

Origins of Life and the RNA World: Evolution of RNA-Replicase Recognition

Laura Guogas, James Hogle & Lee Gehrke

Department of Microbiology and Molecular Genetics, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

Abstract. Central to understanding the origin of life is the elucidation of the first replication mechanism. The RNA World hypothesis suggests that the first self-replicating molecules were RNAs and that DNA later superceded RNA as the genetic material. RNA viruses were not subjected to the same evolutionary pressures as cellular organisms; consequently, they likely possess remnants of earlier replication strategies. Our laboratory investigates how members of the RNA virus family Bromoviridae can have structurally distinct 3' end tags yet are specifically recognized by conserved replication enzymes. This work addresses the idea that 3' tRNA tails were functionally replaced in some viruses by an RNA-protein complex. These viruses may serve as a timeline for the transition from the RNA world to DNA and protein based life.

1. Introduction

Alfalfa mosaic virus (AMV) and ilarviruses are single-stranded positive-sense RNA viruses with three genomic RNAs and a subgenomic RNA (see Fig. 1). These viruses are unique among the Bromoviridae because, first, the viral RNAs lack the 3'-terminal transfer RNA-like structure important for replicase recruitment and second, the genomic RNAs are infectious only when the viral coat protein is present in the inoculum. The interaction of RNA with coat protein is not fully understood, but it may activate an RNA conformational change which tags the RNAs so that they can be recognized by the viral replicase (Baer et al. 1994). In order to more specifically define this interaction, we are in the process of determining a high resolution three-dimensional structure of the AMV RNA-peptide co-complex.

2. The RNA World and Genomic Tags

The term RNA World is used to denote a time prior to the evolution of proteins and DNA. In modern biology, RNA, DNA and protein are all inextricably linked. DNA contains the genetic information and encodes the proteins necessary for its own replication, while RNA is relegated to the role of intermediary. However, growing evidence supports the idea that this intermediary was in fact the first self

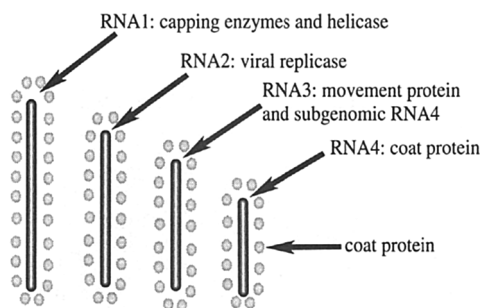


Figure 1. The four virions of AMV are separately encapsidated in bacilliform shaped particles. Virions 1-3 are infectious, however naked RNAs 1, 2 and 3 are not infectious and require the presence of the coat protein particle or the coat protein encoding RNA 4.

replicating molecule. RNA can encode genetic information and has the potential for self catalysis due to its ability to adopt a variety of tertiary structures.

The discovery of self-splicing ribozymes provided experimental impetus for the RNA World, validating the claim of RNA as catalyst (Kruger et al. 1982; Guerrier-Takada et al. 1983). It was also shown that the core of the ribosome is not protein, but an RNA which catalyzes peptide bond formation (Yusupov et al. 2001). Additionally, transfer RNAs, which play a crucial role in protein synthesis, may have first appeared as RNA replication tags (Weiner & Maizels 1999).

We generally think of transfer RNA (tRNA) as part of the protein synthesis machinery. It is possible however, that tRNA was first involved in replication, and that it later evolved to become a central player in protein synthesis. The genomic tag hypothesis, put forth by Weiner and Maizels, suggests that tRNA first evolved as a tail structure for tagging the ends of RNAs for replication. According to this hypothesis, all RNAs were originally tagged this way, yet the only remnants of this system are found in a group of RNA viruses.

Comparisons of the well conserved RNA polymerase sequence have been used to construct phylogenies of RNA viruses (Koonin & Dolja 1993). Based on this phylogeny, AMV is placed in the family Bromoviridae. Within that family, the viruses are divided into two groups, those that have tRNA-like ends (the Bromoviruses) and those that lack the tRNA like end (Alfamoviruses and Ilarviruses). It has become apparent that not all traces of the RNA world have vanished, and that close examination of modern systems may allow us to piece together the evolutionary role of this pivotal molecule.

The type member of the Bromovirus genus, Brome Mosaic Virus (BMV) has a 3' tRNA like tail which is required for replication. This tail serves as a recognition site for the replicase (Kim et al. 2000; Rao et al. 1989). In contrast, the Alfamo and Ilarviruses lack this tRNA tail, and instead rely on an RNA-protein interaction between the 3' untranslated region (UTR) of the viral genomic RNAs with the viral coat protein. This RNA-protein interaction is required in order to initiate replication (see Fig. 2). This suggests that the AMV RNA-protein interaction is functionally equivalent to the tRNA tail found in related viruses (Baer et al. 1994).

