Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

seminars in **VIROLOGY**, Vol 5, 1994 : pp 39-49

cis-acting genomic elements and *trans*-acting proteins involved in the assembly of RNA viruses

Sondra Schlesinger, Shinji Makino† and Maxine L. Linial‡*

There is now considerable evidence that a specific site (or sites) in the genome of an RNA virus interacts with a viral protein to initiate the assembly of the virus ribonucleoprotein or nucleocapsid. We describe the progress that has been made in defining these elements for a number of different viruses: the togavirus, Sindbis virus; the coronavirus, mouse hepatitis virus; influenza A virus; several retroviruses; and the hepadnavirus, hepatitis B virus. The importance of cis-acting elements in packaging has been established for all of these viruses. For Sindbis virus, specificity in the binding of the RNA element to a region of the viral capsid protein in vitro has also been demonstrated.

Key words: packaging signals / RNA-protein binding / RNA structures

THE FOCUS OF this chapter is on the initial event in the assembly of RNA viruses—the interaction between a viral RNA genome and protein leading to the formation of a ribonucleoprotein particle. Specificity in the assembly process was first observed over 30 years ago in the *in vitro* assembly experiments carried out with tobacco mosaic virus (TMV). The efficiency of reconstitution using the TMV protein and TMV RNA was much greater than with other viral RNAs. At that time Caspar noted that 'the RNA may contain the information necessary for it to link up with its own protein, as well as the code for the sequence of this protein'.[1] There are also examples which suggested specificity of assembly *in vivo*. The positive-strand RNA alphaviruses and several families of positive-strand RNA plant viruses synthesize large quantities of both genomic and subgenomic RNAs in infected cells. Usually only the genomic RNA, and not the subgenomic RNA, is packaged. Initially, it was possible to imagine that selectivity in packaging was a consequence of the overall RNA structure which might, for instance, contribute to the stability of the particles formed. Now, however, an increasing number of viral RNA genomes that contain a specific sequence which appears to be important for encapsidation of that RNA—an encapsidation or packaging signal—have been identified.

The assembly of a viral nucleocapsid or ribonucleoprotein (RNP) is an essential step in the formation of a virus. For many viruses, this step may require only the interaction between the genomic RNA and the viral capsid or coat protein. (See also the chapter by J.M. Fox *et al*, this issue.[2]) The capsid protein can play several roles in the life cycle of the virus. It serves as a coat to protect the genome of nonenveloped viruses during the time the particles exist in an extracellular environment. It also acts as the mediator to bind the virus to a susceptible cell. For enveloped viruses, the formation of the infectious virus requires budding of the ribonucleoprotein through cellular membranes.[3] For these viruses, the capsid protein must also interact with other viral proteins in the process of forming infectious virions. For some enveloped viruses, in particular, retroviruses and negative-strand RNA viruses, the RNP is composed not only of the genome and a capsid protein, but also other proteins required for transcription and, for the negative-strand viruses, replication of the RNA. Some of these proteins may also be important in the assembly of the RNP. We are emphasizing the importance of RNA/protein interactions in the initial steps of the assembly process, but these interactions may also be critical in disassembly or uncoating events. In infected cells, the genomic RNA of positive-strand RNA viruses (with the exception of retroviruses) must be able to bind the ribosomes to initiate translation and some mechanisms must exist to permit the ribosome access to the viral RNA.

We discuss here only a few of the viruses for which encapsidation signals have been identified, but they do span several very different virus families. They illustrate examples of specificity and also point out

*From the *Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110-1093, †Department of Microbiology, University of Texas, Austin, TX 78712 and ‡Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98104, USA*

1044-5773/94/010039 + 11$8.00/0

some of the complexities and problems in defining the RNA and protein requirements for encapsidation. Additional examples are covered in other chapters in this issue.

Alphaviruses

Alphaviruses are positive-strand RNA enveloped viruses belonging to the Togaviridae family.[4,5] The genomic 49S RNA serves as the mRNA for translation of a polyprotein precursor for the nonstructural proteins encoded in the 5′ two-thirds of the viral genome. These proteins are required for the replication and transcription of viral RNA. The viral structural proteins are also translated as a polyprotein, but from a subgenomic 26S mRNA identical in sequence to the 3′ terminal one-third of the 49S RNA. The viral capsid protein is translated first and acts as an autoprotease to cleave itself from the nascent polyprotein, leaving the precursor for the envelope proteins. The latter are synthesized and processed in the rough endoplasmic reticulum and are transported through the Golgi network to the plasma membrane of the infected cell. As mentioned above, there is discrimination in packaging, only the genomic 49S RNA is found in virus particles. The 26S mRNA is not packaged.

Weiss *et al* identified a region of the genome that interacts specifically with the viral capsid protein using two different approaches.[6] They first defined a region in the Sindbis virus genomic RNA encompassing nucleotides 684 to 1253 that bound to purified capsid protein immobilized on nitrocellulose. They then demonstrated the importance of this region in encapsidation by showing that defective interfering (DI) RNAs of Sindbis virus containing this region of the genome are packaged, but a DI RNA lacking this region is not. This region lies within the coding region of one of the nonstructural protein genes (nsP1) (Figure 1A). More extensive deletion mapping demonstrated that a fragment containing only 132 nucleotides (from nucleotide 945 to 1076 in the Sindbis virus genome) retained full binding activity (Figure 1B).

The binding of Sindbis virus RNA to the capsid protein triggers the protein/protein interaction that must occur for formation of the viral nucleocapsid. The capsid polypeptide contains 264 amino acids which can be divided into two domains based on the amino acid sequence.[7,8] The basic N-terminal domain consists of about 115 residues. The C-terminal domain has the autoprotease activity and is conserved among alphaviruses. Sequence analysis and mutagenesis studies[9,10] suggested that the capsid protein is a chymotrypsin-like serine protease and this has now been confirmed by X-ray diffraction data.[11] The capsid exists in the crystal structure as a dimer with the monomer-monomer contacts at residues 185 to 190 in a β-strand and also residue 222. The structure of the C-terminal domain (amino acids 114-264) was determined and is similar to the structure of mammalian serine proteases of the chymotrypsin family. In contrast, the N-terminal 113 residues remain largely unstructured in the crystals. The basic nature of this region and its apparent flexibility are reminiscent of the structure of the coat protein of several plant RNA viruses.

