

High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA

Gary W. Nelson,¹ Stephen A. Stohlman^{1,2} and Stanley M. Tahara^{1,2}

Departments of Molecular Microbiology and Immunology¹ and Neurology², USC School of Medicine, 2011 Zonal Avenue, Los Angeles, CA 90033-1054, USA

The nucleocapsid (N) protein of mouse hepatitis virus (MHV) is the major virion structural protein. It associates with both viral genomic RNA and subgenomic mRNAs and has structural and non-structural roles in replication including viral RNA-dependent RNA transcription, genome replication, encapsidation and translation. These processes all involve RNA–protein interactions between the N protein and viral RNAs. To better understand the RNA-binding properties of this multifunctional protein, the N protein was expressed in *Escherichia coli* as a chimeric protein fused to glutathione-S-transferase (GST). Biochemical analyses of RNA-binding properties were performed on full-length and partial N protein segments to define the RNA-binding domain. The full-length N protein and the GST–N protein fusion product had similar binding activities with a dissociation constant (K_d) of 14 nM when the MHV 5′-leader sequence was used as ligand. The smallest N protein fragment which retained RNA-binding activity was a 55 aa segment containing residues 177–231 which bound viral RNA with a K_d of 32 nM. A consensus viral sequence recognized by the N protein was inferred from these studies; AAUCYAAAC was identified to be the potential minimum ligand for the N protein. Although the core UCYAA sequence is often tandemly repeated in viral genomes, ligands containing one or more repeats of UCYAA showed no difference in binding to the N protein. Together these data demonstrate a high-affinity, specific interaction between the N protein and a conserved RNA sequence present at the 5′-ends of MHV mRNA.

Introduction

Coronaviruses are enveloped, positive-sense, ssRNA viruses. The 31 kb, ssRNA mouse hepatitis virus (MHV) genome encodes eight or nine viral genes. Each gene product is translated from a corresponding subgenomic mRNA, with the exception of the gene 1 product(s) translated from genomic RNA. Virions contain four to five structural proteins which include the 120 kDa surface glycoprotein (S), the 23 kDa membrane (M) protein, the 10.5 kDa envelope glycoprotein (E) and the 55 kDa nucleocapsid (N) protein (Lai & Cavanagh, 1997). The latter forms a helical nucleocapsid complex with the genome RNA. The intracellular genome ribonucleoprotein (RNP) complex is contained within a spherical, possibly icosahedral, core consisting mainly of the N and M proteins

(Risco *et al.*, 1996). A fifth component in the envelope of some MHV strains is the 65 kDa HE protein with homologies to haemagglutinin–esterase protein of influenza C virus (Vlasak *et al.*, 1988). It presumably serves a similar role to the influenza viral protein HN in the destruction of the viral receptor following virus entry and release.

Coronaviral mRNAs are synthesized via a discontinuous transcription process. MHV viral leader RNA is transcribed from the 3′-end of the (–)-sense genome RNA as an independent species 50–90 nt in length (Brayton *et al.*, 1982; Spaan *et al.*, 1993; Lai, 1990). Leader RNA is joined to subgenomic mRNA body sequences in a unique process either in *cis* or in *trans* at the initiation of transcription or shortly thereafter. Templates for virus transcription include the full-length genome as well as subgenomic replicons corresponding to viral genes. Several models have been proposed for this process; however, this is still an active area with no consensus (Shieh *et al.*, 1987; Sethna *et al.*, 1989; Sawicki & Sawicki, 1990; Jeong & Makino, 1994). Nonetheless, the result of virus

Author for correspondence: Stanley Tahara (at Department of Molecular Microbiology and Immunology). Fax +1 323 442 1721. e-mail stahara@hsc.usc.edu

transcription is production of a nested set of subgenomic mRNAs which have identical 5'-untranslated regions (UTRs) attached to body sequences with co-terminal 3'-ends.

The N protein is the major structural protein element of coronavirus virions and is encoded by gene 7 of the viral genome. Its mRNA is the most abundant viral mRNA species, which is consistent with high intracellular accumulation of the protein during infection (Lai & Cavanagh, 1997). It is phosphorylated *in vivo* (Stohlman & Lai, 1979) although nothing is known of the physiological consequence of such covalent modification. A primary function of N protein is formation of RNP complexes during assembly; in addition, it has been proposed to be multifunctional with additional roles in replication, transcription and translation (Tahara *et al.*, 1994, 1998; Lai & Cavanagh, 1997).

General nucleic acid-binding activity of the MHV N protein was first shown using an RNA blot overlay (NorthWestern) assay (Robbins *et al.*, 1986). Using a similar approach, it was found that the N protein had high affinity for the leader RNA in the presence of non-specific competitor RNA. These studies localized the specific RNA sequence for N protein association to the 3'-proximal region of leader RNA from bases 56 to 72 (Stohlman *et al.*, 1988), which includes the pentanucleotide repeat (UCUAA) critical for virus transcription (Lai & Cavanagh, 1997). Viral RNA associated with N protein during infection was analysed by co-immunoprecipitation. The data showed that all viral mRNAs, genomic and free leader RNA are equivalent in their ability to associate with N protein; the co-precipitated RNAs, thus, have one or more copies of tightly associated N protein (Baric *et al.*, 1988). Binding of N protein to viral mRNAs and genomic RNA presumably occurs through an interaction with viral leader RNA sequences, since the leader sequence is a common component of these RNA species. N protein was recently shown to bind the viral encapsidation signal found in the 3'-proximal region of gene 1b (Molenkamp & Spaan, 1997). The relationship of this activity of the N protein to its ability to bind leader RNA is not known.

