Structural aspects of encapsidation signals in RNA viruses

Chen, S.C.

Chapter III

A new structure model for the packaging signal in the genome of group IIa Coronaviruses

Shih-Cheng Chen[1], Erwin van den Born[2], Sjoerd H.E. van den Worm[2], Cornelis W.A. Pleij[1], Eric J. Snijder[2], and René C.L. Olsthoorn[1]

[1]Leiden Institute of Chemistry, Leiden University,
[2]Department of Medical Microbiology, Leiden University Medical Center

(*Journal of Virology*, 2007, 81: 6771-6774)

Abstract

A 190-nt packaging signal (PS) located in the 3' end of ORF1b in Mouse hepatitis virus (MHV), a group IIa coronavirus, was previously postulated to direct genome RNA packaging. Based on phylogenetic data and structure probing, we have identified a 95-nt hairpin within the 190-nt PS domain, which is conserved in all group IIa coronaviruses, but not in SARS-CoV (group IIb) or group I and III coronaviruses. The hairpin is composed of 6 copies of a repeating structural subunit that consists of 2-nt bulges and 5-bp stems. We propose that repeating AA-bulges are characteristic features of group IIa PSs.

For many plus-stranded RNA viruses, a specific structural element within the genome RNA has been identified that serves as a starting point in the assembly of the viral nucleocapsid. For the Coronavirus family of RNA viruses (7), which includes important pathogens such as Severe Acute Respiratory Syndrome Coronavirus (SARS CoV), the process of genome encapsidation, which precedes the assembly of the enveloped virus particle, is poorly understood. Previous studies on Mouse hepatitis virus (MHV), a group IIa coronavirus (4), have led to the identification of a so-called packaging signal (PS) in the 3' end of the replicase ORF1b (7, 12). A region of 190 nt was sufficient to drive the encapsidation of defective interfering (DI) RNAs of MHV. Subsequently, a 69–nt fragment of this PS was found to be able, albeit less efficiently, to direct packaging of heterologous transcripts (3). A secondary structure for MHV PS was proposed (3) but its validity was never experimentally confirmed nor was it supported by an analysis of the closely related Bovine coronaviruses (2).

We have now used BLAST (1) to search for MHV PS homologs in currently available group IIa coronavirus sequences: Human coronavirus OC43 (HCoV, Genbank accession number NC_005147), Porcine hemagglutinating encephalomyelitis virus (HEV, NC_007732), Bovine coronavirus (BCoV, NC_003045), Human coronavirus HKU1 genotype A (NC_006577) and genotype B (AY884001). The only common sequence appeared to be a 18-nt stretch that was previously predicted to form the top of the putative BCoV PS (2). Subsequently, we folded a 150-nt region surrounding this motif in all group IIa coronaviruses using MFOLD (13) and adjusted the structure manually based on a phylogenetic analysis. The final model, shown in Fig. 1, is supported by at least 9 naturally occurring co-variations. It partly resembles the Cologna & Hogue structure but contains several previously unnoticed features such as the repeating AA or GA bulges at its 3' side. This feature follows from conserved AGC/GUAAU motifs (Fig. 1), which are repeated 4 times at 10–bp intervals. Another, though less well, conserved element is the 2-nt bulge at the 5' side of the stem-loop which is also repeated every 10 bps. An internal loop seems to divide the PS into symmetric upper and lower parts, in terms of orientation and number of conserved motifs. Interestingly, the ORF1b sequences of HCoV-OC43 and HEV have a GGU codon insertion compared to the other group IIa viruses (Fig. 1, stippled box). The extra three nt are part of the internal loop and maintain the secondary structure of the PS.