Two different studies have identified a region in this domain that interacts with RNA. Geigenmüller-Gnirke *et al* analyzed a variety of deletion mutants of the capsid protein for their ability to bind to the Sindbis virus encapsidation signal.[12] They identified a 32 amino acid region (amino acids 76-107) that is essential for RNA binding. They also demonstrated that a 68 amino acid peptide (lacking residues 11-74 and 133-264) retains almost the same binding activity as the intact protein. The assay measured the binding of *in vitro* translated proteins to RNA based on the migration of the protein with RNA during electrophoresis in an agarose gel. Wengler *et al*[13] took a different tack to identify the region of the capsid protein that binds to ribosomes. They[13] and Singh and Helenius[14] have proposed that the binding of alphavirus nucleocapsids to ribosomes triggers the disassembly process. Wengler *et al* showed that a proteolytic fragment of the capsid protein purified from virions containing amino acids 94 to 264 binds to ribosomes, but a fragment extending only from residue 106 to 264 lacks this activity.[13] The stretch of amino acids between residues 94 and 106 lies within the region involved in the binding of the capsid polypeptide to Sindbis virus RNA. This overlap suggests that at least some of the amino acids are important in the assembly and also in the disassembly of the nucleocapsid.

Coronaviruses

The study of coronavirus nucleocapsid assembly and its regulation is still in its infancy. To date, the data in this area centers on the identification of the putative packaging signal of the prototypic coronavirus, mouse hepatitis virus (MHV). Coronaviruses are enveloped RNA viruses that contain a 27 to 32 kb single-stranded positive-sense

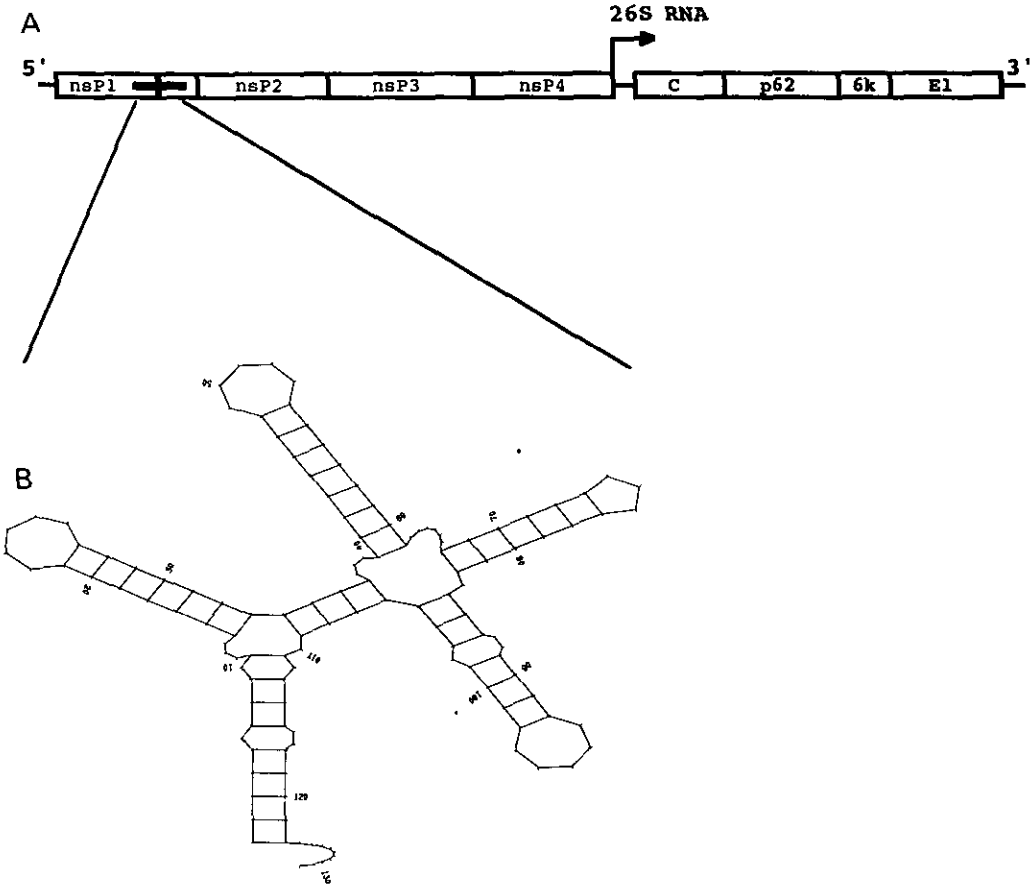

Figure 1. A. The RNA genome of Sindbis virus. The open boxes are the coding regions; nsP refers to the nonstructural proteins, C, p62, 6k and E1 are the structural proteins; C is the capsid protein. The thin lines represent noncoding regions. The horizontal arrow refers to the start of the 26S subgenomic RNA and indicates the location of the promoter for the transcription of this RNA. The cross in nsP1 is the location of the packaging signal. B. The optimal secondary structure of the 132 nucleotide capsid binding region located from nucleotides 945 to 1076 in the Sindbis virus genome. Some of the single-stranded constraints on this structure are based on chemical modifications (B. Weiss, S. Schlesinger, submitted for publication).

genomic RNA.[15-17] The genome interacts with the nucleocapsid (N) protein to form a helical nucleocapsid structure.[18] In infected cells there are six to eight subgenomic mRNA species and one genomic-sized mRNA-mRNA 1.[19] These mRNAs form a 3'-coterminal nested-set. All coronavirus mRNA 5'-ends originate with a 60 to 80 nucleotide-long leader sequence.[20,21] MHV efficiently packages only the genomic RNA.[16] Another coronavirus, transmissive gastroenteritis virus, does contain a small amount of subgenomic RNA species.[22] It is not clear if this represents selective RNA packaging.

Experiments using MHV defective interfering (DI) RNAs provided the first evidence that a *cis*-acting RNA element (a packaging signal) is necessary for the packaging of coronavirus RNA. The MHV strain JHM has several naturally occurring DI RNAs, not all of which are packaged. A 3.6 kb-long DI RNA, called DIssF, is packaged,[23] whereas others, both larger and smaller than DIssF, are not efficiently packaged.[24] This finding suggested that the size of MHV RNA does not determine packaging efficiency. Instead, some RNA sequence or structure in the genomic RNA must be important for MHV RNA encapsidation.