Comparison of the N proteins of five MHV strains revealed that conserved amino acid residues are found in three distinct domains (numbering and net charge, in parentheses, refer to MHV-A59 N protein): domain I (1–139, basic); domain II (163–380, basic); and domain III (406–455, acidic) (Parker & Masters, 1990). Strain-specific sequence variations were found primarily in gaps between the three domains, suggesting that sequence conservation is needed to retain strain-independent N protein functions. Analysis of these conserved domains may yield insight into the many activities exhibited by this single viral protein. A general RNA-binding domain (RBD) of N protein was initially determined to reside within residues 136–397 (Masters, 1992). Subsequently, a specific RBD was determined by Nelson & Stohlman (1993) to reside between residues 169–308 of the A59 strain of MHV (MHV-A59) and confirmed via a molecular recombination approach. Analysis of chimeric N proteins derived from bovine coronavirus (BCoV)

and MHV suggests that RNA-binding activity resided in residues 163–380 (Peng *et al.*, 1995). The basic charge characteristic of this region is consistent with an RNA-binding function; however, it does not have homology to previously described RNA-binding motifs. Analysis of the RNA-binding properties of the N protein may provide insight into its functions as a potential *trans*-acting factor of viral mRNA translation and as a participant in other virus processes requiring RNA–protein interactions, e.g. transcription and encapsidation.

The data in this report demonstrate that the RBD of the N protein resides within a central 55 aa tract. In addition, the minimum RNA sequence serving as ligand for N protein has been identified based on specific, high affinity binding. This relationship of the N protein to its RNA ligand has important implications for virus processes which target this RNA sequence during infection.

Methods

■ **Bacterial expression of JHMV N protein.** All gene 7 constructions encoding the N protein and its derivatives were derived from the JHM strain of MHV (JHMV) (see Fig. 1 for designation of N protein domains).

Constructions of the N protein and its derivatives in pGEX-GL1 expression vectors (Gene Labs) were accomplished by PCR amplification of the desired domain with primers listed in Table 1. All primer pairs were designed to contain 5'-*Nco*I and 3'-*Sma*I/*Bam*HI sites. Amplified DNAs were double-digested with *Nco*I and *Sma*I or *Bam*HI and subcloned into pGEX-GL1, similarly digested with the same enzyme pair. For constructions using the *Sma*I site, the insert was first ligated to the *Nco*I site followed by ligation to the Klenow-blunted *Eco*RI site of the vector. This strategy positioned the N protein coding regions behind and in-frame with the glutathione-S-transferase (GST) gene. Fidelity of the product was examined by both radioimmunoprecipitation and Western blot assays using N protein-specific MAbs (Fleming *et al.*, 1983; Nelson & Stohlman, 1993; Stohlman *et al.*, 1994).

Full-length JHMV N protein was subcloned from pTM1-JN (Nelson, 1996), which contains full-length JHMV N protein in the pTM1 expression vector (Elroy-Stein & Moss, 1990), using primers GLN5' and N3'. The resulting N gene was modified by replacement of Ser² with Arg² and carboxy-terminal extension of the full-length protein by the dipeptide Pro-Gly. Scission of the GST moiety of the fusion protein by thrombin results in an N protein with an amino-terminal extension of Gly-Ser, since cleavage does not occur precisely at the GST–N protein junction. The DNA fragment corresponding to aa 1–169 (N^{1–169}; domain A) was amplified with primers GLN5' and A3'. The GST–A fragment had the same Ser² to Arg² substitution as GST–N and terminated with an additional Gly residue after Pro¹⁶⁹. Both inserts were cloned into *Nco*I/*Sma*I-digested vector. The DNA fragment corresponding to aa 177–295 (N^{177–295}; domain B) was amplified from pTM1-JN using primers GLB'5' and 3'B3'. In B, Thr¹⁷⁷ was preceded by Met-Ala; Asn²⁹⁵ was followed by Pro-Gly. The DNA fragment corresponding to aa 177–231 of N protein (N^{177–231}; domain B1) was amplified from pTM1-JN using primers GLB'5' and 5'B3'. B1 had the same amino-terminal extension as B and Pro²³¹ was followed by an Arg residue. The DNA fragment corresponding to aa 241–295 of N (N^{241–295}; domain B2) was amplified from pTM1-JN using primers 3'B5' and 3'B3'. B2 had a Met residue preceding Val²⁴¹ and Asn²⁹⁵ was followed by Pro-Gly. These

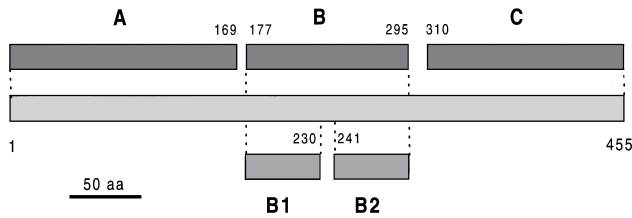


Fig. 1. Diagram of N protein segments expressed as GST fusion proteins. Domains A, B, and C of the N protein are indicated. Numbers refer to amino acid residues. All segments of N protein shown, including the full-length protein, were constructed and expressed as GST fusion products. Inclusive amino acid end-points are indicated.

three segments were subcloned between the *NcoI* and Klenow-blunted *EcoRI* sites of the vector. The DNA fragment corresponding to aa 310–455 (N^{310–455}; domain C) was subcloned from pTM1-C (Nelson, 1996). The *NcoI/BamHI* fragment from pTM1-C containing this fragment was cloned between the corresponding sites in the vector.

The N protein and several subdomains were also subcloned into pGEX-3X (Pharmacia Biotech). This vector differs from pGEX-GL1 in that a Factor Xa cleavage site separates the GST moiety from the carboxy-terminal fusion partner. Otherwise, the amino- or carboxy-terminal heterogeneities described above were maintained in this vector. The entire N protein was amplified from pTM1-JN using primers 3X-N5' and N3'. Domain A was amplified from pTM1-JN with primer 3X-N5' and A3'. Domain B was amplified from pTM1-JN using primers 3X-B'5' and B2.