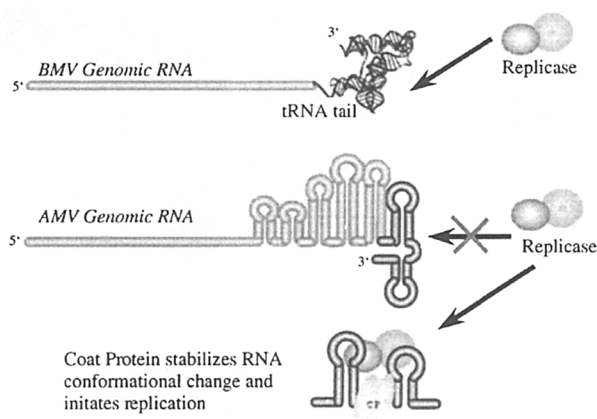


Figure 2. Brome mosaic virus genomic RNAs contain a tRNA like end required for initiation of replication (top). Alfalfa mosaic virus RNAs lack this tRNA like tail and instead require the binding of viral coat protein to initiate replication (bottom). The RNA-coat protein complex may serve as a functional equivalent to the 3' terminal tRNA structure found in numerous viral RNAs.

3. AMV Coat Protein: A Functional Homolog?

The proposed structure of the AMV 3' UTR, based on biochemical observations and computer modeling, contains seven hairpin loops flanked by (A/U)UGC motifs. The minimal length of RNA required for coat protein binding was identified as the 3' terminal 39 nucleotides (Houser-Scott et al. 1994; Ansel-McKinney & Gehrke 1998).

The coat protein (CP) N-terminus is necessary and sufficient for RNA binding. N-terminal peptides are also functionally equivalent to full length CP in activating the viral genome. Substitutions and deletions within the AMV coat protein revealed arginine 17 as crucial for binding. Sequence alignments centered around R17 revealed a putative RNA binding consensus sequence (Ansel-McKinney et al. 1996).

4. Structure Determination

We are in the process of determining a high resolution molecular structure in order to examine whether the RNA protein complex shares structural features with tRNA tails that might further define its role as an intermediate in the evolution of RNA-replicase recognition.

We have crystallized the RNA-peptide binding complex. Our crystals are small ($50 \times 100 \mu\text{m}$) and require that data be collected at a high energy synchrotron source in order to obtain maximum resolution. Diffraction studies were carried out at the Advanced Photon Source (APS), Argonne, Ill. Native data was collected to 3 Å resolution, sufficient for structure solution, however we have been unable to collect derivative data with sufficient phasing power.

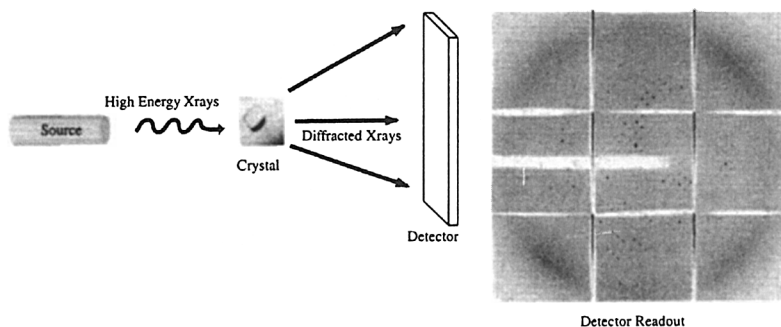


Figure 3. To acquire structural data, the RNA-peptide crystal was bombarded by high energy x-rays. The position and intensity of the x-rays scattered by the crystal are recorded on a CCD detector. The image above shows one frame of detector readout. To obtain a 3D structure, multiple images are recorded as the crystal is rotated.

Derivative crystals contain heavy metal compounds which introduce a strong electron scatterer. Comparison of native and derivative data sets is done to determine phase information necessary for structure solution. Likely the current derivatives contain metals bound to a disordered section of the molecule. Present work is focused on identifying derivatives.

References

- Ansel-McKinney, P., & Gehrke, L. 1998, *J. Mol. Biol.*, 278, 767
- Ansel-McKinney, P., Scott, S. W., Swanson, M., Ge, X., & Gehrke, L. 1996, *EMBO*, 15(18), 5077
- Baer, M. L., Houser, F., Loesch-Fries, L. S., & Gehrke, L. 1994, *EMBO*, 13(3), 727
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman S. 1983, *Cell*, 35, 849
- Houser-Scott, F., Baer, M. L., Leim, K. F., Cai, J., & Gehrke, L. 1994, *J. Virol.*, 68, 2194
- Kim, C. H., Kao, C. C., & Tinoco, I. Jr. 2000, *Nat. Struct. Biol.*, 7(5), 415
- Koonin, E. V., & Dolja, V. V. 1993, *Crit. Rev. Biochem. and Mol. Biol.*, 28(5), 375
- Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., & Cech, T. R. 1982, *Cell*, 31, 147
- Rao, A. L. N., Dreher, T. W., Marsh, L. E., & Hall, T. C. 1989, *Proc. Nat. Acad. Sci. USA*, 86, 5335
- Weiner, A. M., & Maizels, N. 1999, *Biol. Bull.*, 196(3), 327
- Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., & Noller, H. F. 2001, *Science*, 292, 883