DIssF contains 1400 nucleotides that are not present in another DI RNA, DIssE, a 2.2 kb DI RNA that is packaged inefficiently.[23,25] Evidence that the packaging signal lies within these sequences came from insertion of these DIssF-unique sequences into a DIssE cDNA construct; the *in vitro*-transcribed RNA from this hybrid construct is efficiently packaged into viral particles.[23] In another study,

van der Most *et al* showed by deletion analysis of a DI RNA derived from another strain of MHV, MHV-A59, that the packaging signal maps within 650 nucleotides located near the 3′ region of gene 1.[26] This 650 nucleotide region is included within the 1400-nucleotides identified in the MHV-JHM DI RNA.[23,26] Deletion analysis of DIssF located the packaging signal to a 61 nucleotide sequence—a sequence that is found 1381 to 1441 nucleotides upstream of the 3′ end of gene 1 in the MHV genomic RNA (Figure 2).[22] Mutational analysis of this region suggests that the RNA secondary structure is important for function.[27] It is not known if this 61 nucleotide region is sufficient for packaging of MHV RNA and further studies with DI RNAs should determine if this sequence (or structure) is necessary and sufficient for packaging of MHV RNA.

The protein or proteins that bind to the MHV packaging signal have not been identified. Although the MHV N protein is a likely candidate, it does bind to the leader sequence of all MHV subgenomic mRNA species[28] and only mRNA 1, not the subgenomic mRNAs, forms a nucleocapsid structure in infected cells.[29] N protein also binds to non-MHV RNA.[30] Therefore, the binding of N protein to leader sequences does not seem to be sufficient for specific encapsidation of genomic RNA and some other factor must be required for discrimination in packaging.

Influenza viruses

The influenza A virus genome consists of eight single-stranded RNA segments of negative polarity. An infectious virion contains at least eight unique helical ribonucleoproteins (RNPs) composed of an RNA molecule, the RNA polymerase proteins, and the nucleoprotein (NP). In contrast to the paramyxovirus RNP, the RNAs in the influenza virus RNPs are partially susceptible to RNase digestion,[31,32] indicating that at least some regions are exposed on the surface of RNP structure.[33]

Palese and his colleagues described the location of an influenza virus *cis*-acting packaging signal.[34] They constructed an influenza virus cDNA in which the viral termini were maintained but the chloramphenicol acetyltransferase (CAT) gene replaced the viral NS gene-coding sequence. This cDNA was transcribed *in vitro* into RNA, which was then reconstituted with purified viral proteins to form a viral RNP. The RNP was transfected into cells which had also been infected with influenza virus. CAT activity was detected in these cells, but more importantly, it was also detected in cells infected with virus released from the transfected/infected cells. This result supports the conclusion that the signals for transcription, replication and packaging of an RNA are located in the 22 5′-terminal and the 26 3′-terminal nucleotides of the influenza A virus RNA.[34] Influenza virus genomic RNAs exist in a circular conformation that is based on a terminal panhandle of approximately 15 base pairs (Figure 2C).[35] It is possible that this terminal panhandle structure plays a role in influenza virus RNA packaging.[35]

The proteins required for packaging of influenza virus have not been identified. Influenza virus NP protein can bind any RNA longer than 15 nucleotides.[36] The RNA polymerase binds approximately 12-15 nucleotides at the 3′ end of each influenza virus RNA.[37] As described above, short stretches of nucleotides at both termini of influenza virus RNAs are sufficient for RNA packaging and contribute to the panhandle structure. It may be the RNA polymerase that confers the specificity to packaging of influenza virus RNAs.

One interesting question about the packaging of influenza virus RNAs is how a complete set of each of the RNAs is selected during assembly of influenza virus. Specific selective packaging of the full complement of eight RNAs into each particle would require complex molecular interactions and recognition signals. Although several observations support the selective packaging mechanism,[38-40] a number of observations are consistent with random packaging of more than eight RNAs. Estimates of the ratio of infectious to noninfectious particles based on calculations of random packaging are compatible with published values for the infectivity ratio of virus particles.[41,42] A heterozygotic influenza virus with two copies each of RNA3 and RNA6 was reported.[43] Furthermore, an influenza virus containing nine different RNA segments rather than eight RNA species was identified.[44] Currently the model that assumes influenza virus forms infectious virus particles via random packaging of more than eight viral RNA segments seems to be in favor.

Retroviruses and hepadnaviruses

Specific encapsidation of genomic molecules is critical to the survival of all viruses. In the case of viral