Expression plasmids were transformed into BL21-DE3 (*ompT, lon*). For production of recombinant proteins, bacteria were induced with 0.5 mM IPTG, upon attaining an A_{600} of 0.2. Cells were grown for an additional 1–2 h at 37 °C prior to harvest. Bacterial pellets were resuspended in 9 ml ice-cold PBS (20 mM sodium phosphate pH 7.3, 150 mM NaCl containing 50 mM EDTA, aprotinin (100 kallikrein units/ml) and 2 mM PMSF, snap-frozen in dry ice/ethanol and rapidly thawed at 37 °C followed by sonication on ice with three 45 s bursts from a Branson sonicator. Lysates were adjusted to 1% Triton X-100 and incubated on ice for 20 min. Following centrifugation at 10 000 *g* for 10 min, supernatants were stored at –20 °C for 4 h, thawed and centrifuged at 10 000 *g* for 10 min to remove any cryoprecipitate. Clarified supernatants were mixed with 1 ml settled volume of glutathione–Sepharose 4B (Pharmacia) and

poured into a column constructed from a 10 ml disposable syringe. Column beds were washed with four, 10 ml aliquots of PBS containing 1% Triton X-100 and twice with 10 ml aliquots of TBS (50 mM Tris–HCl pH 7.3, 150 mM NaCl). Fusion proteins were eluted with 10 mM reduced glutathione (GSH) in TBS. Alternatively, bound fusion proteins were cleaved with the appropriate protease and eluted. Protein concentrations were determined by the Bradford method using BSA as standard (Bradford, 1976). Total expression of GST–N from 100 ml bacterial culture was typically 3–4 mg. Identities of expressed proteins were confirmed by comparison of actual vs expected sizes and immunoreactivity with N protein-specific MAbs (Stohlman *et al.*, 1994). After purification, fusion products were ≥ 90% pure (data not shown).

■ **RNA probe preparation.** [α -³²P]UTP (NEN; 3000 Ci/mmol) labelled RNA probes were synthesized by transcription of *HindIII*-linearized pBSL with T7 RNA polymerase, as previously described (Stohlman *et al.*, 1988). The run-off transcript is 153 nt in length and contains the first 114 nt of gene 6 mRNA of JHMV. Leader RNA was transcribed from ϕ XGL-1 (Tahara *et al.*, 1994) after linearization with *NcoI*. The resulting transcript was 92 nt in length and contained bases 1–72 of the 5'-leader sequence of gene 6 mRNA from JHMV.

For construction of the UCUAA repeat templates, oligoDNA molecules were synthesized with the sequences AGCTTAAGTTAG-ATTGAGCT (1R) and AGCTTAAGTTAGATTAGATTAGATTGAGCT (3R) cloned into *SacI/HindIII*-digested pGEM-3Zf(–). For preparation of RNA probes, the vector was linearized with *HindIII* and transcription was performed with T7 RNA polymerase. The resulting transcripts contained one copy (L1) or three copies (L3) of the pentamer motif and are 32 nt and 42 nt in length, respectively.

■ **Ligand binding assay.** Solution binding assays were performed in 50 or 100 μ l volumes containing 20 mM Tris–HCl (pH 8.0), 50 mM KCl, 2 mM DTT, 25 mM NaCl, 2 mM MgCl₂, 1 mM EDTA and 10 μ g/ml total RNA extracted from DBT cells as previously described (Nelson & Stohlman, 1993). Proteins were assayed at a concentration of 10–40 nM; ³²P-labelled RNA ligand concentrations varied from 0 to 50 nM. Binding was assayed at 22 °C for 10 min and was linear with increasing protein concentration up to 40 nM. Binding reactions in triplicate were terminated by filtering through pre-wetted nitrocellulose filters (25 mm; Schleicher and Schuell) followed by two washes with 1 ml TBS. Filters were dried and radioactivity was quantified using a Beckman LS-200 liquid scintillation spectrometer. Data were analysed using the EZ-Fit program (Frank Perrella, E. I. DuPont de Nemours). Data were fitted to a single

Table 1. OligoDNA PCR primers

Name	Sequence	Description
GLN5'	ATAGGATCCATGGCTTTTGTTCCTGGGCAAGAA	5' primer for N ^{1–454} and A (N ^{1–168})
N3'	TTACCCGGGCACATTAGAGTCATCTTCTAAC	3' primer for N ^{1–454}
A3'	TTACCCGGGGTCCCTTTTCACAATATCAGA	3' primer for A (N ^{1–168})
GLB'5'	ATAGGATCCATGGCTACTAGGTTTGC GCCCGGC	5' primer for B (N ^{176–295}) and B1 (N ^{176–295})
3'B3'	TTACCCGGGATTCTGATTGGGGCCTCTCTT	3' primer for B (N ^{176–295}) and B2 (N ^{241–295})
5'B3'	TTACCCGGGGTTTTACAGTAGAGGCAGGCTG	3' primer for B1 (N ^{176–230})
3'B5'	ATAGGATCCATGGTTTTGGCTAAGCTCGGTAAG	5' primer for B2 (N ^{241–295})
3X-N5'	ATATGGATCCTGTCTTTTGTTCCTGGGCAAGAA	5' primer for A (N ^{1–168}) and N (N ^{1–454}); pGEX-3X vector only
3X-B'5'	ATATGGATCCCTACTAGGTTTTCGCGCCCGGC	5' primer for B2 (N ^{241–295}); pGEX-3X vector only
B2	TTACCCGGGATTCTGATTGGGGCCTCTCTT	3' primer for B2 (N ^{241–295}); pGEX-3X vector only

substrate binding curve: $b = [(B_{\max} * s) / (K_d + s)]$, where b is the experimentally observed amount of bound ligand, s is the ligand concentration, K_d is the dissociation constant of ligand binding and B_{\max} is the amount of bound ligand at saturating ligand concentration.