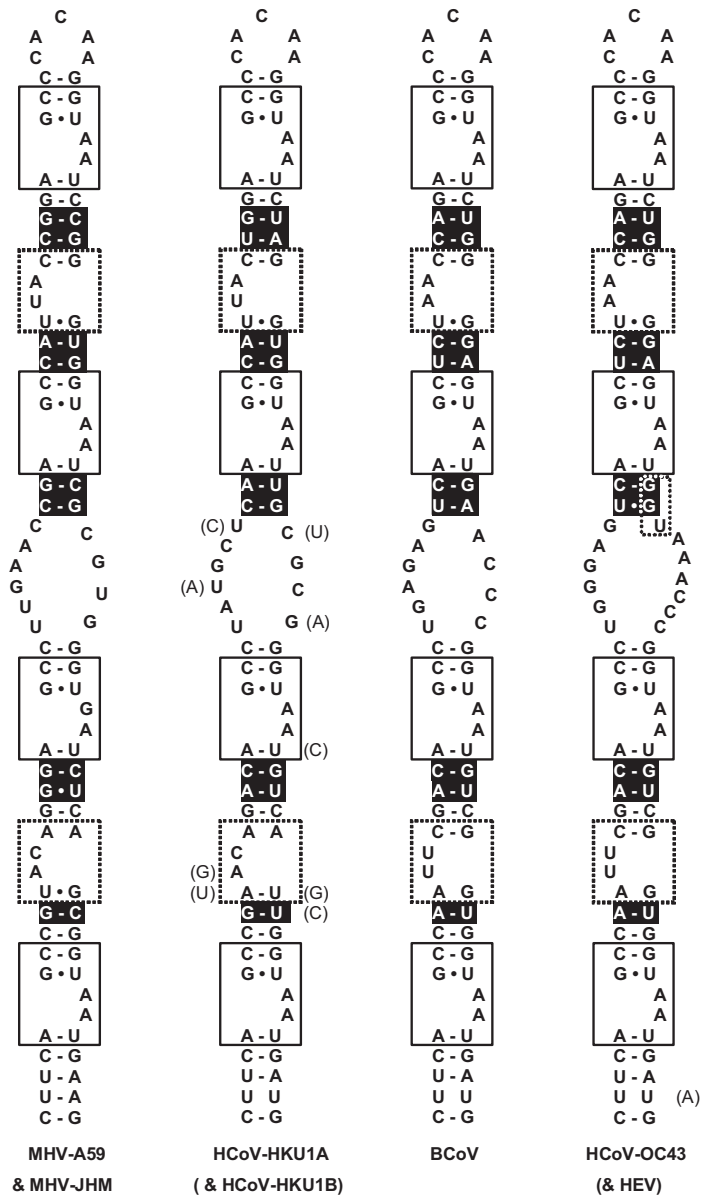

Figure 1. Highly conserved secondary structure of the PS of group IIa coronaviruses. Covariations are highlighted in reverse font. The conserved AGC/GUAAU motifs are boxed by solid lines while the other motif is boxed by dashed lines. The GGU insertion in the internal loop of HCoV-OC43 and HEV is indicated by the stippled box. Residues in brackets indicate the sequence of closely related species or isolates.

To support our model further, we PCR-amplified a DNA fragment containing the MHV-A59 PS using primers T7MH (5'-TTAATACGAC TCACTATAGACGATGTTATCTTCAGCCG-3') and MHV1 (5'-CCGATCGAGGTGTG AAAGAAG-3') using a MHV-A59 cDNA library as template. Subsequent transcription by T7 RNA polymerase (RiboMAX^{TM}, Promega, USA), yielded a

169-nt MHV PS-transcript which was subjected to chemical and enzymatic probing as described in Fig. 2, which shows RNAse T1, RNAse V1, RNase A, and RNAse S1 susceptible positions and the results of dimethylsulphate (DMS) and diethylpyrocarbonate (DEPC) probing. G-residues recognized by RNAse T1 were G23 and G77 (Fig. 2B) in agreement with their bulged-out positions in the helix. Residues C46, A47 and A48 were susceptible to RNAse S1 (Fig. 2B) and DMS (Fig. 2C), in accordance with their exposed positions in the loop. The pentaloop, internal loop and several bulges were also supported by RNAse A digestions (Fig. 2C) and by DMS, DEPC (Fig 2C and D) and lead(II) probing (data not shown). Interestingly, A5, A13, A17, A29 and A41, which are supposedly involved in double-stranded regions, were susceptible to DMS probing at room temperature (Fig. 2C) but also at 0ºC (Fig. 2D, white arrows). Although this may point to "breathing" of these base pairs due to their location at the end of a helix, the absence of DEPC modification of these particular residues (Fig. 2D) strongly suggests the formation of non-Watson Crick base pairs involving N7 rather than N1 of adenosine.

The absence of strong cuts in the internal loop would suggest it to be more structured than presented here. In fact, one can draw 3 additional base pairs (dashed lines in Fig 2A): U21-G73, A24-U72, and C25-G71, creating another bulge at the 5' side of the hairpin that would complete the symmetry of alternating bulges. This is also supported by CMCT probing which modified U22 and G23, but not U21 (data not shown).