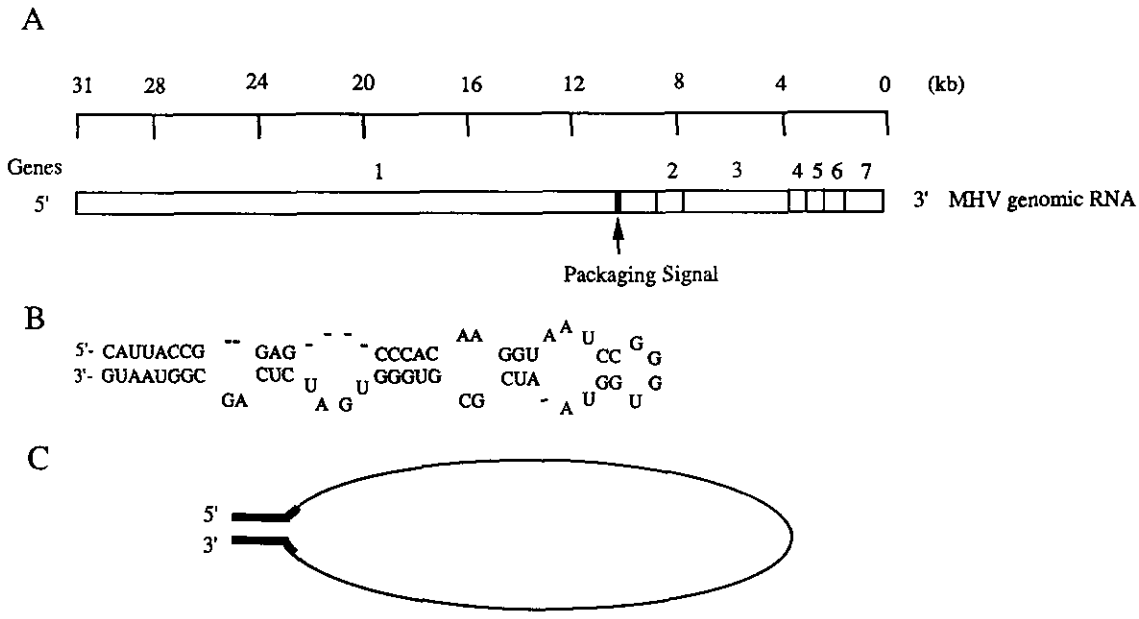

Figure 2. Location and structure of coronavirus and influenza virus packaging elements. A. Location of MHV packaging signal. Genes 1 through 7 represent the seven genes of MHV. The 61 nucleotide packaging signal is shown as a black box. B. Predicted secondary structure of the 61 nucleotide MHV packaging signal. C. Proposed influenza virus RNA structure. The bold lines include the RNA packaging signal. The panhandle is approximately 15 base pairs.

elements which encode reverse transcriptase, sequestration of only genomic RNAs into protein complexes containing active reverse transcriptase activity is also vital for the genetic integrity of the host cell. Cellular mRNAs which are introduced into retroviruses in artificial situations can be reverse transcribed and integrated into the genomes of newly infected cells.[45,46] Promiscuous integration of nontranscribed cDNAs (as processed pseudogene-like structures) could disrupt normal gene function. In addition, many reverse transcriptase viruses do not kill the infected cells, and in some cases transform or immortalize them; thus allowing more chance of such events to occur. Thus for these viruses, absolute packaging specificity is particularly important.

Both retroviruses and hepadnaviruses are enveloped viruses whose genomes are plus-strand RNAs. Both viral groups encode reverse transcriptase and synthesis of a cDNA copy of their genomes is central to their replication. These viruses also encode both genomic and subgenomic RNAs, but package only full length genomic species. In the case of the retroviruses, proviral DNA is synthesized early after viral infection following partial uncoating of the particle. Two identical (+) strand 8-11 kb RNA genomes are found in each particle; reverse transcriptase is associated with the genomes complexed with the tRNA primer at the primer binding site located near the 5′ end of the RNA. Hepadnaviruses have evolved an entirely different time course of reverse transcription. Reverse transcription occurs late in infection prior to viral release such that the encapsidated genome is a partially double-stranded 3.2 kb circular DNA with the reverse transcriptase (P protein) covalently attached to the (−) strand. The retroviruses and hepadnaviruses have developed different strategies to ensure that only sequestered genomic RNAs are available for reverse transcription. In the case of retroviruses, reverse transcriptase is synthesized as a fusion protein with the product of the first open reading frame, the *gag* gene. Synthesis of Gag-Pol is less frequent than that of Gag and occurs via suppression of a stop codon or ribosomal frameshifting (reviewed in ref 47). The activity of the Gag-Pol fusion protein varies among the retroviruses; in some cases there is no activity until cleaved by the viral protease which is encoded within either *pol* or *gag*. Protease itself is only activated as a dimer and this only occurs when the concentration is high, in the viral particle after assembly and budding. Changes in the protease structure may also be required

(reviewed in ref 48). Only after viral maturation is Pol released and bound to genomic RNA. In a viral mutant which encapsidates cellular mRNAs, reverse transcription of these RNAs is efficient.[49] In retroviruses, cDNA synthesis does not occur until after infection and partial uncoating of the viral core, and cDNA is not commonly found within mature particles. In the yeast retrotransposons such as Ty1 and Ty3 which also synthesize a fusion Gag-Pol reverse transcriptase precursor, particles do not bud from the cell and particles are found to contain cDNA products as well as genomic RNA (reviewed in ref 50). Although not well characterized, non-LTR (long terminal repeat) retroposons such as LINE elements, appear to encode a potential open reading frame upstream of reverse transcriptase which could encode a Gag-like protein. When expressed as a fusion protein with Ty1 Gag, a human LINE-1 orf2 does encode functional polymerase.[51] All existing data suggests that in the case of retroviruses and retrotransposons, packaging of the Pol protein is through the Gag domain. In contrast, the reverse transcriptase of the hepadnaviruses (the P protein) is not synthesized as a fusion protein. P is translated inefficiently from full length RNA; the major translation product is the amino terminal core protein.[52] Binding of P to its own mRNA (also genomic length) at a specific sequence in *cis* during synthesis appears to be required for packaging and is the nucleation signal for particle formation;[53-55] recently reviewed in more detail in ref 56. This pathway ensures the specific packaging of both RNA and reverse transcriptase.

Thus for hepadna- and retroviruses, survival depends upon a requirement for genomic RNA to be marked with a unique element which can interact specifically with a viral protein and ensure its preferential encapsidation. This element has been called ϵ in the case of the hepadnaviruses and ψ or E in the case of the retroviruses (see Figure 3). The structure of ϵ has been more precisely determined than that of ψ. In the case of human hepatitis B, a 94 nucleotide sequence at the 5′ end of the genomic RNA is necessary and sufficient for encapsidation[54,57] (Figure 3C). The ϵ sequence is also found at the 3′ end of the genomic RNA but does not function as a packaging signal in that position.[54,55] The predicted structure of the hepatitis B ϵ region is a stem loop with an upper and lower portion separated by a 6 nucleotide bulge and topped with a 6 nucleotide loop (see Figure 3D). Chemical probing and mutational analysis suggest that the predicted structure is important in encapsidation.[57] It is interesting that recent evidence suggests that the same stem-loop structure is also where P protein first binds and is the template for synthesis of a short DNA primer which then translocates to the complementary repeated region at the 3′ end of the genome where DNA synthesis then continues. Thus packaging seems to be intimately associated with the ability of the virus to initiate reverse transcription.[58] Packaging and reverse transcription are clearly separable however since P protein lacking enzymatic activities still functions in encapsidation.[53] The viral core protein is also important for encapsidation and an arginine-rich region in the carboxy-terminus of the core protein plays a role in this process.[59]