■ **NorthWestern RNA blot analysis.** N protein, formic acid-cleaved fragments of N protein, GST-N protein fusion products and the corresponding cleaved products were separated by PAGE and transferred to nitrocellulose membranes as described previously (Nelson & Stohlman, 1993). ^{32}P -labelled transcripts were prepared from pBSL as described above. Competitor RNA, isolated from uninfected DBT cells, was added to the binding reaction at increasing concentrations to a maximum of 50 $\mu\text{g}/\text{ml}$ (Nelson & Stohlman, 1993). Bound RNA was detected by autoradiography with Kodak XAR X-ray film.

Results

Binding of leader RNA to the N protein

Binding constants for MHV RNA binding to N protein were obtained using a solution phase assay. Purified protein was incubated with increasing amounts of radiolabelled leader RNA. Typical primary binding data of such an experiment and double reciprocal plot are shown in Fig. 2(a). Binding of leader RNA to the N protein was saturable, showed first-order binding kinetics and was consistent with a single RNA-binding site per N monomer. Quantitative treatment of the data yielded an apparent K_d of 14 ± 4 nM ($n = 5$). The apparent B_{\max} of binding varied among experiments and typically gave values of 5–20% of the actual protomer input. Reduced and variable B_{\max} values were attributed to rapid protein inactivation after expression and isolation. Purified fusion proteins lost all activity within 7–10 days, even when stored at -70 °C. Therefore, the B_{\max} of a typical experiment represented the level of active protein in an individual preparation. A comparison of the binding activities of GST-N and purified N protein showed that the GST moiety had no effect on RNA binding (Fig. 2b). Hill coefficients for binding yielded values consistent with a single RNA-binding site per N protein monomer suggesting the absence of protomer cooperativity (Segel, 1976). Although higher multimers of N protein have been reported from virions (Robbins *et al.*, 1986; Nelson, 1996), no evidence of aggregation on non-denaturing polyacrylamide gels was observed (data not shown). These data provide the quantitative evidence for a high affinity, specific interaction between leader RNA and N protein and indicate that the N protein, consistent with previous data (Stohlman *et al.*, 1988; Nelson & Stohlman, 1993), contains a single RNA-binding site.

The RBD

N protein segments designated A, B and C, each containing approximately 30% of the N protein sequence, were examined for specific RNA binding to define the high affinity RBD. These segments correspond to fragments of N protein obtained after formic acid hydrolysis (Nelson & Stohlman, 1993). Only B

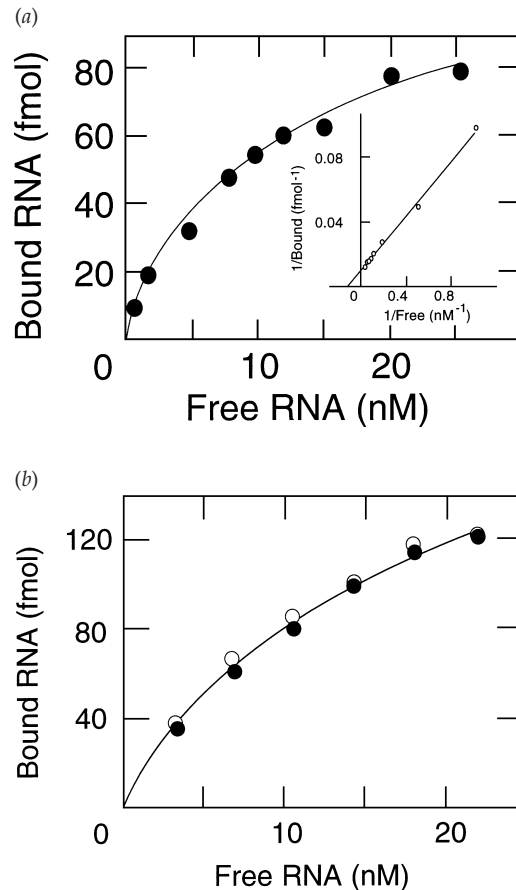


Fig. 2. RNA-binding activity of bacterially expressed N protein. (a) Bacterially expressed N protein was assayed for MHV leader RNA-binding activity by filter binding assay. The amount of protein in each binding reaction was 0.1 μg . Data presented in the primary binding plot are the average of assays in triplicate. The double reciprocal plot of the primary data is shown in the inset graph. An experimentally determined K_d of 14 nM was obtained. (b) Binding activity was compared between intact GST-N (\circ) and free N protein (\bullet) obtained by cleaving GST-N with Factor Xa and purification as described in Methods. Binding conditions were the same as those described in (a). Equimolar concentrations of GST-N and free N protein were used in binding reactions.

($\text{N}^{177-295}$) had RNA-binding activity; no RNA-binding activity was observed for either A (N^{1-169}) or C ($\text{N}^{309-255}$) segments (see Fig. 3), which is consistent with previously published results (Nelson & Stohlman, 1993). The B segment showed significant binding activity with a K_d of 32 nM (vs 14 nM for the N protein; see Fig. 2). Specific RNA binding by the GST-B fusion protein was compared to that by the GST-N protein. Both fusion proteins showed similar binding behaviour with control ligands. Neither GST-B or GST-N had any significant binding activity specific for DNA, non-viral RNAs, human α -globin 5'-UTR (Fig. 4), polylinker transcripts, the reverse complementary sequence of the MHV leader or yeast tRNA (not shown). These data demonstrate that the RNA-binding activity associated with B ($\text{N}^{177-295}$) is similar to that observed for full-length N protein.