We were unable to detect RNAse V1 cuts in this transcript, although the enzyme was active on control RNAs (data not shown). Lack of V1 cuts could be due to the relatively short stem regions of just 5 bp, or to steric hindrance by the AA bulges (6). Altogether, the probing data were in good agreement with the MHV PS model based on structure prediction and phylogenetic analysis (Fig. 1). Preliminary NMR data on a synthetic RNA corresponding to bases 37-59 is in agreement with the proposed structure (data not shown).

For comparison, Fig. 3A shows the secondary structure of the 69-nt core PS originally proposed by Fosmire *et al*. (3). Our proposed structure comprises these 69 nt and also falls within the original 190-nt PS identified by Makino *et al*. (7) and van der Most *et al*. (12). Fosmire *et al.* (3) used mutagenesis of the 69-nt PS to identify structural elements important for packaging. However, they obtained some unexpected results that were difficult to explain on the basis of their structure model. We have tried to reconcile these data with our revised PS model. For example, mutations (FB3-FB6) that disrupted the secondary structure in the Fosmire model would also strongly interfere with the folding of the "AA-bulge" hairpin identified in this study, and not surprisingly, interfere with DI RNA packaging. However, stem inversion mutants, FB7 and FB8, which were designed to preserve the secondary structure of the Fosmire model (Fig. 3A) but would disrupt the "AA-bulge" hairpin (Fig. 3B), showed a considerable reduction in packaging. Likewise, of three mutants lacking only 3 bps at the

bottom of the stem according to the Fosmire model and exhibiting equal thermodynamic stability, two (FB12 and 14), showed significant packaging reduction while one (FB13) showed nearly wild type efficiency.