The location and putative structure of the retroviral ψ sequence is not as clearly defined as ϵ. There are many differences in the *cis*-acting packaging signals among the viruses in which these have been most studied; the avian and murine oncoviruses, and human immunodeficiency virus (HIV). However, in all cases the packaging sequences have been located to regions near the 5′ end of the genomic RNA (Figure 3A and B). Such studies have employed both deletion analysis and transfer of sequences to heterologous RNAs (reviewed in ref 60). Several groups have used computer-assisted folding programs to generate potential structures and have used *in vitro* RNase probing to determine whether such structures are formed.[61-64] However thus far, the relationship of these *in vitro* generated RNA secondary structures to the minimal packaging sequences required *in vivo*, has not been determined. In the case of Rous sarcoma virus (RSV), an avian retrovirus, a 270 nucleotide sequence between the major 5′ splice site and the start of the *gag* gene has been found to be sufficient to allow encapsidation when placed 3′ of a heterologous RNA sequence.[65] More recent studies suggest that 192 nucleotides are sufficient in this assay (A, Yeo, M.L. Linial, unpublished observations). The 192 nucleotide ψ sequence does not contain the putative purine rich motifs hypothesized to be important for *in vitro* dimerization.[66] Although all retroviruses contain two genomic RNAs in a dimeric structure, whether or not dimerization precedes encapsidation is not yet known. It has been suggested that there may be two types of dimers, early dimers which are less stable to heat dissociation than those that form later and could involve RNA/protein interactions (W. Fu, A. Rein, submitted). In the case of mutant retroviruses

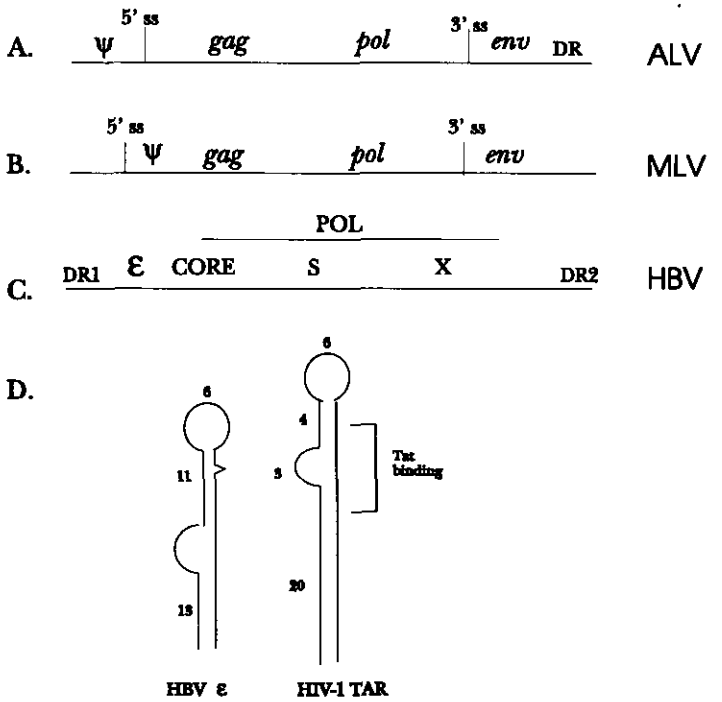

Figure 3. Location and structure of retro- and hepadnaviral packaging elements. A. Avian leukosis virus (ALV) RNA genome showing location of ψ and the three genes encoding structural proteins (*gag*), reverse transcriptase (*pol*) and glycoproteins (*env*). The major 5′ and 3′ splice sites (ss) are shown, as well as the DR (direct repeat) element which is required for optimal packaging. B. Murine leukemia virus (MLV) genome—symbols as in A. C. Human hepatitis virus B (HBV) RNA genome show location of ϵ in relation to the known genes for the core protein, polymerase (Pol), surface protein (S) and X of unknown function. DR1 and DR2 are repeats at the ends important for reverse transcription but not packaging. D. Comparison of the putative HBV packaging element to that of HIV-1 TAR (Tat responsive element). The numbers refer to the number of nucleotides in each portion of the proposed structure. Data taken from ref 57 for HBVϵ and ref 85 for HIV-1 TAR.

which package cellular RNAs, there is some evidence that these are in a higher order complex.[49] Further studies on encapsidated heterologous RNAs containing ψ sequences devoid of dimerization signals are necessary to clarify this point. In avian retroviruses, the minimal ψ sequences are found on both genomic and subgenomic RNAs, yet the latter are not encapsidated (see Figure 3A). Sequences located at the 3′ ends of both the genomic and mRNAs are also involved in packaging;[67,68] it is likely that a complex interaction between the ends is required for packaging and this is only achieved in the full length genome. The situation for murine leukemia virus (MLV) is much simpler in that 5′ sequences required for packaging are removed by splicing and not found in subgenomic mRNAs.[69] HIV packaging signals seem to be complex. Deletion analyses originally located a small region between the 5′ splice site and *gag*.[70-72] However, this region does not appear to be required for packaging genomic RNA (P.P. Lee, M.L. Linial, submitted for publication). It is interesting to note that the TAR loop at the 5′ of the HIV genome has some structural similarity to the hepadnavirus ϵ structure (see Figure 3D), but the relationship of the TAR stem-loop to HIV-1 packaging is currently not known.

While the initial interaction of RNA and protein in hepadnaviruses appears to require only the viral polymerase, the situation in retroviruses is less clearly defined. By the process of elimination, the precursor of the retroviral structural proteins, PrGag, is thought to specifically interact with ψ; expression of *gag* alone leads to particles containing RNA. (See the chapter by E. Hunter, this issue.[73]) In the case of HIV specific binding of PrGag to 5′ genomic RNA has been demonstrated in an *in vitro* assay.[74,75] The major retroviral RNA binding protein NC, binds to RNA with high affinity but no specificity,[76] and

is only released from precursor PrGag after viral assembly. Protease activity is not required for packaging.[77,78] The domains in PrGag which actually interact with RNA and confer specificity are not known. The Zn finger motifs found in most (but not all) retroviral and retrotransposon NC proteins have been implicated in specific binding and packaging,[76,79-82] but in the case of RSV, specific RNA packaging can occur if they are deleted.[65] In retroviruses it is highly unlikely that Gag-ψ interaction is a trigger for viral assembly because under certain conditions particles devoid of genomic RNA can be produced[83] and viral mutants have been found which do not package genomic RNAs but produce particles. For instance, mutants of RSV lacking the terminal protease domain of PrGag produce particles which lack genomic or ψ-containing RNAs.[65,77,89] MLV with mutations in the nucleocapsid domain also are deficient in RNA encapsidation. Interestingly, some mutants continue to package other cellular RNAs related to the MLV genome which contain ψ-like sequences.[82] While details of the requirements and specificities of retroviral protein/genomic RNA interactions have thus far proved elusive, the great interest in developing antiviral agents for HIV and the exquisite specificity of this step in the viral life cycle, makes intervention at this step very attractive, and should be an impetus for more detailed examination.