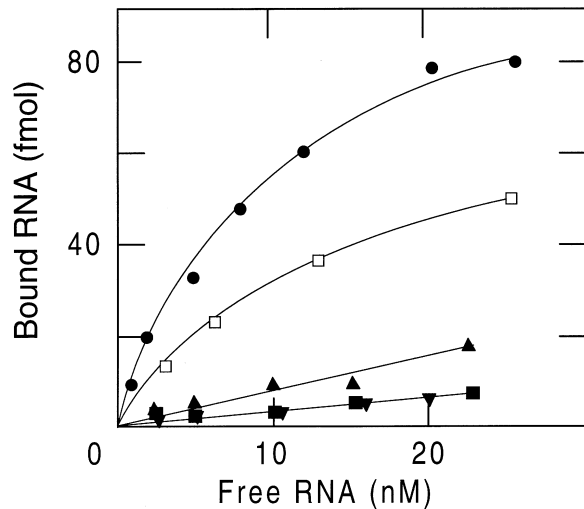


Fig. 3. RNA-binding activity of N protein fragments. GST fusion proteins for N protein, A, B and C were cleaved with thrombin and isolated for binding experiments. RNA-binding assays were as described in Methods. N (●) and B (□) were assayed at a concentration of 20 nM; A (■), C (▲) and GST (▼) were assayed at a concentration of 40 nM. A primary plot of binding data is shown.

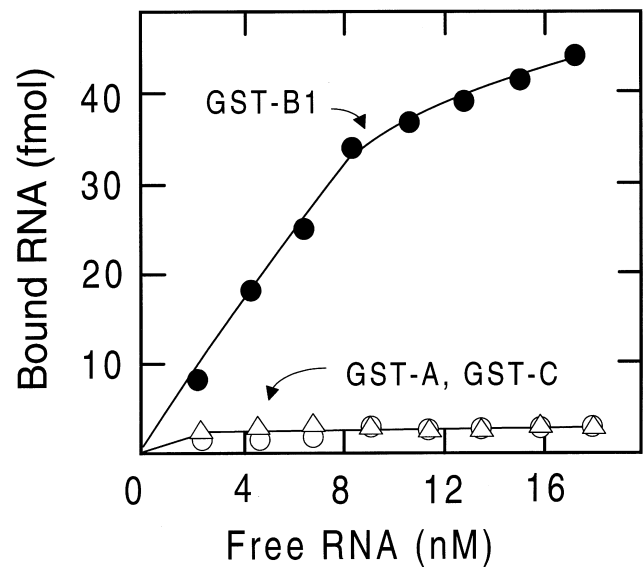


Fig. 5. RNA-binding activity of N protein subdomains. GST-A (○), GST-C (△) and GST-B1 (●) were assayed for RNA-binding activity as described above. The amount of each purified fusion protein in the binding reaction was 20 nM. The primary plot of binding data is shown.

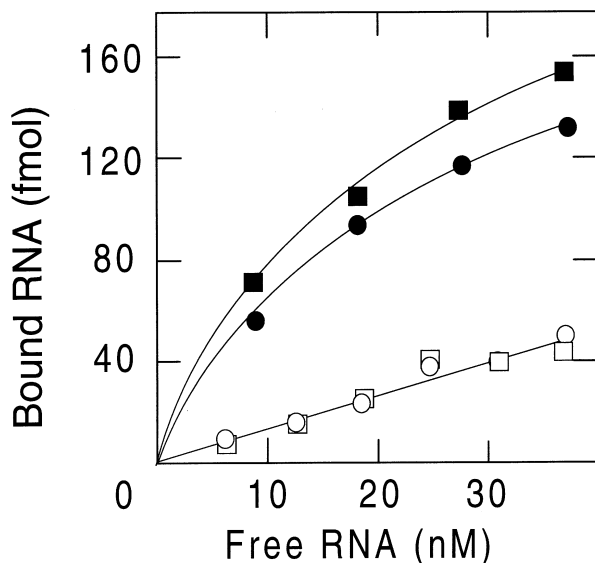


Fig. 4. Sequence specificity of RNA binding to N protein. GST-N (○, ●) and GST-B (■, □) were assayed with different RNA ligands. Binding to negative control human α -globin 5'-UTR (open symbols) was compared to MHV 5'-leader (closed symbols).

Location of the RBD

The B segment of N protein is highly conserved among coronaviruses (Britton *et al.*, 1988; Parker & Masters, 1990; Motokawa *et al.*, 1996; Homberger, 1995). In order to localize the RBD, the B segment (N¹⁷⁷⁻²⁹⁵) was subdivided into smaller segments, B1 (N¹⁷⁷⁻²³¹) and B2 (N²⁴¹⁻²⁹⁵). Both domains are highly conserved in coronavirus N proteins and have no apparent homologies to any previously described RNA-

binding proteins (Peng *et al.*, 1995). The GST-B1 and GST-B2 preparations were initially assayed for RNA-binding activity by NorthWestern blot assay. RNA-binding activity was observed for GST-B1 and GST-N but not for GST-B2 (not shown). These results suggested that RNA-binding activity resided within the B1 segment. Solution binding assay with the GST-B1 fusion protein demonstrated an apparent K_d of 14 nM for the viral leader RNA (Fig. 5), which is identical to the value obtained for the N protein. Neither the GST-A or GST-C segments exhibited any significant RNA-binding activity that was similar to that for GST-B or GST-B1. Based on the similar RNA-binding activities of N protein, B and B1, these results suggest that the amino-terminal subfragment B1, within domain B (N¹⁷⁷⁻²³¹), contains the leader RNA-binding site.

