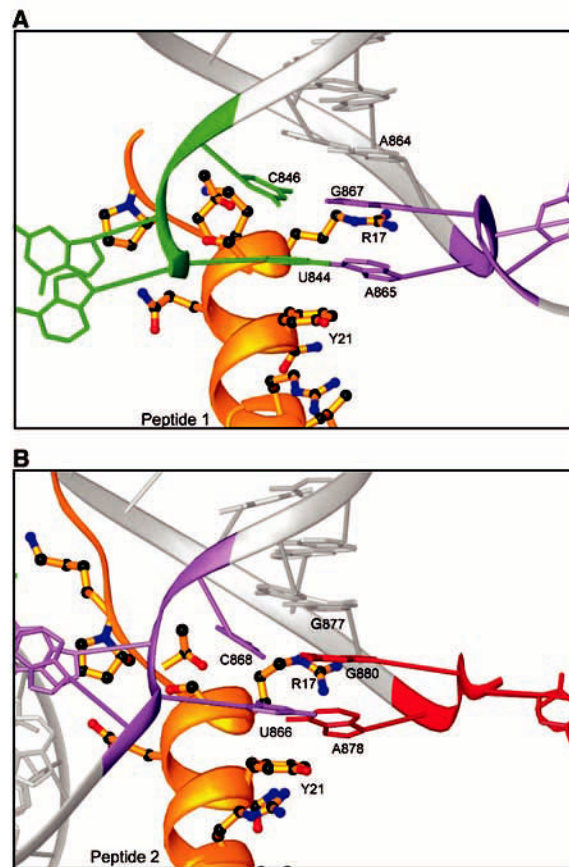


Fig 2. Formation of inter-AUGC base pairs. (A) RNA-peptide interactions at the base of duplex 1. The CP peptide a helix and sidechains are shown in gold. The RNA is shown in gray. The 5' AUGC sequence is colored green, and the interhelical AUGC sequence is colored purple. (B) Similar RNA-peptide interactions are seen at the base of duplex 2. The interhelical AUGC sequence is highlighted in purple and the 3' AUGC sequence is colored red. R, Arg; Y, Tyr.

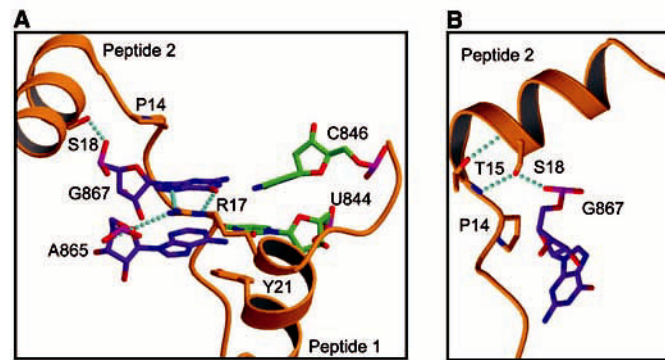


Fig 3. Interactions between the AMV RNA and CP26 peptides. (A) View of the two additional base pairs at the base of duplex 1 and their interaction with peptide sidechains. The carbon atoms of base pairing members at the base of duplex 1 from the 3' AUGC [U(844) and C(846)] are colored green, whereas members from the interhelical AUGC [A(865) and G(867)] are colored purple. Hydrogen bonds are indicated in cyan. (B) View of peptide 2 making the L-shaped turn. The carbon atoms of peptide 2 are colored gold, whereas those of G(867) are colored purple. P, Pro; R, Arg; S, Ser; T, Thr; Y, Tyr.

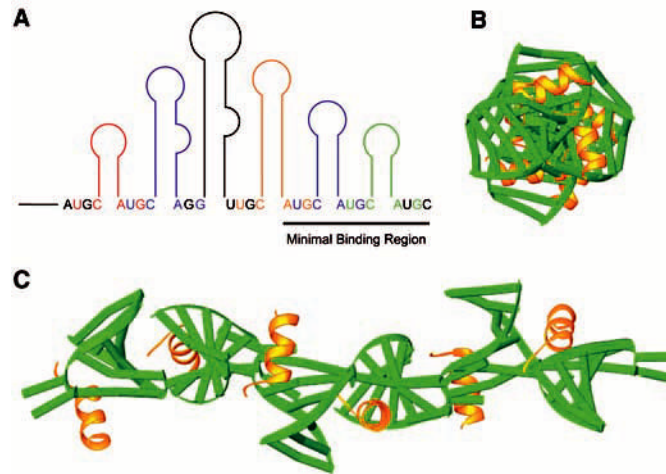


Fig 4. Proposed structure of the entire 3'UTR. (A) Diagram of the 3'UTR with color-coded U-C-hairpin-A-G CP-binding sites. (B) End-on view of the simulated 3'UTR. (C) View of the 3'UTR along its length. RNA is colored green; peptide is colored yellow.