Perspectives

Many of the chapters in this issue describe nucleic acid sequences or structures required in the encapsidation of a variety of different viruses. These nucleic acid/protein interactions are specific and, for those viruses associated with disease, encapsidation may be an important target for antiviral agents. An additional goal, particularly pertinent in the study of RNA/protein interactions, is the attempt to determine if there are general principles that can be applied in defining these interactions. Although there are many different DNA-binding proteins, it is often possible to identify a specific DNA-binding domain within the protein. Several different classes of DNA-binding domains have been identified and structural studies have accurately defined the DNA-amino acid contacts.[86] It has been more of a challenge to define domains in RNA binding proteins. X-ray crystallography and NMR data are available for a very few RNA/protein structures and only a few specific types of interactions have been identified. A large family of RNA-binding proteins share a consensus sequence or an RNA recognition motif.[87,88] Thus far these sequences have been found most frequently in SnRNPs, splicing factors and hnRNPs.[87,88] Some of the most detailed studies are those involving the interactions of the HIV-1 Tat and Rev proteins with their RNA elements TAR and RRE respectively (See Figure 3D). Studies of Tat-TAR interactions have demonstrated a crucial role for arginine in the interaction.[89,90] These studies, as well as those with Rev-RRE,[91] suggest an important role for the sugar phosphate backbone of the RNA in recognition.

The concept that RNA structure plays a crucial role in the recognition process has also received support in the RNA/protein interactions leading to assembly of RNA viruses. Extensive mutagenesis of the packaging signals of R17[92] and TMV[93] led to the conclusion that the structure of the RNA is essential in the recognition process. More recently, a bulged stem-loop structure was proposed to be the critical element for the encapsidation of the RNA of flock house virus, a small nonenveloped insect virus.[94] The importance of structure is now being explored further in the interaction of the viral RNAs and the viral proteins described in this and other chapters.

Acknowledgements

The research described in the authors' laboratories is funded by NIH grants AI11377 (S.S.), AI29984 (S.M.), CA 18282 (M.L.L.) and AI27291 (M.L.L.). We would like to thank Alan Rein, Bill Mason, Christoph Seeger and Peter Palese for their comments.

References

1. Caspar DLD, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spr Harb Symp Quant Biol 27:1-24
2. Fox JM, Johnson JE, Young MY (1994) RNA/protein interactions in icosahedral virus assembly. Semin Virol 5:51-60
3. Dubois-Dalcq M, Holmes KV, Rentier B (1984) Assembly of enveloped RNA viruses. Springer, New York
4. Schlesinger S, Schlesinger M (1990) Replication of Togaviridae and Flaviviridae, in Virology, vol 1 (Fields BN, Knipe DM, eds), pp 697-711. Raven Press, New York
5. Strauss EG, Strauss JH (1986) Structure and replication of the alphavirus genome, in The Togaviridae and Flaviviridae (Schlesinger S, Schlesinger MJ, eds), pp 35-90. Plenum Press, New York