Characterization of the N protein-binding site on leader RNA

Previous results indicated that the 3'-proximal sequence of leader RNA is required for binding to the N protein (Baric *et al.*, 1988; Stohman *et al.*, 1988). A likely candidate ligand sequence within the 3'-proximal sequence is the conserved pentanucleotide repeat present within the 5'-UTRs of all MHV mRNAs. As shown in Fig. 6, deletion of 3'-proximal sequences containing two copies of the UCYAA pentanucleotide resulted in no specific N protein binding, thus providing direct evidence for a previously suggested 3'-proximal ligand domain (Stohman *et al.*, 1988). Other characteristics of the 3'-proximal sequences of leader RNA for N protein binding are shown in Fig. 6. Synthetic ligand RNAs containing one or three repeats

Leader Bases	K_d
1-114	16.1 ± 6.4 nM
1-72	18 ± 5.4 nM
1-32	>1.2 × 10 ⁵ nM
L1	14.7 ± 3.1 nM
L3	14 ± 4 nM

Fig. 6. Comparison of binding constants for different MHV RNA ligands. RNA-binding activity of GST-N was assayed with different MHV RNA ligands containing zero to three repeats of the UCUAA pentamer under standard binding conditions. Construction of templates for one (L1) and three repeat (L3) ligands is described in Methods. Ligands containing two copies of the UCUAA consensus were derived from JHMV mRNA 6 sequence. This mRNA has UCUAAUCCAA instead of a tandem repeat of UCUAA. The truncated leader (nt 1–32) ligand was prepared by linearizing the transcription template plasmid with *Sna*BI. A second two repeat ligand from JHM mRNA 6 leader sequence was used which contained only nt 1–72 of the viral leader. Actual sequences of ligands L1 and L3 are indicated. Dissociation constants (K_d) for binding of each ligand RNA to GST-N were determined from standard binding conditions with varying ligand concentrations. Primary binding data were analysed using the EZ-fit program as described earlier for determination of binding constants. Black regions indicate MHV leader sequences; positions of pentamer sequence are indicated as open boxes. Light grey shading indicates the truncated coding region of the JHM-M gene. Positions of pentamer repeats in L1 and L3 are underlined.

of the UCUAA pentamer were compared for N protein binding. It was found that a single copy of the UCUAA pentamer (L1) was sufficient for N protein binding; additional repeats of the UCUAA pentamer in the ligand (L3) had no effect on binding affinity. Binding of the smaller UCUAA monomer (L1) and trimer (L3) ligands, which have no significant secondary structure, with the same affinity as full-length, native leader containing two tandem copies of the UCUAA pentamer is consistent with an RNA-protein interaction based on sequence recognition. By comparison of the ligand sequences tested, it appeared that a minimum consensus RNA sequence for N protein binding was AAUCYAAAC.

Discussion

Comparisons of the N protein genes from a variety of coronaviruses suggest three highly conserved regions of amino acid sequences (Britton *et al.*, 1988; Parker & Masters, 1990; Motokawa *et al.*, 1996). Amino acid divergence is primarily found within two 'spacer regions' separating these domains (Parker & Masters, 1990). The functions of the less well conserved amino- and carboxy-terminal domains of the N protein remain unknown. However, the central third of the protein is responsible for RNA-binding activity (Nelson & Stohman, 1993).

Independent confirmation of the importance of the central region (B) for virus viability was obtained by analysis of BCoV and MHV recombinant viruses (Peng *et al.*, 1995). Functional N

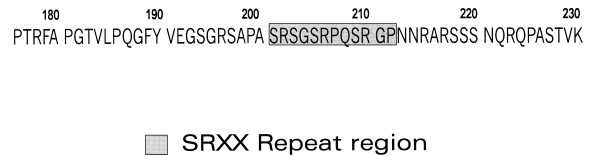


Fig. 7. RNA-binding domain of the N protein.

proteins were inferred by recovery of viable MHV. Interestingly, no viable MHV were obtained containing the central domain of the BCoV N protein, suggesting that these amino acids are required for strain-specific replication. It was initially proposed that aa 194–227 of the N protein (present in B1) bear homology to the SR superfamily of RNA-binding proteins (Parker & Masters, 1990). This motif is important in protein-protein interactions and is present in many splicing regulatory proteins (Fu, 1995). Data suggest that the N protein central domain was crucial for interactions with other strain-specific viral proteins and as a site for phosphorylation (Peng *et al.*, 1995). However, the N protein SR-rich region has limited homology to other SR proteins; therefore, it may be a subset of the SR family of proteins or totally unrelated. In contrast, the results of the present paper demonstrate the central domain, containing the SR-rich region is important in RNA-protein interactions. The K_d values determined for domain B1 (N^{177–231}; see below) indicate high affinity, RNA-binding activity specific for the 3'-proximal end of the MHV leader sequence. A distinguishing feature of this RNA-binding region is a tandem, triple repeat of the sequence SRXX located within residues 201–212 of N protein (Fig. 7). This motif is found in all coronavirus N proteins (Britton *et al.*, 1988; Parker & Masters, 1990; Homberger, 1995), supporting the notion of a conserved function within this amino acid sequence. In support of its role in RNA binding, preliminary studies using MAb J.3.1 (Fleming *et al.*, 1983; Stohman *et al.*, 1994), which recognizes an epitope contained within aa 171–196 of the MHV N protein, inhibited RNA-binding activity, whereas an isotype-matched control MAb specific for the S protein had no such effect (data not shown). Inhibition by J.3.1 is consistent with the RNA-binding site of the N protein located within or adjacent to the MAb-binding epitope.

The central amino acid domain of the N protein exhibits high affinity binding specific for a target ligand present on both genomic and viral subgenomic RNAs. RNA recognition by N protein provisionally requires AAUCYAAAC in the ligand and is independent of secondary or tertiary RNA structure. These binding data are consistent with the previous report that N protein-specific MAb co-precipitates RNAs containing this sequence from infected cells (Stohman *et al.*, 1988). Furthermore, the presence of the target ligand at the 5'-ends of all MHV mRNAs is consistent with the hypothesis that binding of N protein to viral mRNA is important in its function as a translational augmenter (Tahara *et al.*, 1994, 1998). The ability of N protein to bind to a consensus AAUCYAAAC sequence

suggests that this interaction of N with MHV RNA is not the rate-limiting step of virus transcription. Additional tandem copies of the UCYAA pentamer in intergenic sequences correlate with increased rates of initiation of virus transcription (Zhang *et al.*, 1994; van Marle *et al.*, 1995). Thus, our results are consistent with data indicating that other viral or cellular proteins are required for sequence discrimination during virus transcription (Furuya & Lai, 1993; Zhang & Lai, 1995; Li *et al.*, 1997).