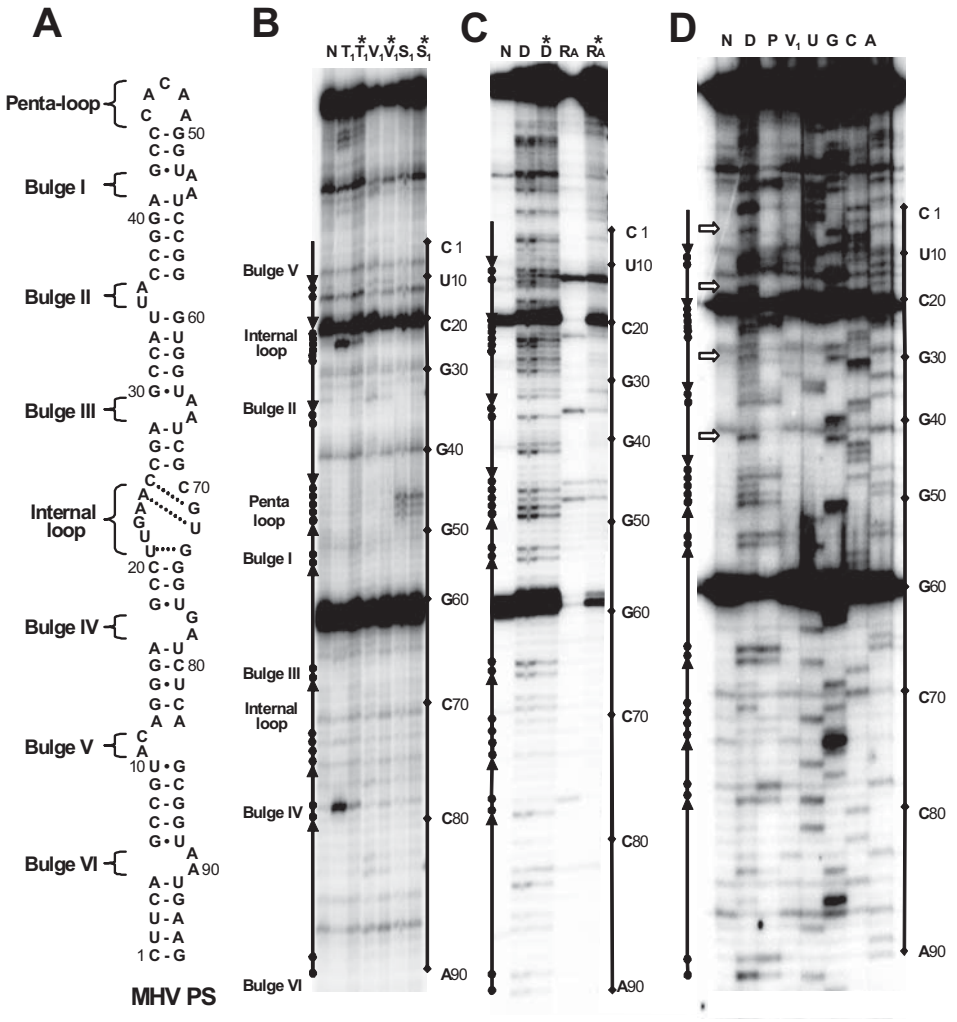

Figure 2. Structure probing of the MHV PS. Secondary structure model of the MHV PS was proven by structure probing. (A) The nucleotides are numbered from the 5'-end of the first cytosine in the PS, and bulges, internal-loop, and penta-loop were indicated. Alternative base parings of internal-loop are represented with dot-line. (B, C and D) Structure probing was done by 0.1 µg aliquots of RNA transcripts treated with nuclease-free water (N), 0.0001U of RNAse T1 (T_1), 0.0001U of RNAse V1 (V_1), 4U of RNase S_1 (S_1), 10 pg of RNAse A (R_A), 0.01% (v/v) DEP (P) and 0.05% (v/v) DMS (D) for 20 min at room temperature in a total volume of 50 µl (DMS and DEP probing shown in Fig.2D is done on ice). The incubation buffer contained 10 mM Tris, 100 mM KCl, 10 mM $MgCl_2$ and, optionally, 10 mM of zinc acetate (indicated by the asterisk above these lanes), pH 6.0. Primer extension reactions were carried out with 0.01 µg of treated transcripts, 0.5 µl of 0.1 mM MHV1 primer, 1 µl of 5 mM dGAT, 1 µl of 25 µM dCTP, 0.1 µl of α-^{32}P labeled dCTP (10 mCi/ml), 1 µl of 5X RT Buffer (Promega, USA), and 20 U of M-MLV reverse-transcriptase (Promega, USA). Susceptible nucleotides were identified by a reference sequencing ladder (not shown in B and C).

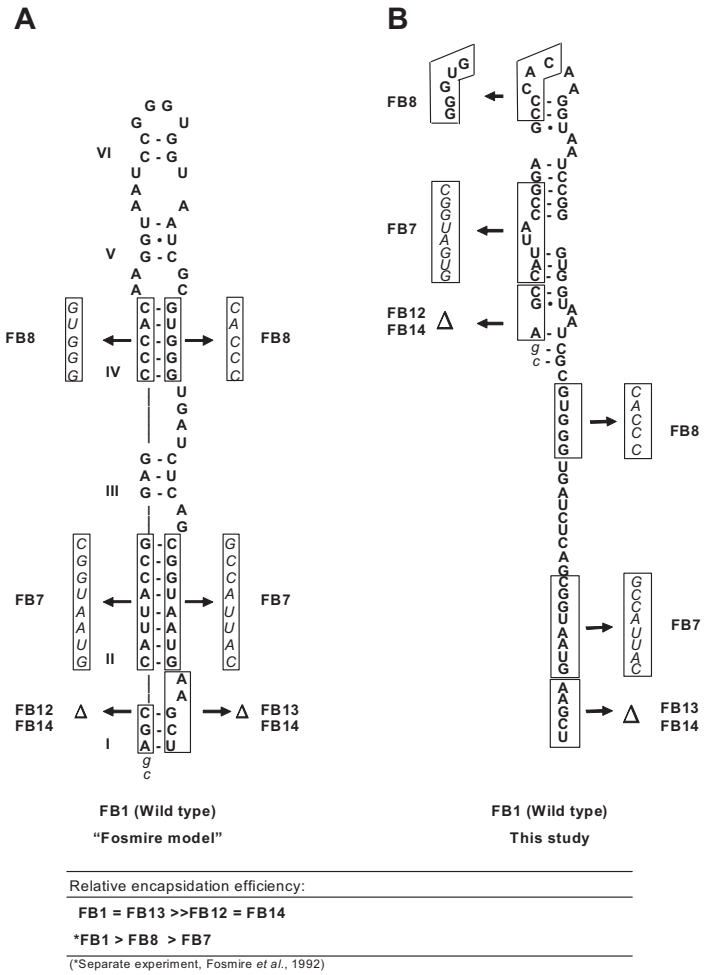

Figure 3. Predicted effects of mutations on the structure of the 69-nt MHV PS. Structure of the wild type 69-nt MHV PS: (A) according to Fosmire *et al*. (1992) and (B) redrawn to the new model. The boxed residues indicate the sequences of FB series mutants, FB 7-8 and 12-13, which were constructed by Fosmire *et al*. (1992), whereas Δ stands for deletion. The two extra residues (cg) at the 5'-end were vector-derived and stabilize the structure shown in B.