6. Weiss B, Nitschko H, Ghattas I, Wright R, Schlesinger S (1989) Evidence for specificity in the encapsidation of Sindbis virus RNAs. J Virol 63:5310-5318
7. Garoff H, Frischauf A-M, Simons K, Lehrach H, Delius H (1980) The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino terminal region. Proc Natl Acad Sci USA 77:6376-6380
8. Rice CM, Strauss JH (1981) Nucleotide sequence of the 26S mRNA of Sindbis virus and deduced sequence of the encoded virus structural proteins. Proc Natl Acad Sci USA 78:2062-2066
9. Hahn CS, Strauss EG, Strauss JH (1985) Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid protein autoprotease. Proc Natl Acad Sci USA 82:4648-4652
10. Hahn CS, Strauss JH (1990) Site-directed mutagenesis of the proposed catalytic amino acids of the Sindbis virus capsid protein autoprotease. J Virol 64:3069-3073
11. Choi H-K, Tong L, Minor W, Dumas P, Boege U, Rossmann MG, Wengler G (1991) Structure of Sindbis virus core protein reveals a chymotrypsin like serine protease and the organization of the virion. Nature 354:37-43
12. Geigenmüller-Gnirke U, Nitschko H, Schlesinger S (1993) Deletion analysis of the capsid protein of Sindbis virus: identification of the RNA binding region. J Virol 67:1620-1626
13. Wengler G, Würkner D, Wengler G (1992) Identification of a sequence element in the alphavirus core protein which mediates interaction of cores with ribosomes and the disassembly of cores. Virology 191:880-888
14. Singh I, Helenius A (1992) Role of ribosomes in Semliki Forest virus nucleocapsid uncoating. J Virol 66:7049-7058
15. Boursnell MEG, Brown TDK, Foulds IJ, Green PF, Tomley FM, Binnis MM (1987) Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. J Gen Virol 68:57-77
16. Lai MMC, Stohlman SA (1978) RNA of mouse hepatitis virus. J Virol 26:236-242
17. Lee H-J, Shieh C-K, Gorbalenya AE, Eugene EV, La Monica N, Tuler J, Bagdzhadzhyan A, Lai MMC (1991) The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. Virology 180:567-582
18. Sturman LS, Holmes KV, Behnke J (1980) Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. J Virol 33:449-462
19. Lai MMC, Brayton PR, Armen RC, Patton CD, Pugh C, Stohlman SA (1981) Mouse hepatitis virus A59: mRNA structure and genetic localization of the sequence divergence from hepatotropic strain MHV-3. J Virol 39:823-834
20. Lai MMC, Baric RS, Brayton PR, Stohlman SA (1984) Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, cytoplasmic RNA virus. Proc Natl Acad Sci USA 81:3626-3630
21. Spaan W, Delius H, Skinner M, Armstrong J, Rottier P, Smeekens S, van der Zeijst BAM, Siddell SG (1983) Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. EMBO J 2:1939-1944
22. Sethna PB, Hofmann MA, Brian DA (1991) Minus-strand copies of replicating coronavirus mRNAs contain antileaders. J. Virol 65:320-325
23. Makino S, Yokomori K, Lai MMC (1990) Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal. J Virol 64:6045-6053
24. Makino S, Fujioka N, Fujiwara K (1985) Structure of the intracellular defective viral RNAs of defective interfering particles of mouse hepatitis virus. J Virol 54:329-336
25. Makino S, Shieh C-K, Soe LH, Baker SC, Lai MMC (1988) Primary structure and translation of a defective interfering RNA of murine coronavirus. Virology 166:550-560
26. van der Most RG, Bredenbeek PJ, Spaan WJM (1991) A domain at the 3′ end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs. J Virol 65:3219-3226
27. Fosmire JA, Hwan, K, Makino S (1992) Identification and characterization of a coronovirus packaging signal. J Virol 66:3522-3530
28. Stohlman SA, Baric RS, Nelson GN, Soe LH, Welter LM, Deans RJ (1988) Specific interactions between coronavirus leader RNA and nucleocapsid protein. J Virol 62:4288-4295
29. Perlman S, Ries D, Bolger E, Chang L-J, Stolzfus CM (1986) MHV nucleocapsid synthesis in the presence of cycloheximide and accumulation of negative strand MHV RNA. Virus Res 6:261-272
30. Robbins SG, Frana MF, McGowan JJ, Boyle JF, Holmes KV (1986) RNA-binding proteins of MHV: detection of monomeric and multimeric N protein with an RNA overlay-protein blot assay. Virology 150:402-410
31. Pons MW, Schulze IT, Hirst GK (1969) Isolation and characterization of the ribonucleoprotein of influenza virus. Virology 39:250-259
32. Kingsbury DW, Webster RG (1969) Some properties of influenza virus nucleocapsids. J Virol 4:219-225
33. Lamb RA (1983) The influenza virus RNA segments and their encoded proteins, in Genetics of Influenza Viruses (Palese P, Kingsbury DW, eds), pp 21-69. Springer, Vienna
34. Luytjes W, Krystal M, Enami M, Parvin JD, Palese P (1989) Amplification, expression, and packaging of a foreign gene by influenza virus. Cell 59:1107-1113
35. Hsu M-T, Parvin JD, Gupta S, Krystal M, Palese P (1987) Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. Proc Natl Acad Sci USA 84:8140-8144
36. Yamanaka K, Ishihama A, Nagata K (1990) Reconstitution of influenza virus RNA-nucleoprotein complexes structurally resembling native viral ribonucleoprotein cores. J Biol Chem 265:11151-11155
37. Honda A, Ueda K, Nagata K, Ishihama A (1987) Identification of the RNA polymerase-binding site on genome RNA of influenza virus. J Biochem 102:1241-1249
38. Smith GL, Hay AJ (1982) Replication of the influenza virus genome. Virology 89:506-516
39. Akkina RK, Chambers TM, Nayak DP (1984) Mechanism of interference by defective interfering particles of influenza virus: differential reduction of intracellular synthesis of specific polymerase proteins. Virus Res 1:687-702
40. Nakajima K, Ueda M, Sugiura A (1979) Origin of small RNA in von Magnus particles of influenza virus. J Virol 29:1142-1148
41. Hirst GK (1973) Mechanism of influenza recombination. I. Factors influencing recombination rates between temperature-sensitive mutants of strain WSN and the classification of mutants into complementation-recombination groups. Virology 55:81-93
42. Lamb RA, Choppin PW (1983) The gene structure and replication of influenza virus. Annu Rev Biochem 467-506
43. Scholtissek C, Rohde W, Harms E, Rott R, Orlich M, Boschek DB (1978) A possible partial heterozygote of an influenza A virus. Virology 89:506-516
44. Enami M, Sharma G, Benham C, Palese, P (1991) An influenza virus containing nine different RNA segments. Virology 185:291-298
45. Linial M (1987) Creation of a processed pseudogene by retroviral infection. Cell 49:93-102