The high affinity of the N protein for its RNA motif implies that genomic intergenic sequences, if not masked by RNA secondary structure or occupied by other proteins with significantly higher affinity, should also be available as N protein ligands. This suggests that multiple copies of N protein may bind to viral RNAs along their length at each intergenic sequence as protein levels rapidly accumulate during infection. The N protein has RNA-binding activity as a monomer; however, this does not preclude binding to larger N protein complexes. Robbins *et al.* (1986) reported that trimers of N protein present in virions exhibit nucleic acid-binding activity, suggesting that N protein self-association may be important for initiating RNP formation leading to encapsidation. Initial events in packaging may require: (1) binding of N protein to intergenic sequences in genome RNA; (2) collapse of this RNP into a more compact form via protein-protein interactions between N protomers; and (3) direction of this RNP assemblage to the ER for encapsidation via interaction with the M protein (Sturman *et al.*, 1980). Thus, the encapsidation signal, located at the 3'-proximal end of gene 1b in genomic RNA (Makino *et al.*, 1990; van der Most *et al.*, 1991; Fosmire *et al.*, 1992) would distinguish genome-sized RNPs from subgenomic-sized RNPs. The N protein binds the encapsidation signal (Molenkamp & Spaan, 1997); however, it is not known if the RBD we described for the N protein also recognizes this viral RNA feature or if binding is at an adjacent RNA domain.

The MHV N protein is required primarily for virion assembly. However, it has also been implicated in both RNA-dependent RNA transcription and translation (Lai & Cavanagh, 1997). MAb specific for the carboxy terminus of MHV N protein co-precipitates the following: leader RNA, all viral mRNAs and the 31 kb RNA genome from lysates of MHV-infected cells (Baric *et al.*, 1988; Stohlman *et al.*, 1994). These results suggested high avidity of N protein for a ligand common to each of these RNA species, i.e. the viral leader RNA. Analysis of RNA binding clearly establishes both high affinity and specificity of N protein for sequences within viral leader RNA and intergenic sequences of genomic RNA and subgenomic mRNA.

This work was supported by the NIH (NS 30880, NS 18146, AI 07078) and the Margaret E. Early Medical Research Trust. We thank Michael Lai, Therese Dietlin and Ite Laird-Offringa for critically reading the manuscript and providing helpful comments. We also thank Cornelia

Bergmann and Emmanuel Dimacali for sequencing the L1 and L3 plasmids.

References

- Baric, R. S., Nelson, G. W., Fleming, J. O., Deans, R. J., Keck, J. G., Casteel, N. & Stohlman, S. A. (1988). Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription. *Journal of Virology* **62**, 4280–4287.
- Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of proteins utilizing the principle of dye binding. *Analytical Biochemistry* **72**, 248–254.
- Brayton, P. R., Lai, M. M.-C., Patton, C. D. & Stohlman, S. A. (1982). Characterization of two RNA polymerase activities induced by mouse hepatitis virus. *Journal of Virology* **42**, 847–853.
- Britton, P., Cármenes, R. S., Page, K. W., Garwes, D. J. & Parra, F. (1988). Sequence of the nucleoprotein gene from a virulent British field isolate of transmissible gastroenteritis virus and its expression in *Saccharomyces cerevisiae*. *Molecular Microbiology* **2**, 89–99.
- Elroy-Stein, O. & Moss, B. (1990). Cytoplasmic expression system based on constitutive synthesis of bacteriophage T7 RNA polymerase in mammalian cells. *Proceedings of the National Academy of Sciences, USA* **87**, 6743–6747.
- Fleming, J. O., Stohlman, S. A., Harmon, R. C., Lai, M. M.-C., Frelinger, J. A. & Weiner, L. P. (1983). Antigenic relationships of murine coronaviruses: analysis using monoclonal antibodies to JHM (MHV-4) virus. *Virology* **131**, 296–307.
- Fosmire, J. A., Hwang, K. & Makino, S. (1992). Identification and characterization of a coronavirus packaging signal. *Journal of Virology* **66**, 3522–3530.
- Fu, X.-D. (1995). The superfamily of arginine/serine-rich splicing factors. *RNA* **1**, 663–680.
- Furuya, T. & Lai, M. M.-C. (1993). Three different cellular proteins bind to complementary sites on the 5'-end-positive and 3'-end-negative strands of mouse hepatitis virus RNA. *Journal of Virology* **67**, 7215–7222.
- Homberger, F. R. (1995). Sequence analysis of the nucleoprotein genes of three enterotropic strains of murine coronavirus. *Archives of Virology* **140**, 571–579.
- Jeong, Y. S. & Makino, S. (1994). Evidence for coronavirus discontinuous transcription. *Journal of Virology* **68**, 2615–2623.
- Joo, M. & Makino, S. (1992). Mutagenic analysis of the coronavirus intergenic consensus sequence. *Journal of Virology* **66**, 6330–6337.
- Lai, M. M.-C. (1990). Coronaviruses: organization, replication and expression of genome. *Annual Review of Microbiology* **44**, 303–333.
- Lai, M. M.-C. & Cavanagh, D. (1997). The molecular biology of coronaviruses. *Advances in Virus Research* **48**, 1–100.
- Li, H. P., Zhang, X., Duncan, R., Comai, L. & Lai, M. M.-C. (1997). Heterogeneous nuclear ribonucleoprotein A1 binds to the transcription-regulatory region of mouse hepatitis virus RNA. *Proceedings of the National Academy of Sciences, USA* **94**, 9544–9549.
- Makino, S., Yokomori, K. & Lai, M. M.-C. (1990). Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal. *Journal of Virology* **64**, 6045–6053.
- Masters, P. S. (1992). Localization of an RNA-binding domain in the nucleocapsid protein of the coronavirus mouse hepatitis virus. *Archives of Virology* **125**, 141–160.
- Molenkamp, R. & Spaan, W. J. (1997). Identification of a specific interaction between the coronavirus mouse hepatitis virus A59 nucleocapsid protein and packaging signal. *Virology* **239**, 78–86.