However, according to our model, mutant FB13 adopts the same secondary structure as wild type (FB1), while in mutants FB12 and 14 the stem is truncated in a manner that likely allows the formation of only one of the AA-bulges (Fig. 3B). Thus, reconciling these data with our model suggests that the 2-nt bulges at either side of the hairpin are important elements for the RNA encapsidation of MHV and presumably of other group IIa viruses in which these bulges are conserved. Finally, we note that the wild type 69-nt PS used by Fosmire *et al*. (3) can merely form the upper half of the "AA-bulge" hairpin and probably adopts the structure proposed by these authors most of the time. However, in the initially identified 190-nt PS domain the entire "AA-bulge" hairpin can be expected to form, as was also demonstrated by structure

probing of our related 169-nt fragment. This discrepancy in conformations may well explain the inability of the 69-nt PS to interact with the MHV N protein (9) and its 5-fold lower packaging efficiency (3).

The role of N protein in the encapsidation process is currently debated as virus-like particles could be obtained in the absence of N protein (10) and a prominent role for M protein has been suggested instead (11). Also, Enjuanes and coworkers have recently shown that TGEV N protein is a non-specific RNA-binding protein that acts as an RNA chaperone (14). Indeed, we observed that purified MHV N protein is able to bind RNA containing a minimal PS (nts C27-G69) but also tRNA (data not shown). Thus, how PS interacts with different structural proteins and initiates virion assembly is still poorly understood. The improved knowledge of the secondary structure of the MHV PS addressed in this study hopefully leads to a better insight into the encapsidation process for this group of coronaviruses.

Recently, Hsieh *et al*. (5) have proposed that a stem-loop structure similar to that proposed by Fosmire *et al*. (3) can be found in the 3' part of SARS-CoV ORF1b. They showed that this region could promote encapsidation of a reporter RNA into virus-like particles. However, we were unable to identify a hairpin in this region that conforms to the model shown in Fig. 1. In fact, preliminary data suggests that the SARS-CoV encapsidation signal is group I-like (Olsthoorn, unpublished data). *In vivo* studies to support this hypothesis are currently in progress.

References

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403-410.
2. **Cologna, R., and B. G. Hogue.** 2000. Identification of a bovine coronavirus packaging signal. J. Virol. **74:** 580-583.
3. **Fosmire, J. A., K. Hwang, and S. Makino.** 1992. Identification and characterization of a coronavirus packaging signal. J. Virol. **66:**3522-3530.
4. **Gorbalenya, A. E., E. J. Snijder, and W. J. M. Spaan.** 2004. Phylogeny of SARS coronavirus: toward consensus. J. Virol. **78:**7863-7866.
5. **Hsieh, P.-K., S.-C. Chang, C.-C. Huang, T.-T. Lee, C.-W. Hsiao, Y.-H. Kou, I-Y. Chen, C.-K. Chang, T.-H. Huang, and M.-F. Chang.** 2005. Assembly of Severe Acute Respiratory Syndrome Coronavirus RNA Packaging Signal into Virus-Like Particles Is Nucleocapsid Dependent. J. Virol. **79:**13848–13855.
6. **Lowman, H. B., and D. E. Draper.** 1986. On the recognition of helical RNA by cobra venom V_1 nuclease. J. Biol. Chem. **261:**5396-5403**6.**
7. **Makino, S., K. Yokomori, and M. M. C. Lai.** 1990. Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal. J. Virol. **64:**6045-6530.
8. **Masters, M. S.** 2006. The molecular biology of coronaviruses. Adv. Virus Res. **66:**193-292.
9. **Molenkamp, R., and W. J. M. Spaan.** 1997. Identification of a specific interaction

between the coroavirus mouse hepatitis virus A59 nucleocapsid protein and packaging signal. Virology. **239:**78-86.

10. **Narayanan K., C.-J. Chen, J. Maeda, and S. Makino**. 2003. Nucleocapsid-Independent Specific Viral RNA Packaging via Viral Envelope Protein and Viral RNA Signal. Virol. **77**: 2922-2927
11. **Narayanan, K., and S. Makino.** 2001. Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging. J. Virol. **75:**9059-9067
12. **van der Most, R. G., P. J. Bredenbeek, and W. J. Spaan.** 1991. A domain at the 3' end of the polymerase gene is essential for coronavirus defective interfering RNAs. J. Virol. **65:**3219-3226.
13. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:** 3406-3415.
14. **Zúñiga, S., I. Sola, J. L. Moreno, P. Sabella, J. Plana-Durán and L. Enjuanes.** 2007. Coronavirus nucleocapsid protein is an RNA chaperone. Virology. **357**:215-27.