46. Dornburg R, Temin HM (1988) Retroviral vector system for the study of cDNA gene formation. Mol Cell Biol 8:2328-2334
47. Jacks T (1990) Translational suppression in gene expression in retroviruses and retrotransposons, in Retroviruses Strategies of Replication (Swanstrom R, Vogt PK, eds), pp 93-124. Springer, Berlin
48. Oroszlan S, Luftig RB (1990) Retroviral proteinases, in Retroviruses Strategies of Replication (Swanstrom R, Vogt PK, eds), pp 153-186. Springer, Berlin
49. Linial M, Medeiros E, Hayward WS (1978) An avian oncovirus mutant (SE21Q1b) deficient in genomic RNA: biological and biochemical characterization. Cell 15:1371-1381
50. Boeke JD, Sandmeyer SB (1991) Yeast transposable elements, in The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Genome Dynamics, Protein Synthesis and Energetics (Pringle J, Jones E, Haber J, eds), pp 193-261. Cold Spring Harbor Press, New York
51. Mathias SL, Scott AF, Kazazian Jr HH, Boeke JD, Gabriel A (1991) Reverse transcriptase encoded by a human transposable element. Science 254:1808-1810
52. Huang MJ, Summers J (1991) Infection initiated by the RNA pregenome of a DNA virus. J Virol 65:5435-5439
53. Bartenschlager R, Junker-Niepmann M, Schaller H (1990) The P gene product of hepatitis B virus is required as a structural component for genomic RNA encapsidation. J Virol 64:5324-5332
54. Junker-Niepmann M, Bartenschlager R, Schaller H (1990) A short *cis*-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. EMBO J 9:3389-3396
55. Hirsch RC, Loeb DD, Pollack JR, Ganem D (1991) *cis*-acting sequences required for encapsidation of duck hepatitis B virus pregenomic RNA. J Virol 65:3309-3316
56. Nassal M, Schaller H (1993) Hepatitis B virus replication. Trends Microbiol, in press
57. Pollack JR, Ganem D (1993) An RNA stem-loop structure directs hepatitis B virus genomic RNA encapsidation. J Virol 67:3254-3263
58. Wang G-H, Seeger C (1993) A novel mechanism for reverse transcription in hepatitis B viruses. J Virol, in press
59. Nassal M (1993) The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. J Virol 66:4107-4116
60. Linial ML, Miller AD (1990) Retroviral RNA packaging: sequence requirements and implications. Curr Top Microbiol Immunol 157:125-152
61. Alford RL, Honda S, Lawrence CB, Belmont JW (1991) RNA secondary structure analysis of the packaging signal for moloney murine leukemia virus. Virology 183:611-619
62. Harrison GP, Lever AML (1992) The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. J Virol 66:4144-4153
63. Tounekti N, Mougel M, Roy C, Marquet R, Darlix J-L, Paoletti J, Ehresmann B, Ehresmann C (1992) Effect of dimerization on the conformation of the encapsidation Psi domain of Moloney murine leukemia virus RNA. J Mol Biol 223:205-220
64. Konings DA, Nash MA, Maizel JV, Arlinghaus RB (1992) Novel GACG-hairpin pair motif in the 5' untranslated region of type C retroviruses related to murine leukemia virus. J Virol 66:632-640
65. Aronoff R, Hajjar AM, Linial ML (1993) Avian retroviral RNA encapsidation: reexamination of functional 5' RNA sequences and the role of nucleocapsid cys-his motifs. J Virol 67:178-188
66. Marquet R, Baudin F, Gabus C, Darlix J-L, Mougel M, Ehresmann C, Ehresmann B (1991) Dimerization of human immunodeficiency virus (type 1) RNA: stimulation by cations and possible mechanism. Nucl Acids Res 19: 2349-2337
67. Sorge J, Ricci W, Hughes SH (1983) *cis*-acting RNA packaging locus in the 115-nucleotide direct repeat of Rous sarcoma virus. J Virol 48:667-675
68. Aronoff R, Linial ML (1991) Specificity of retroviral RNA packaging. J Virol 65:71-80
69. Mann R, Baltimore D (1985) Varying the position of a retrovirus packaging sequence results in the encapsidation of both unspliced and spliced RNAs. J Virol 54:401-407
70. Lever A, Gottlinger H, Haseltine W, Sodroski J (1989) Identification of a sequence required for efficient packaging of Human Immunodeficiency Virus type 1 RNA into virions. J Virol 63:4085-4087
71. Clavel F, Orenstein JM (1990) A mutant of Human Immunodeficiency Virus with reduced RNA packaging and abnormal particle morphology. J. Virol 64:5230-5234
72. Aldovini A, Young RA (1990) Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging results in production of noninfectious virus. J Virol 64:1920-1926
73. Hunter E (1994) Macromolecular interactions in the assembly of HIV and other retroviruses. Semin Virol 5:71-83
74. Luban J, Goff SP (1991) Binding of human immunodeficiency virus type 1 (HIV-1) RNA to recombinant HIV-1 gag polyprotein. J Virol 65:3203-3212
75. Hayashi T, Shioda T, Iwakura Y, Shibuta H (1992) RNA packaging signal of human immunodeficiency virus type 1. Virology 188:590-599
76. Fu X, Katz RA, Skalka AM, Leis J (1988) Site-directed mutagenesis of the avian retrovirus nucleocapsid protein, pp 12. J Biol Chem 263:2140-2145
77. Stewart L, Schatz G, Vogt VM (1990) Properties of avian retrovirus particles defective in viral protease. J Virol 64:5076-5092
78. Oertle S, Bowles N, Spahr PF (1992) Complementation studies with Rous sarcoma virus gag and gag-pol polyprotein mutants. Virology 66:3873-3878
79. Gorelick RJ, Henderson LE, Hanser JP, Rein A (1989) Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a 'zinc finger-like' protein sequence. Proc Natl Acad Sci USA 85:8420-8424
80. Gorelick RJ, Nigida Jr SM, Bess Jr JW, Arthur LO, Henderson LE, Rein A (1990) Noninfectious Human Immunodeficiency Virus type 1 mutants deficient in genomic RNA. J Virol 64:3207-3211
81. Meric C, Gouilloud E, Spahr P-F (1988) Mutations in Rous sarcoma virus nucleocapsid protein p12 (NCV): deletions of Cys-His boxes. J Virol 62:3328-3333
82. Meric C, Goff SP (1989) Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys-His box of the nucleocapsid protein. J Virol 63:1558-1568
83. Levin JG, Grimley PM, Ramseur JM, Berezesky IK (1974) Deficiency of 60 to 70S RNA in murine leukemia virus particles assembled in cells treated with Actinomycin D. J Virol 14:152-161
84. Oertle S, Spahr PF (1990) Role of the gag polyprotein precursor in packaging and maturation of Rous sarcoma virus genomic RNA. J Virol 64:5757-5763
85. Roy S, Delling U, Chen C-H, Rosen C, Sonenberg N (1990) A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated *trans*-activation. Genes Dev 4:1365-1373

86. Harrison SC (1991) A structural taxonomy of DNA-binding domains. Nature 353:715-719
87. Bandziulis RJ, Swanson MS, Dreyfuss G (1989) RNA-binding proteins as developmental regulators. Genes Dev 3:431-437
88. Kenan DJ, Query CC, Keene JD (1991) RNA recognition: towards identifying determinants of specificity. Trends Biochem Sci 16:214-220
89. Calnan BJ, Tidor B, Biancalana S, Hudson D, Frankel AD (1991) Arginine-mediated RNA recognition: the arginine fork. Science 1167-1171
90. Tao J, Frankel AD (1992) Specific binding of arginine to TAR RNA. Proc Natl Acad Sci USA 89:2723-2726
91. Bartel DP, Zapp ML, Green MR, Szostak JW (1991) HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. Cell 67:529-536
92. Uhlenbeck OC, Wu H-N, Sampson JR (1987) Recognition of RNA by proteins, in Molecular Biology of RNA: New Perspectives (Inouye M, Dudock BS, eds), pp 285-294. Academic Press, San Diego
93. Turner DR, Joyce LE, Butler PJG (1988) The tobacco mosaic virus assembly origin RNA. Functional characteristics defined by directed mutagenesis. J Mol Biol 203:531-547
94. Zhong W, Dasgupta R, Rueckert R (1992) Evidence that the packaging signal for nodaviral RNA2 is a bulged stem-loop. Proc Natl Acad Sci USA 89:11146-11150