- Motokawa, K., Hohdatsu, T., Hashimoto, H. & Koyama, H. (1996).** Comparison of the amino acid sequence and phylogenetic analysis of the peplomer, integral membrane and nucleocapsid proteins of feline, canine and porcine coronaviruses. *Microbiology and Immunology* **40**, 425–433.
- Nelson, G. W. (1996).** *RNA binding characteristics of the MHV nucleocapsid protein*. PhD thesis, University of Southern California.
- Nelson, G. W. & Stohlman, S. A. (1993).** Localization of the RNA-binding domain of mouse hepatitis virus nucleocapsid protein. *Journal of General Virology* **74**, 1975–1979.
- Parker, M. M. & Masters, P. S. (1990).** Sequence comparison of the N genes of five strains of the coronavirus mouse hepatitis virus suggests a three domain structure for the nucleocapsid protein. *Virology* **179**, 463–468.
- Peng, D., Koetzner, C. A., McMahon, T., Zhu, Y. & Masters, P. S. (1995).** Construction of murine coronavirus mutants containing interspecies chimeric nucleocapsid proteins. *Journal of Virology* **69**, 5475–5484.
- Risco, C., Anton, I. M., Enjuanes, L. & Carrascosa, J. L. (1996).** The transmissible gastroenteritis coronavirus contains a spherical core shell consisting of M and N proteins. *Journal of Virology* **70**, 4773–4777.
- Robbins, S. G., Frana, M. F., McGowan, J. J., Boyle, J. F. & Holmes, K. V. (1986).** RNA-binding proteins of coronavirus MHV: detection of monomeric and multimeric N protein with an RNA overlay-protein blot assay. *Virology* **150**, 402–410.
- Sawicki, S. G. & Sawicki, D. L. (1990).** Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. *Journal of Virology* **64**, 1050–1056.
- Segel, I. H. (1976).** *Biochemical Calculations*, 2nd edn, pp. 309–312. New York: John Wiley.
- Sethna, P. B., Hung, S. L. & Brian, D. A. (1989).** Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proceedings of the National Academy of Sciences, USA* **86**, 5626–5630.
- Shieh, C.-K., Soe, L. H., Makino, S., Chang, M.-F., Stohlman, S. A. & Lai, M. (1987).** The 5′-end sequence of the murine coronavirus genome: implications for multiple fusion sites in leader-primed transcription. *Virology* **156**, 321–330.
- Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B. A. & Siddell, S. G. (1983).** Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO Journal* **2**, 1839–1844.
- Stohlman, S. A. & Lai, M. M.-C. (1979).** Phosphoproteins of murine hepatitis viruses. *Journal of Virology* **32**, 672–675.
- Stohlman, S. A., Baric, R. S., Nelson, G. N., Soe, L. H., Welter, L. M. & Deans, R. J. (1988).** Specific interaction between coronavirus leader RNA and nucleocapsid protein. *Journal of Virology* **62**, 4288–4295.
- Stohlman, S. A., Bergmann, C., Cua, D., Wege, H. & van der Veen, R. (1994).** Location of antibody epitopes within the mouse hepatitis virus nucleocapsid protein. *Virology* **202**, 146–153.
- Sturman, L. S., Holmes, K. V. & Behnke, J. (1980).** Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. *Journal of Virology* **33**, 449–462.
- Tahara, S. M., Dietlin, T. A., Bergmann, C. C., Nelson, G. W., Kyuwa, S., Anthony, R. P. & Stohlman, S. A. (1994).** Coronavirus translational regulation: leader affects mRNA efficiency. *Virology* **202**, 621–630.
- Tahara, S. M., Dietlin, T. A., Nelson, G. W., Stohlman, S. A. & Manno, D. J. (1998).** Translation effector properties of mouse hepatitis virus nucleocapsid protein. *Advances in Experimental Medicine and Biology* **440**, 313–318.
- van der Most, R. G., Bredenbeek, P. J. & Spaan, W. J. M. (1991).** A domain at the 3′-end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs. *Journal of Virology* **65**, 3219–3226.
- van Marle, G., Luytjes, W., van der Most, R. G., van der Straaten, T. & Spaan, W. J. M. (1995).** Regulation of coronavirus mRNA transcription. *Journal of Virology* **69**, 7851–7856.
- Vlasak, R., Luytjes, W., Leider, J., Spaan, W. & Palese, P. (1988).** The E3 protein of bovine coronavirus is a receptor-destroying enzyme with acetyltransferase activity. *Journal of Virology* **62**, 4686–4690.
- Zhang, X. & Lai, M. M.-C. (1995).** Interactions between the cytoplasmic proteins and the intergenic (promoter) sequence of mouse hepatitis virus RNA: correlation with the amounts of subgenomic mRNA transcribed. *Journal of Virology* **69**, 1637–1644.
- Zhang, X., Liao, C.-L. & Lai, M. M.-C. (1994).** Coronavirus leader RNA regulates and initiates subgenomic mRNA transcription both in *trans* and in *cis*. *Journal of Virology* **68**, 4738–4746.

Received 26 April 1999; Accepted 7 September 1999